UNIVERSITY OF MARYLAND

GLENN L. MARTIN INSTITUTE OF TECHNOLOGY
A. JAMES CLARK SCHOOL OF ENGINEERING
DEPARTMENT OF MECHANICAL ENGINEERING

2181 Glenn L. Martin Hall
College Park, MD 20742
301.405.2410 TEL 301.314.9477 FAX
http://www.enme.umd.edu

April 10, 2025

Re: Previously Published Materials appearing in Thesis or Dissertation

Dean of the Graduate School,

The student Camille Levine (UID 115250581) has:

☐ NOT INCLUDED any previously published works within their thesis or dissertation.

☒ INCLUDED one or more previously published works within their thesis or dissertation. This letter certifies that the examining committee for the student has determined that the student made a substantial contribution to the previously published work. The inclusion of the previously published work has the approval of the thesis or dissertation advisor and the Graduate Director.

Sincerely,

Katrina Groth
Director of Graduate Studies
Reliability Engineering
University of Maryland

Sincerely,

Katrina Groth
Advisor for Camille Levine

ABSTRACT

Title of Dissertation: EXPANDING THE CAUSAL LOGIC FOUNDATIONS OF HUMAN RELIABILITY ANALYSIS

Camille S. Levine
Doctor of Philosophy, 2025

Dissertation Directed by: Associate Professor Katrina M. Groth
University of Maryland Center for Risk and Reliability

Human reliability analysis (HRA) seeks to identify opportunities for operator error in complex engineering systems and quantify the probabilities of these errors occurring. Achieving accuracy and realism in HRA is a challenging process due to the lack of validated models and data available to analysts. First, it is necessary to design HRA methods capable of addressing the complexity of human cognition and the wide variety of operational scenarios where human errors are critical. HRA models should capture the multitudes of causal pathways to failure and their probabilistic nature. Second, error probabilities quantified through data-driven methods result in more accurate, traceable estimates; it is important that HRA methods completely and transparently incorporate human reliability data into the error quantification process.

This project develops robust causal logic models with a strong basis in cognitive psychological research that are quantified using human reliability data and engineering literature. This research has produced three literature-substantiated Bayesian network models that characterize human-machine error pathways and the mechanisms through which they occur. The network

structures are parameterized with a variety of data to render them capable of quantifying human failure event (HFE) probabilities. Several data resources were synthesized, including nuclear power plant training simulator data, expert knowledge, HRA dependency idioms, and psychological literature. The contextual information in these quantitative resources allowed derivation of conditional probabilities for performance influencing factors (PIFs), human failure mechanisms, and crew failure modes (CFMs). The applicability of these models is demonstrated through a multi-stage validation. The model structures are validated through expert discussions and against pre-analyzed nuclear event narratives from the ATHEANA HRA method, while the quantitative outputs are validated against German nuclear power plant operational experience. Then, the models are holistically evaluated against nuclear power plant simulator data and other established method outputs from the U.S. HRA Empirical Study. Lastly, the models are applied to an external flooding hazard scenario as an extension of the cognitively focused Phoenix HRA method. This work presents a novel application of new and existing HRA data sources to create validated causal models via Bayesian network structures. The resulting models provide a robust technical basis suitable for scenarios beyond those covered in conventional HRA, and are intended for integration into the Phoenix HRA method.

Expanding the Causal Logic Foundations of Human Reliability Analysis

by

Camille S. Levine

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2025

Advisory Committee:
Associate Professor Katrina M. Groth, Chair/Advisor
Associate Professor Michelle Bensi, Dean's Representative
Professor Mohammad Modarres
Professor Peter Sandborn
Assistant Professor Yunfei Zhao

Acknowledgments

This material is based upon work supported under a Department of Energy, Office of Nuclear Energy, University Nuclear Leadership Program Graduate Fellowship. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the Department of Energy Office of Nuclear Energy. I would also like to thank the A. James and Alice B. Clark Foundation for their financial and professional support through the Clark Doctoral Fellows program at the University of Maryland.

I could not have completed this journey without the unfailing guidance and support of my advisor, Dr. Katrina Groth. Thank you for always advocating for my work, and for always pushing me just enough out of my comfort zone so that I could grow and thrive as a researcher. You have provided me with countless valuable opportunities– both technical and non-technical– to cultivate my knowledge base and professional network, for which I am eternally thankful.

Thank you also to my committee members Dr. Michelle Bensi, Dr. Mohammad Modarres, Dr. Peter Sandborn, and Dr. Yunfei Zhao. Your classes have illuminated my understanding of the core principles and the subtleties of reliability engineering, shedding light on the breadth of topics in this field. I am extremely grateful for your thoughtful questions and constructive criticism that have helped to shape my work.

Thank you to my technical collaborators, both from within the SyRRA Lab and from outside the lab. Without the mentorship and assistance of Dr. Vinnie Paglioni and Dr. Ahmad

Al-Douri, my understanding of HRA and the models resulting from this project would be yet incomplete. Many thanks to Garill Coles and the risk assessment team at PNNL, for providing me with necessary data, requesting my input on regulatory guidance documents, and welcoming me to Richland for the summer.

And finally, a sincere and special thanks to the friends and family who stood by my side these past four years. To my brilliant boyfriend Ryan, you have provided emotional stability, companionship, and levity, exactly when I needed it most. I love you to R_H and back. To my mom, dad, Grandma JoAnn, and Grandpa Bill, thank you for supporting my academic endeavors since day one and for always telling me I could do anything that I set my mind to. To the Limestone lads: Eileen, Heidi, Kelly, Max, and Zoee- you've witnessed all the highs and lows of my grad school experience, and College Park would not have been "home" without you. To my friends from the SyRRA Lab, especially Vinnie, Andres, Colin, Ahmad, Sami, Cris, Ava, Ruochen, Suroosh, Victoriia, and Lauren, it was a great pleasure to work alongside you, and with every day that passes, I am even more proud to count myself among such an accomplished group.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ATHEANA	A Technique for Human Error Analysis
BN	Bayesian network
CCW	component cooling water
CDF	core damage frequency
CESA	Commission Errors Search and Assessment
CFM	crew failure mode
CPT	conditional probability table
CREAM	Cognitive Reliability and Error Analysis Method
CRT	crew response tree
EFW	emergency feedwater
EOC	error of commission
EOO	error of omission
ET	event tree
FT	fault tree
GRS	Gesellschaft für Anlagen- und Reaktorsicherheit
HEART	Human Error Assessment and Reduction Technique
HEP	human error probability
HFE	human failure event
HRA	human reliability analysis
HSI	human-system interface
IDAC	Information-Decision-Action in Crew Context
IDHEAS	Integrated Decision-Tree Human Event Analysis System
IDHEAS-ECA	Integrated Human Event Analysis System for Event and Condition Assessment
LERF	large early release frequency
LOCA	loss of coolant accident
LOFW	loss of feedwater

MCF	major crew function
MMA	morale, motivation, and attitude
NPP	nuclear power plant
NRC	Nuclear Regulatory Commission
PDP	positive displacement pump
PIF	performance influencing factor
PPA	personal and physical abilities
PRA	probabilistic risk assessment
RCP	reactor coolant pump
SACADA	Scenario Authoring Characterization and Debriefing Application
SAMG	severe accident managament guidelines
SF	situational factor
SGTR	steam generator tube rupture
SPAR-H	Standardized Plant Analysis Risk- Human Reliability Analysis
SSCs	structures, systems, and components
THERP	Technique for Human Error Rate Prediction
TOE	Training Objective Element
XHPRA	external hazards probabilistic risk assessment

Chapter 1: Introduction

Human reliability analysis (HRA) is the study of human-machine teams' interaction within complex engineering systems. HRA was originally developed to quantify the probability of human errors in nuclear power plant (NPP) control rooms as part of the probabilistic risk assessment (PRA) process. Identifying opportunities for human failure events (HFEs), and mitigation strategies to address deficient conditions, is essential to maintaining system safety.

HFEs are caused by an wide variety of factors and many different causal pathways, but current HRA methods do not model the full range of causal chains relevant to HFEs. With advances in human factors psychology and computing capacity, there is a need to incorporate more psychological realism and new data into causal HRA models. This work leverages these new tools to enhance HRA's causal logic basis.

1.1 Motivation

A large share of accidents across a variety of industries can be attributed to HFEs, or contexts conducive to human error. HRA is used in nuclear power (e.g., [1, 2]), aviation (e.g., [3], healthcare (e.g., [4]), and many other fields to characterize the potential causes, likelihood, and downstream effects of these errors. However, there has been an ongoing opportunity for the development of cognitively realistic, yet quantifiable, models that account for the full range of

contexts inherent to human operation [5–12]. While older models frequently lack this psychological realism, newer models may lack validated quantification schemes that are feasible in practice. Current research has sought to enhance this theoretical basis [2, 7, 13, 14], but is incomplete in its lack of explicit acknowledgment of causal chains leading to HFEs. This fundamental issue with HRA modeling manifests through three main research gaps. A literature review documented in Chapter 2 revealed additional technical gaps which will be explored in more depth and specificity.

1.1.1 GAP 1: Need for a causally robust theoretical basis

HRA methods in current use vary greatly in the strength of their causal basis and accuracy. Older HRA methods do not base their model structure in scientific theory or literature. Methods like Technique for Human Error Rate Prediction (THERP) and Standardized Plant Analysis Risk-HRA (SPAR-H) [1, 15] have extensive psychological literature reviews which are not incorporated into their resultant HRA models. Many newer models have a stronger theoretical basis that incorporates psychological literature, but there remains a need for treatment of causality [6, 7, 14]. It is important to document the causal pathways to failure, not only for improved method traceability and accuracy, but also so HRA practitioners can leverage results to develop more targeted human reliability improvement strategies for complex engineering systems.

1.1.2 GAP 2: Need for user-friendly, data-driven quantification schemes

Another opportunity for improving current HRA models is that many of their quantification schemes have been simplified, losing fidelity in order to maintain method usability. THERP has a human error probability (HEP) estimate for each task type, and SPAR-H assumes multipliers

for each performance influencing factor (PIF) state, but neither of these quantification methods are based in HRA data. These methods ease the mathematical burden on the analyst, and their low computational requirements contribute to their ubiquity in current PRA and HRA practice.

Although they have a stronger psychological basis, HRA methods such as Cognitive Reliability and Error Analysis Method (CREAM) [16] or Integrated Decision-Tree Human Event Analysis System (IDHEAS) [17], tend to be harder to quantify by hand. Software interfaces have very recently been developed for these methods ([17, 18]) but it will take time to adopt these approaches in regulatory practice.

The Information-Decision-Action in Crew Context (IDAC) method [19] has a strong psychological basis, but like CREAM and IDHEAS, has a necessarily complicated quantification process. The Phoenix method [10] is a newer method that strikes a good balance between solid theoretical foundations and straightforward quantification procedures, but still needs a more robust underlying causal model and a more intuitive quantification model to be practical for analyst use. There is a need for usable, analyst-friendly quantification schemes derived from HRA data so that resulting HEP estimates reflect the true likelihood of human error.

1.1.3 GAP 3: Need for verified model applicability in a range of scenarios

Finally, there is a need for HRA models that apply to a broad array of scenarios. Tasks that are assumed to be skill-of-the-craft, such as maintenance tasks or external hazard mitigation actions, represent a significant opportunity for human failure events. However, the vast majority of HRA methods were designed for nuclear power plant control room actions, which only represent a fraction of the potential human-machine failures possible in complex engineering systems. The

opportunities and consequences of HFEs both inside and outside the control room vary greatly. Data collection for ex-control room scenarios presents significant logistical challenges that has delayed the development of quantitative HRA methods capable of addressing ex-control room actions [20]. An HRA method should be able to accommodate an expanded repository of actions, and should likewise be validated for use in a variety of applications.

1.2 Research Objectives

This work has three objectives to address these specific gaps in the field of HRA. By addressing these research gaps, we construct HRA models that are accurate, causally based, and useful to analysts. An ideal HRA method would be comprehensive, research-based, adaptable, and multipurpose [6]. In this work, we achieve these desirable method qualities by addressing three main objectives.

1.2.1 OBJECTIVE 1: Identify and model causal logic pathways leading to human failure events

Currently, there are no HRA methods that fully acknowledge all of the causal pathways through which human failure events are brought about. The IDAC model [21] recognizes a need for documenting the various influences that contextual factors exert on human performance, but cautions that the analysis is preliminary. The Phoenix method [10] seeks to address this through its causal models, but leaves room for further consideration of error pathways that this work aims to solve. For an HRA method to provide accurate HEP values, the model must first be scientifically accurate and complete. To address this gap through Objective 1, a wide body of

HRA methods and literature are synthesized and used to develop human performance models with a strong basis in causal logic.

1.2.2 OBJECTIVE 2: Leverage HRA data resources to parameterize causal models

Objective 2 builds upon these causal logic structures to advance models capable of quantifying HEPs. Ideally, an HRA method should also be able to predictively quantify the likelihood of different human failure mechanisms and crew failure modes (CFMs), as well as retrospectively identify factors affecting error probabilities. A method capable of identifying specific failure modes and their contributing factors would allow an analyst to select more targeted mitigation strategies. This has been a long-standing gap in HRA, as the available data sources to draw upon are few, sparse, and not standardized. To address this gap through Objective 2, we parameterize the networks with a variety of rigorously selected and processed data.

1.2.3 OBJECTIVE 3: Validate and demonstrate usability of the causal models

To ensure that the networks are both valid to real-world scenarios and usable for analysts, Objective 3 is to validate the models against several types of operational data. The models were validated qualitatively, using descriptions of adverse event progression, and quantitatively, using error data from both emergency and non-emergency scenarios. Due to the growing breadth of HRA scope, it is essential that a method should be as scenario-agnostic as possible without sacrificing accuracy. To address Objective 3, we validated the models against operational experience narratives and data, and against simulator data and other HRA methods' output to benchmark

their performance in several use cases.

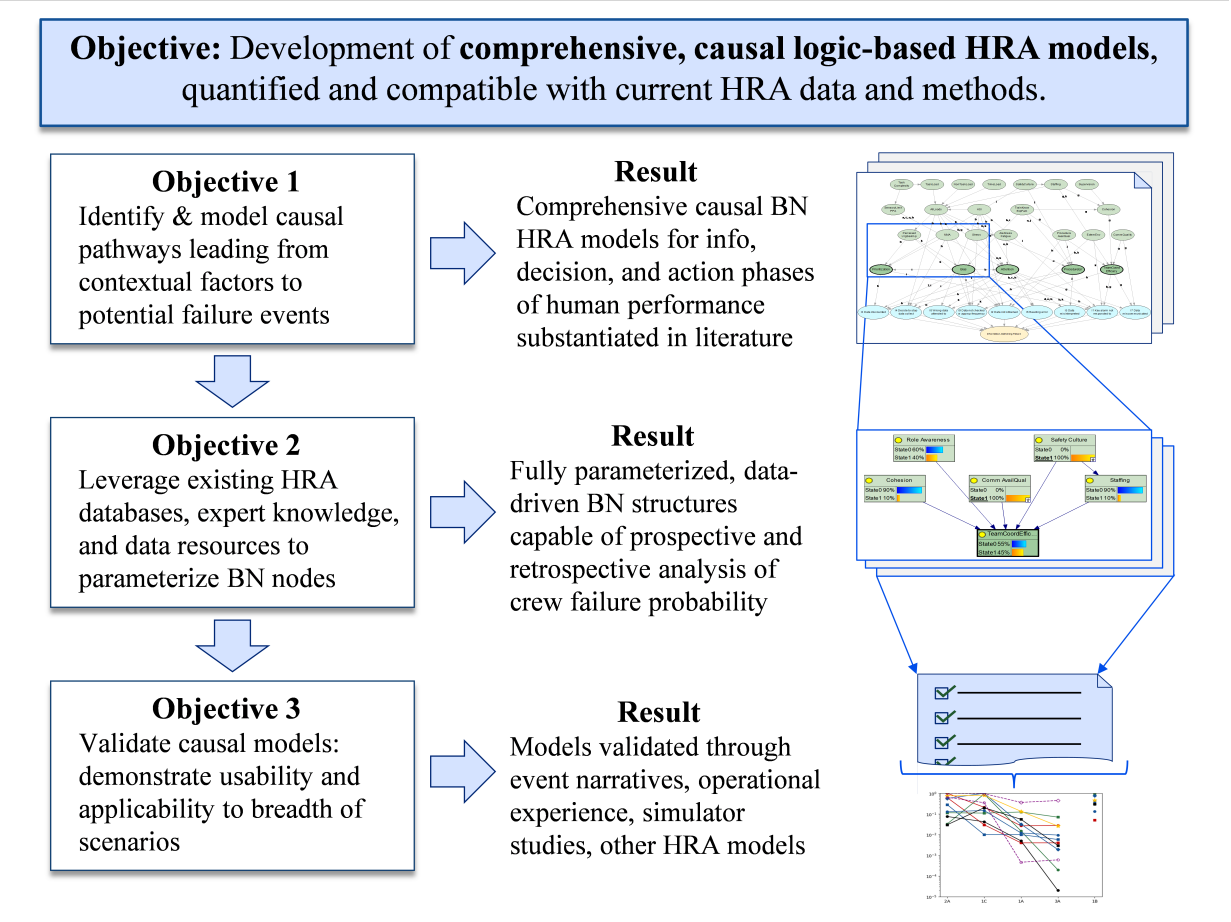

Figure 1.1: Summary of research objectives and results associated with each objective.

1.3 Results and Scientific Impact

The result of addressing each objective is summarized in Figure 1.1. The identification of causal pathways leading from scenario context to CFMs resulted in the development of Bayesian network (BN) models for human performance. These models, rooted in causal logic, comprehensively represent each cognitive stage of operator response. Each model is linked to a different stage of cognition, with a BN for each of the Information, Decision, and Action cognitive phases in accordance with [19]. CFMs from the Phoenix method [10], alongside two additional CFMs

identified in [22], are fully documented in their causal origins. This thorough treatment of causes and effects allows conceptual connections to be drawn between the antecedent conditions of each failure mode.

HRA data resources were leveraged to parameterize the network nodes so that the structures are capable of both prospective and retrospective analysis of human failure events. The models have improved traceability over other current HRA methods, with a strong basis in literature and data to deliver results grounded in reality. The model validation exercises demonstrated their accuracy to real-world events, usability to the analyst, and compatibility with current systems thinking. These models serve to improve upon traditional HRA, enabling plant safety assessment and improvement in a broader variety of scenarios.

The main scientific impact of this work is the most robust– and usable– HRA causal models to date. These are the first HRA models developed that explicitly incorporate causal logic concepts, cognitive pathways to failure, and recent HRA data. The resulting models demonstrate the synthesis of a wide variety of comprehensive literature and HRA methods, such as the Phoenix HRA method [10, 23] and Groth & Mosleh's taxonomy of performance influencing factors (PIFs) [24], and it has been the goal of these efforts to devise a similarly comprehensive method. Additional scientific impacts of this work involve a novel application of existing HRA databases, HRA dependency idioms, and other resources to quantify Bayesian logic structures, and a novel application of the Phoenix HRA framework to ex-control room actions to support the employment of this work in analyses beyond conventional HRA.

1.4 Structure of the Dissertation

The remainder of this dissertation describes the construction, capabilities, and validation of three related human reliability models. Chapter 2 details the methodology and gaps in HRA modeling identified from a literature review focused on methods for identifying and characterizing human failure events. Chapter 2 also provides an overview of Bayesian network methods and their use in HRA.

Chapter 3, Chapter 4, and Chapter 5 present the development and structure of information-gathering, decision-making, and action execution human performance Bayesian network models, respectively. Chapter 3 describes the method of developing the model structures, relevant data, and model parameterization in the greatest detail. In addition to the model construction method, Chapter 3 contains the structure, quantitative capabilities, and a structural validation of the information-gathering model against pre-analyzed operational event narratives. Chapter 4 describes the structure and quantitative capabilities of the decision-making model, and presents a quantitative validation exercise against NPP operational experience. Chapter 5 likewise describes the structure and quantitative capabilities of the action execution network, concluding with a summative validation against NPP simulator data and other HRA methods' output from the U.S. HRA Empirical Study. Chapter 6 presents a case study application of the BNs and the Phoenix method to manual actions for external flood hazard mitigation. Chapter 7 summarizes the contributions of this work and opportunities for future work. Finally, material used to substantiate the model structures is enumerated in Appendices A, B, and C, and all marginal and conditional probability values for the models are presented in Appendix D.

Chapter 2: Background and Literature Review

The majority of this chapter is reproduced from [25].

Camille S. Levine, Ahmad Al-Douri, Vincent Philip Paglioni, Michelle Bensi, and Katrina M. Groth. "Identifying human failure events for human reliability analysis: A review of gaps and research opportunities." In: *Reliability Engineering & System Safety* 245 (May 2024), p. 109967. ISSN: 0951-8320. DOI: 10.1016/j.ress.2024.109967.

2.1 Introduction

This chapter presents the methodology for the project's initial literature search, which focused on finding the gaps and opportunities in the identification of human failure events. First, the causal models of several currently used HRA methods are discussed, followed by a description of the HRA variables to be used in the rest of this proposal. Next, the methodology of the literature search is described, and the identified research opportunities enumerated. Lastly, Bayesian networks are introduced as a tool for HRA modeling.

HRA is a complex and dynamic field with both quantitative and qualitative means for assessing and addressing human operators' potential for contributing to failure in engineered systems. HRA is frequently used as a component in PRA for quantifying human error probabilities (HEPs). However, the PRA typically leverages HRA quantitative outputs, and the qualitative

analyses generated are rarely employed outside of the HRA itself [26]. Engineering knowledge about error causes, interventions, and consequences is generated from the qualitative side of HRA, through the identification and characterization of human failure events (HFEs). This yields valuable insights beyond numerical probabilities. As such, the accuracy and completeness of qualitative HRA is of the utmost importance, as it not only underpins the accuracy of HEPs but also brings light to the myriad of potential accident sequences involving human actions, and the requisite system safety improvement strategies.

Human behavior is complex in nature, and is rendered even more complex in the context of an engineered system. This makes it impractical for HRA practitioners to enumerate every potential failure mechanism. Failures can be caused by both erroneous actions and also what appear to be reasonable, or expected, human actions. Errors may be obvious in hindsight but arise from reasonable actions given the operators' mental model and information available at the time [5]. A more complete picture of the causal factors and the context leading up to the event would allow for better identification of human failure events with root causes beyond the observable action level.

Traditionally, HRA has been used as a component within PRA for complex systems with human operators. Each PRA method guides the development of the level of detail of its associated HRA methods. For internal consistency, the basic failure events are generally equivalent for both humans and hardware. This means that a complete technical basis of HRA has not been developed independently from PRA, and there is a lack of standardization for many concepts [27]. Notably, the HFE is a typical unit of analysis across HRA methods but there is no established definition or procedure for use. This yields flexibility that allows analysts to tailor the definition to the needs of the scenario, however, this leads to significant inter-analyst and inter-method vari-

ability, particularly for complex HFEs [11, 12, 27]. The U.S. and International HRA Empirical Studies found that this variation stems from method-driven factors (e.g., the method's ability to capture underlying influences and properly characterize the analyst's understanding) as well as analyst-driven factors (e.g., their expertise and the depth of the qualitative analysis they choose to undertake) [12].

A modern HRA method should comprehensively address the full range of causal factors and contexts, build upon state-of-the-art research, adapt to source material changes, and serve multiple qualitative and quantitative purposes [6, 7]. Other key qualities of a method include traceability, credibility, and reproducibility [6, 7]. However, previous efforts have struggled with integrating these needs throughout the development of new methods and data sources. By enhancing HRA's causal basis in current reliability engineering, psychological, and organizational literature, analysts will be able to more easily justify their modeling choices, which should reduce inter-analyst variability. A quantification scheme tied to this causal basis will produce more credible HEP results, and in turn, more accurate PRA analyses. Research efforts in the near future should seek to address this next major challenge in the field of HRA.

This review synthesizes existing knowledge to articulate the major gaps and opportunities in HFE identification and highlight the need for an underlying causal basis and traceability in HRA. By enumerating the gaps and opportunities related to HFE identification, the associated technical basis can be systematically improved. HRA methods in practice are not capable of identifying all significant HFE sequences, as the safety significance of human actions is directly tied to the specific context and circumstances surrounding the event [28]. Addressing these long-standing issues in the treatment of HFEs will expand the range of scenarios that HRA is capable of characterizing and increase overall traceability and transparency.

2.2 Overview and Procedures for HFE Identification

2.2.1 Definition of the Human Failure Event

There have been multiple methods developed that include guidance for identifying HFEs in HRA scenarios, including A Technique for Human Error Analysis (ATHEANA), SPAR-H, Phoenix, and IDHEAS [1, 10, 17, 29]. However, before discussing these methodologies, it is beneficial to define the key concepts that underlie the study of HFEs. These include the HFE itself and the related causal constructs that decompose the HFE into an interpretable causal network.

The first important concept is the HFE, which has become the standard qualitative and quantitative unit of analysis in HRA since the introduction of the ATHEANA method in the mid-1990s. Previously, "human failure event" was used in PRA interchangeably with "human error," and was defined as randomly occurring operator error rather than the product of the event context [29]. Despite its widespread use, the HFE remained formally undefined until very recently [27], with HRA methods ascribing unique meaning and guidance for what an HFE should be in the given context. A joint ASME/ANS standard provides a definition for an HFE as used in the nuclear power PRA context [30]: "a basic event that represents a failure or unavailability of a component, system, or function that is caused by human inaction or an inappropriate action." This definition is useful for understanding how the HFE fits into the PRA event tree, but is of limited utility when attempting to identify an HFE directly from a scenario narrative, task analysis, or HRA model.

To rectify this issue and provide a robust definition of the HFE for HRA purposes, another definition was proposed in 2022 to unify HRA theory and practice. This definition states that

the HFE is the highest-level failure considered in an HRA model, "the failed state of some overarching objective... the culmination of the failure process, which is composed of at least one failed function-level element (e.g., MCF)." [27]. The definition neatly defines the HFE as the combination of a failed task objective state and the process leading up to that failure.

2.2.2 Cognitive Structure of HFEs and Associated Terminology

Additional terminology is needed to describe the building blocks of an HFE. Specific real-world cognitive processes are described by *major crew functions* (MCFs), which are high-level mental activities used to accomplish a goal [31]. Failure of a major crew function may initiate the occurrence of an HFE. MCFs build upon the idea of *macrocognitive functions*, representing a system-specific objective that if failed, leads to an HFE [27].

A wide variety of models have been created to categorize macrocognitive functions and describe their interactions, but a majority of these models include operator functions such as Detecting/Noticing, Understanding/Sensemaking, Decision-Making, Action-Taking, and Coordination/Teamwork in their framework [32]. The Information-Decision-Action in Crew Context (IDAC) HRA method [19] considers information-gathering, decision-making, and action-taking in its underlying model of cognition, supplemented with a simplified crew model to account for the Communication and Coordination functions. The model includes these three phases of operator response, with further nested cognitive activities in each phase that also correspond to Information, Decision, and Action tasks.

These failures of MCFs are caused by *crew failure modes* (CFMs), which are specific failure pathways, i.e., how the MCF failed state occurs. CFMs leading to the failure of a MCF are

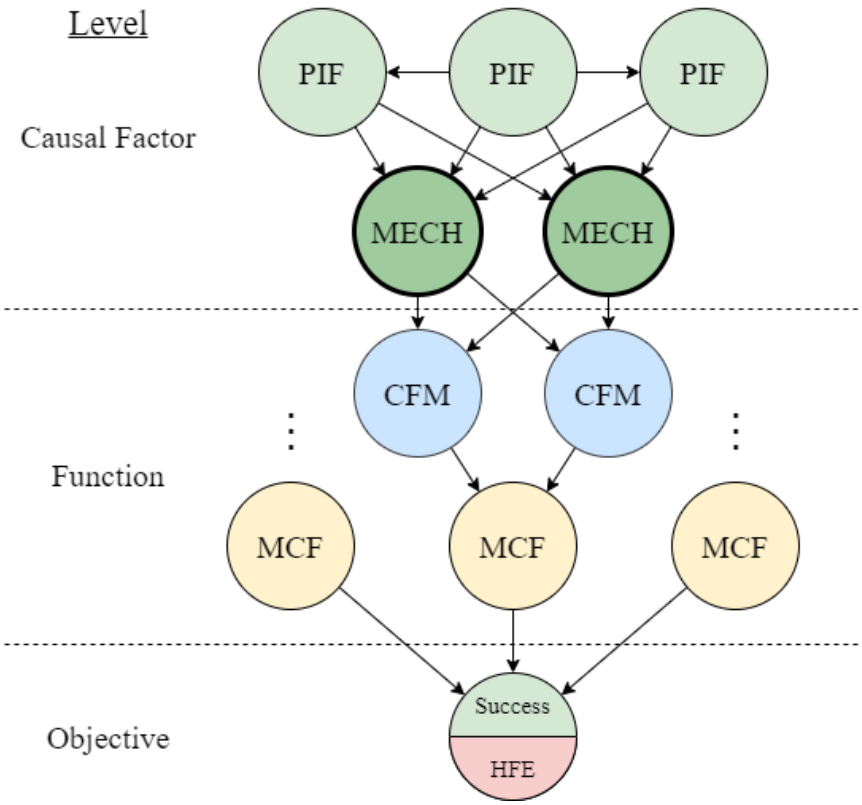

Figure 2.1: Causal structure of the HFE shown in a Bayesian Network framework, connecting PIFs as causes of crew failure mechanisms, in turn causing crew failure modes, which lead to MCFs, which define whether crews fail key objectives (HFEs).

orthogonal and mutually exclusive. CFMs describe a lower level of cognition than MCFs, and represent observable failure modes that manifest due to the human-machine team's error [31].

Human failure mechanisms represent the pathways through which CFMs might be brought about. Mechanisms represent real psychological or physical processes that explain the underlying reasons for failure, and are frequently represented via contextual factors [33].

The context and environment in which these failure mechanisms and failure modes operate are accounted for by *performance influencing factors* (PIFs). PIFs modify the human error probability (HEP) by modifying the probability that cognitive mechanisms fail [32]. These factors may be internal or external to the operator, and may fall broadly into the categories of organization-,

team-, person-, situation-, or machine-based causes [24]. Among existing HRA models, there are a wide variety of identified PIFs and PIF classification schemes, a discussion of which is outside the objectives of this dissertation. PIFs with immediate, direct effects on the operator are the most important to consider, because latent or time-delayed factors are implicitly modeled in the more immediate ones [34].

Recent efforts have sought to standardize lexicographical and mathematical definitions of dependency among HFE components [31], as well as an exhaustive, orthogonal set of relationships that describe all possible dependency relationships between these components [35]. Standard terminology is important in discussions of dependent and causal factors to facilitate compatibility between HRA methods. A common vocabulary also reduces the variability of results, improving a method's traceability and reproducibility by standardizing the analysts' treatment of scenario components [31]. This ensures that HFEs between scenarios are decomposed to the same level, and that this level can be quantified using standard tools available.

2.2.3 Qualitative Identification of HFEs through HRA Methods

As mentioned above, ATHEANA [29] is considered the origin of HFEs as the standard HRA unit of analysis. It defines HFEs as a representation of the failure of a function, system, or component resulting from unsafe human actions that places the plant in a worse condition. Within the ATHEANA HRA method, HFEs may result from one of many unsafe acts which, in turn, may result from one of several error mechanisms and error-forcing contexts. Different error mechanisms are associated with different phases of operator response to an abnormal event. ATHEANA describes the four phases of operator response corresponding to the macrocognitive

functions of detection, situation assessment, response planning, and response implementation. The error mechanisms associated with these phases are influenced by different sets of error-forcing contexts, which may be PIFs, plant conditions, accident scenarios, and scenario timings. Its authors state that applying the ATHEANA method will result in identifying HFEs previously unmodeled in HRA contexts, especially errors of commission (EOCs), for which they advocate for further definition in future works. A major limitation of ATHEANA is that it is concerned with post-initiator events only and allows the definition of a HFE to be determined by the level (functional, system, or component) of the PRA scenario being considered, limiting the scope of analysis and introducing the potential for variability of results.

The SPAR-H method [1] uses the ASME definition of the HFE as a basic event that represents a failure or unavailability of a component, system, or function that is caused by human inaction or an inappropriate action. Similar to ATHEANA, SPAR-H does not recommend a level of abstraction and suggests that the analyst use their own judgment to decompose the HFEs while applying a consistent level of decomposition. Defining HFEs as appropriate to the scenario being examined allows analysts to ensure that the HRA methods interface with the PRA method at an appropriate level of analysis. However, this would lead to a high level of variability amongst both different methods and different analysts (inter-analyst variability).

The Phoenix method [10] attempts to address this by clearly defining HFEs as a high-level objective failure that can be further decomposed into a set of defined CFMs and PIFs. Still, the method allows for HFEs to be identified as either part of an existing PRA model or in an iterative process as part of the HRA qualitative analysis. By allowing the definition of a HFE to be linked to the parent PRA method, inter-analyst variability is still highly likely. A newer method, the Integrated Human Event Analysis System for Event and Condition Assessment (IDHEAS-ECA)

[17], defines HFEs as the failure of one or more of the following five macrocognitive functions [32]: detection, understanding, decision-making, action execution, and interteam coordination. The level of abstraction of these macrocognitive functions is chosen by the analyst. Like the Phoenix method, IDHEAS-ECA defers to the parent PRA method and the analyst's judgment for identifying and defining the HFEs. This lack of guidance on identifying HFEs may lead to omitting pre-initiator and initiator events, discussed in depth in Section 2.5.5, which have also been lacking in prior methods.

2.3 Approach

A two-pronged approach was used to identify gaps in the identification of HFEs in HRA. First, a literature search was conducted in order to characterize the state of current research on HFEs and identify the most recent gaps on the topic. The primary method was a keyword search for "human failure event" on Google Scholar, which returned 470 results from sources including scholarly journals, conferences, and government reports from relevant federal agencies. Other filters appended to this search included "crew failure mode," "human reliability analysis," and "dependency". Secondary searches for "human failure event" were conducted in Web of Science and ScienceDirect, searching both the title and abstract fields for the phrase.

Although the Google Scholar search prioritized English-language papers that were more recent or relevant, papers were first filtered by source, including reliability journals and conferences known for publishing HRA content. This included the journals Reliability Engineering and Systems Safety, Journal of Risk and Reliability, Risk Analysis, and Safety Science; as well as the conferences European Safety and Reliability Conference (ESREL), and the International

Association on Probabilistic Safety Assessment and Management (PSAM). Materials published by IEEE, and DOE national laboratories through the Office of Science and Technical Information (OSTI) database, were also found to be relevant. A particular interest was taken in papers published by collaborators at the University of Maryland, University of California, Los Angeles, the U.S. Nuclear Regulatory Commission (NRC), Sandia National Laboratories, and Idaho National Laboratory, as these HRA developers and analysts tend to have standardized definitions and tools among themselves. The search was also centered on these resources because this work aims to improve upon U.S.-developed HRA methods.

A multi-layer approach was used to determine the relevance of each paper. Due to the large number of search results, the title and abstract were first scanned to ascertain whether the paper's content was relevant to the identifying and characterizing HFEs. If the publication appeared relevant after this first inspection, the full text was examined for the number of mentions of "HFE" or "human failure event," and papers with very few or sporadic mentions were discarded as not relevant. Finally, the full text was read for content, and selected for inclusion if it discussed the identification or level of detail of the HFE.

The second part of the approach involved ongoing project activities on identifying uncertainties in modeling human actions for external hazards probabilistic risk assessment (XHPRA) [36, 37]. During the development of HRA models for this project, several of the literature-identified gaps were identified and corroborated. Final project synthesis discussions also centered around how uncertainties associated with external hazard modeling propagate to identifying HFEs.

2.4 Overview of Identified Gaps

The results from the two activities described in Section 2.3 led to the synthesis of five major gaps in identifying human failure events in HRA models. Figure 2.2 shows a breakdown of the number of relevant publications highlighting each gap since 2004. Some publications identified multiple gaps as areas of need for research in HFE identification.

These gaps are summarized here, and Section 2.5 contains a more detailed discussion of each gap. First, quantitative, checklist-based HRA methods commonly used in PRA studies lack comprehensive HFE identification schemes. Second, while some progress has been made to develop research-based HRA methods, the majority of existing methods lack the scientific literature substantiation necessary for accurate and traceable identification and characterization of HFEs. Third, most HRA methods do not capture the complex nature of human cognition which leads to inaccuracies and inconsistencies in the HFE identification process. Fourth, HFE identification is primarily conducted in the context of single-unit PRA studies, but the increasing frequency of events in multi-unit NPPs and external hazards highlights the need for developing methods to understand the sources of, and model, HFEs in those contexts. Finally, given that most HRA methods do not explicitly provide for identifying EOCs, HRA analysts lack a standard approach towards identifying HFEs stemming from EOCs.

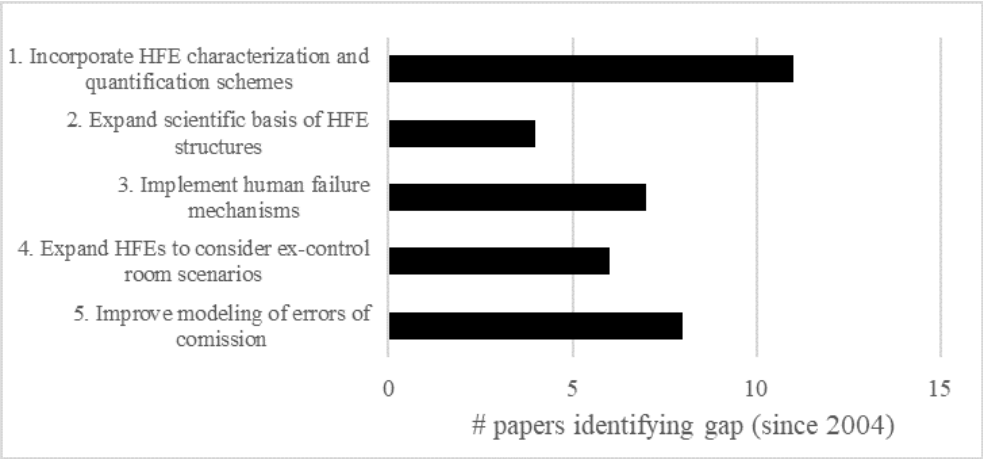

Figure 2.2: Number of relevant publications to the five main opportunities in HFE identification and characterization for HRA.

2.5 Gaps and Opportunities for Research

2.5.1 Incorporate robust HFE characterization and usable, accurate quantification schemes into HRA methods

A long-standing gap in HRA research is that the methods most accessible to analysts tend to lack a strong theoretical basis, and existing methods that allow for the comprehensive identification of HFEs tend to lack a straightforward quantification scheme. The "checklist" methods are the easiest for analysts to use during quantification, but they precariously leave a considerable portion of the qualitative analysis up to individual judgment. Although these methods allow for an expedited HRA process, they sever the bridge between the qualitative and quantitative functions HRA serves in PRA. Both of these purposes are extremely important to plant safety: characterization of HEPs allows for the inclusion of the human element in the quantitative assessment of overall plant risk, and identification of specific contextual factors leading to HFEs enables performance-improving measures to be taken [6].

Older HRA methods address a need for a simple, computationally cheap quantification scheme, such as the THERP, SPAR-H, or Human Error Assessment and Reduction Technique (HEART) methods [1, 15, 38]. These methods are the most readily accessible and easiest to understand due to their simplicity. As such, they have been used for the longest time, and are the most widely understood and accepted. Although some of these methods have a relatively strong qualitative HFE identification scheme compared to others, findings from the U.S. HRA Empirical Study noted that qualitative analysis is a shared weakness across all HRA methods [39]. HRA researchers have sought to address this gap [13], but current approaches are still lacking in this area.

Despite the strength of simple, quick quantification, it is dangerous to reduce HRA to the multiplication of assumed contextual factor values to find an HEP. Results can vary by orders of magnitude between methods and even between analysts [39]. While assigning PIF levels and multiplying their associated factors might seem straightforward, it is not so in practice. Analysts may choose to identify different HFEs, for reasons ranging from different task decomposition procedures, to different prioritization of the tasks in question, or even to different operational definitions of the HFE. Even if significant practical guidance is provided in other respects, this effect has been observed due to different analyst goals and scenario contexts in case study applications of the Petro-HRA method, which is based upon the SPAR-H method [40]. This large variation between the HFEs that are investigated leads to fundamentally different analyses, undermining any efforts to strive towards correctness in assessing PIF effects.

Compounding this issue is a lack of availability of professional HRA software tools in the public domain, unlike the myriad of open-source tools available to PRA practitioners [18]. Newer HRA methods tend to involve estimating conditional HEPs and evaluating uncertainties. These

are non-trivial tasks, especially for the required scope of most HRA [41]. It can be cumbersome to perform such calculations by hand, which is another reason why analysts might choose to employ checklist methods instead.

2.5.2 Expand basis of HFE structures via scientific studies

Another research opportunity arising from the current state of HRA methods is a lack of experimental and theoretical literature to substantiate the causal paths through which human failure occurs. Any modern method should be (1) rooted in data and models across multiple engineering disciplines and (2) able to incorporate multiple types of sources and information. Furthermore, it should be possible to update the method at regular intervals with new data, structures, or methodologies as research evolves. HRA is a relatively new science, incorporating aspects of cognitive psychology, organizational science, and systems engineering [6]. An ideal method would provide a framework for synthesizing a wide variety of content and updating its guidance based on the most recent advances in research. The accurate identification and characterization of HFEs is dependent on a method structure that is grounded in research.

The IDAC method [21] has the one of the most extensive literature-substantiated causal models of any method, with 29 information-preprocessing, 80 decision-making, and nine action execution human tendencies identified, each associated with one or two PIFs and substantiated with a literature citation. The IDAC literature survey is noted as being preliminary groundwork for subsequent refinement, but this refinement has not since been addressed. IDAC also does not denote each human cognitive tendency leading to higher HEPs as a mechanism for failure, which will be further expanded upon in the next section. Despite this extensive literature basis, there

is no guidance provided regarding the level at which to analyze HFEs. IDAC describes a dual method for identifying errors: *external*, referring to behaviors that do not reflect system/procedure needs, and *internal*, referring to cognition errors [19]. This error identification scheme may double-count errors, and may identify HFEs at mismatched task or objective levels.

IDHEAS [2] also has a robust literature basis, but still lacks complete traceability in the inclusion and exclusion of HFE components such as cognitive failure modes, mechanisms, and PIFs [14]. The Phoenix method [10] has a strong theoretical basis, and captures the probabilistic nature of cognition through Bayesian networks (BNs) but still falls short of capturing all the causal paths leading to failure, i.e., the interactions between PIFs across various layers of influence [14]. Otherwise, most HRA methods available to experts do not provide comprehensive, publicly available literature documentation for their causal models. This shortage of substantiation among methods results in a lack of traceable, reproducible results, and may be under- or over-emphasizing the impact of certain factors on the failure process and ultimate likelihood of HFEs.

2.5.3 Implement human failure mechanisms to better describe the failure process

Most HRA models consider only interactions between PIFs and task failure, even though it is clear that there are more failure mechanisms and modes at play behind the scenes [33]. For scientific accuracy and completeness, these mechanisms need to be enumerated and their effects documented so that the HFE can be identified as the culmination of the failure process, and not simply as an incorrect or omitted action [31]. Rather than considering the HFE as a merged block

of operator tasks, it is important to consider individual tasks and mechanisms so more targeted mitigation strategies can be used to refine system design [42, 43].

These mechanisms can be observed in describing how PIFs could cause a particular failure mode to arise. For example, an operator may misinterpret data due to a failure in *attention*, which in turn can be affected by stress, workload, or internal biases. Workload can also affect the operator's perceived severity of the scenario, which then causes *bias* toward incorrect interpretations of the data. It is clear that there are many causal pathways leading to the failure mode of misinterpretation stemming from different personal, environmental, and task-related causes, and that these factors interact with each other to increase or decrease the likelihood of failure [44]. Certain PIFs, particularly broad processes like *attention* and *bias*, tend to act as pinch points for the downstream effects of more indirect environmental factors such as task loads, perceived severity, and the operator's abilities and motivations, suggesting mechanistic effects for these particular PIFs [33]. By identifying these cognitive processes as human failure mechanisms, this may provide a rationale for the identification of more important PIFs to human performance [12].

HRA research should strive towards greater fidelity in psychological modeling to better reflect the causation inherent in HFEs. By considering human failure mechanisms, the HFE identification process can be rendered simpler and more modular. Furthermore, a complex, true-to-life model framework can be scaled down to a checklist to aid in the plant PRA process, but a less accurate model cannot be scaled up during implementation. As more human reliability data is collected, the modular nature of mechanisms allows the model to be continually augmented. It is also possible to derive causal HRA model structures from real-world data, which has the power to either justify or challenge current expert judgment or laboratory experiments [45].

Modeling scenario complexity presents a non-trivial research challenge. Higher model

fidelity can be achieved by acknowledging causal pathways beyond just the consequences of PIFs on overall human error probabilities [6]. In a recent model-based HRA method for nuclear power plant security, this is done through consideration of different PIF hierarchy layers, which form causal paths to the occurrence of CFMs, described by a Bayesian network [46]. However, this multi-layer structure does not fully exploit the reasoning capabilities of BNs, as interactions between PIFs are not part of the recognized causal paths [6].

Similarly, the IDAC method's causal model considers cognitive tendencies for each stage of cognition, analogous to mechanisms of failure [21]. Each tendency is described and associated with one or two PIFs from the set identified earlier in the method [34]. This collection of mechanisms is not treated as a complete set, and interactions between mechanisms are not considered (e.g., an operator's bias towards familiar information increases the likelihood that they will also lock into a strategy and ignore new information as it unfolds). Mechanisms are also not leveraged in the HEP quantification stage of the model [47]. Instead, the operator response is simulated at a much higher, behavioral level, while keeping track of several key PIF states. HFEs are not considered in IDAC's simulation sequence described in [47] either; the output describes operator decision-making strategy as the scenario progresses in time.

Other models, such as CMS-BN [48], have acknowledged these probabilistic aspects of cognition. CMS-BN uses BNs to directly model signals from the outside world, perception of these signals, and retrieval of beliefs from the operator's knowledge base to respond to these cues. Activation functions are used to determine success or failure at each stage of response, and stress and fatigue levels are updated at each time step. This bypasses the use of mechanisms, or an expanded set of PIFs to account for generic contextual influences. The majority of HFEs considered in this method are failures to correctly match mental models with reality during a

given task. Although this is a robust entity-based simulation of complex cognitive processes, the model does not yet take into account the full range of failure mechanisms and PIFs that act as causes of HFEs.

A complete, scientific treatment of human failure mechanisms means that HFEs can be more accurately characterized within this framework once they have been identified.

2.5.4 Expansion of HFEs to consider ex-control room scenarios, such as multi-unit accidents and external hazard conditions

Most PRA studies are restricted to single reactor units and thus referred to as single-unit PRAs (SUPRAs) [49]. However, the Fukushima Daiichi disaster in March 2011 highlighted the need to pay greater attention to inter-unit dependencies across two or more reactor units in a plant site. Proper characterization of these dependencies is critical to obtaining accurate risk profiles of a nuclear power plant site, and a recent review by Zhou, Modarres, and Droguett [50] detailed some of the methodological developments and challenges in multi-unit PRA studies (MUPRAs). One of these challenges was identified as the lack of methods for modeling human and organizational factors to support MUPRA studies. To address this challenge, a recent Canadian Nuclear Safety Commission workshop identified developing appropriate HRA method(s) for multi-unit context as an important and urgent priority [51]. St. Germain et al. [52] identified two sources for HFEs in multi-unit contexts that need to be modeled: (1) human actions that arise due to inter-unit dependencies, and (2) human actions necessary due to the additional challenges imposed by those accidents on operators. Examples of these challenges that may give rise to new HFEs include those due to a shared main control room (MCR), decision-making on deploying

shared portable equipment (e.g. emergency diesel generator), shared emergency response organization and personnel, and the effects of radiation or contamination damage from one unit on human actions performed at other units.

In addition, Park et al. [53] developed an approach to treat the human and organizational factors for a multi-unit HRA based on the SPAR-H method. The authors considered six multi-unit task types obtained from observed practice and a review of procedures in South Korean NPPs. Further work focused on developing dependency evaluation elements for multi-unit HFEs based on a review of single-unit HRA methods such as THERP, SPAR-H, Fire-HRA, and K-HRA [54]. Additional research activities have recognized the inadequacy of HRA methods in addressing human and organizational factors for multi-unit scenarios for severe accident management. The review of challenges associated with MUPRAs [50] suggested leveraging and extending current research on the reliability of operator actions following external events such as seismic event PRAs, FLEX and portable equipment, and severe accident management guidelines (SAMG) tasks.

Another opportunity for enhancing the research basis of HFE identification during external hazard conditions is in the treatment of actions preceding the hazardous event. Few task analyses for preparatory mitigation actions are available, and nominal timing estimates are nonexistent. These mitigating actions may be subject to increased likelihood of error due to the additional time needed to compensate for adverse conditions and additional difficulties imposed by these scenarios and previous errors. This may result in compounding error conditions that are not typically addressed in HRA, but could be modeled dynamically throughout the event.

The analysis of ex-control room actions during site-damage conditions presents additional unique challenges. Data collection is difficult due to the higher variability of operator actions,

the higher effort levels required to conduct observations, and the validity issues with simulator studies [20].

2.5.5 Improve identification and treatment of errors of commission

Another issue that has not yet reached a definitive conclusion is the approach for identifying HFEs stemming from EOC. It is a much more direct task to imagine the points at which human actions are *not* taken rather than explicitly listing the multitudes of incorrect actions that could be taken [55], so analyses typically only consider errors of omission (EOO). It is also particularly difficult to identify scenario-initiating human actions, because most prior research has examined pre- and post-initiator human actions [56]. Even for post-initiator events, most current HRA methods do not explicitly identify EOCs; rather, the EOC effect is implicitly included in the HEP [43, 57]. There is a present need for a standardized method that supports the identification of EOC HFEs, as it has been found from operating experience records in Korea that over 90% of events caused by human error since 1991 were commission errors [56, 58].

Various EOC search schemes have been developed to address this. The Borssele-global misdisagnosis method [59] begins by assuming that an adverse scenario has occurred, and then searches the response procedure for potential erroneous actions. Other methods, like ATHEANA [29], also assume that a scenario has occurred but instead examine the necessary system functions and how operators could cause them to fail. Finally, methods such as Commission Errors Search and Assessment (CESA) [55] list all actions, both those proceduralized and not, then identify events linked to important systems and the most critical scenarios linked to the potential EOCs. There is not yet a consensus on which technique is best. Generally, approaches like CESA are

used for NPP HRA because of the procedure-dominated nature of the environment [56].

A shortcoming of relying on procedure analysis is that it excludes non-proceduralized actions. EOCs are frequently associated with maintenance and abnormal operations [58] which may not have complete procedures, if procedures are available at all. Ex-control room actions are another particular concern, because they are integral to the mitigation of external events, but are often assumed to be skill-of-the-craft. As procedures for these actions are rarely available, it becomes necessary to conduct task analyses through expert group consensus building [60]. During external hazard events, the operators are almost certainly affected by stress, higher than usual task loads, and poor tool and information availability. These factors may combine with other aspects of scenario context to bring about HFEs. EOCs are highly likely to be brought about under these conditions, yet no standard technique exists to identify them.

One potential way to circumvent this difficulty in identifying EOCs alongside potential EOOs might be consideration of all the ways that psychological functions and operator actions could fail. This bottom-up approach provides a better framework for identifying EOCs than a top-down search scheme that prioritizes system functions [61]. The bottom-up, human factors approach was found to be suitable for identifying HFEs for ex-control room actions during external hazard events [22]. Task analyses of flooding hazard mitigation actions were broken down into failures of MCFs, which were broken further into relevant CFMs, enumerating the failures that could potentially be observed. Frameworks such as the Information-Decision-Action in Crew Context method, alongside the implementation of human failure mechanisms, enable a more holistic analysis of human failures than the dichotomy of errors of omission/commission.

One method that employs this approach is the GTT-PIF structure developed for the radiotherapy domain [62]. It uses macrocognitive functions and proximate causes of failure (analogous

to CFMs) to map generic tasks associated with radiotherapy administration to Groth's PIF taxonomy [24]. However, this method introduces some additional complexity in the form of identifying relatively specific radiotherapy-related tasks, and the step of mapping each potential action to a generic task type. The PIFs identified for each task type are not necessarily relevant to similar task types in other domains, necessitating further work if this method were to be generalized to non-radiotherapy applications.

Finally, other error taxonomies like SHERPA and TRACEr [63, 64] support EOC identification via task analysis, although they are more developed with respect to HFEs involving information-gathering and action-taking tasks. The IDHEAS method [17] can be used as a more complete catalog of cognitive errors, considering a wider variety of decision-making failure modes [65]. The Phoenix method has a similarly strong catalogue of potential CFMs for each macrocognitive phase of operation [10]. However, there still remains a lack of standardization surrounding the identification of EOC HFEs through this identification of macrocognitive failure points.

2.6 Operationalizing Identified Gaps into Dissertation Gaps

To address these conceptual gaps in HRA, they were operationalized into the three dissertation research gaps detailed in Chapter 1. The most important challenge to tackle was developing models with a causally robust theoretical basis (Gap 1). This significant research opportunity comprises the first three conceptual gaps: incorporating robust HFE characterization into HRA methods, expanding the theoretical and scientific basis of the models, and implementing human failure mechanisms. After developing models with these characteristics, the next logical step

was implementing a user-friendly, data-driven quantification scheme (Gap 2). This corresponds to part of the first conceptual gap: incorporating a usable, accurate quantification scheme into HRA methods. Lastly, the models' applicability were validated for a variety of scenarios (Gap 3). We chose to conduct a demonstration exercise with an ex-control room task analysis under external flooding hazard conditions due to its complexity beyond the typical control room scenarios where HRA is typically applied. This corresponds to the last two conceptual gaps: the treatment of HFEs in ex-control room scenarios and the identification and treatment of errors of commission.

2.7 Bayesian Networks for HRA

To address these gaps in characterizing HFEs, Bayesian networks are a versatile and intuitive modeling tool to be leveraged. Bayesian network (BN) methods allow for the integration of disparate data sources, which is important in a data-sparse field like HRA, and encode probabilistic causal principles, which is important in modeling stochastic processes like human cognition. This section presents an overview of BN formalism as well as previous applications of BNs to the HRA field.

2.7.1 Overview of Bayesian Methods

Bayesian networks are directed acyclic graphs that consist of nodes representing variables and arcs denoting causation. Nodes may represent the occurrence of an event or evidence, and take on different states to reflect these real processes. Familial terminology is used here to describe the causal relationships between nodes, following conventions used in [66]. In the example

network shown below, there are four nodes and three arcs, connected via the following relationships:

- Nodes A and B are the *parent* nodes of C, and C is the parent of D. A and B can be said to cause C, and C causes D. Mathematically, $pa(C)$ is used to describe the set of nodes that are parents to the node C, which all have outgoing arcs directed into C.

- Likewise, C is the *child* node of A and B, and D is the child of C. This means that C is a direct effect of A and B, and D a direct effect of C.

- Nodes A, B, and C are the *ancestors* of node D. Ancestor nodes comprise the parents of the parent nodes, and so on until a *root node* is reached that has no parent nodes. Ancestors beyond the direct parents of the node are its indirect causes: A and B are the indirect causes of D.

- Nodes C and D are the *descendants* of node A, and similarly C and D are also the descendants of B. Descendants beyond the direct children of the node are its indirect effects: D is the indirect effect of A and B.

These terms are useful for keeping track of a node's direct and indirect causes. One main benefit of BNs is the ability to consider as many parent or child nodes as needed to sufficiently reflect real-world causes and effects.

Beyond the causal logic encoded in graphical structures, Bayesian networks are also a powerful quantitative reasoning tool. Bayes' theorem (Equation 2.1) allows for the calculation of conditional probabilities given prior belief in the likelihood of a parent node, or cause, *H* ($Pr(H)$), and the likelihood that its effects *e* will materialize given that the cause is present

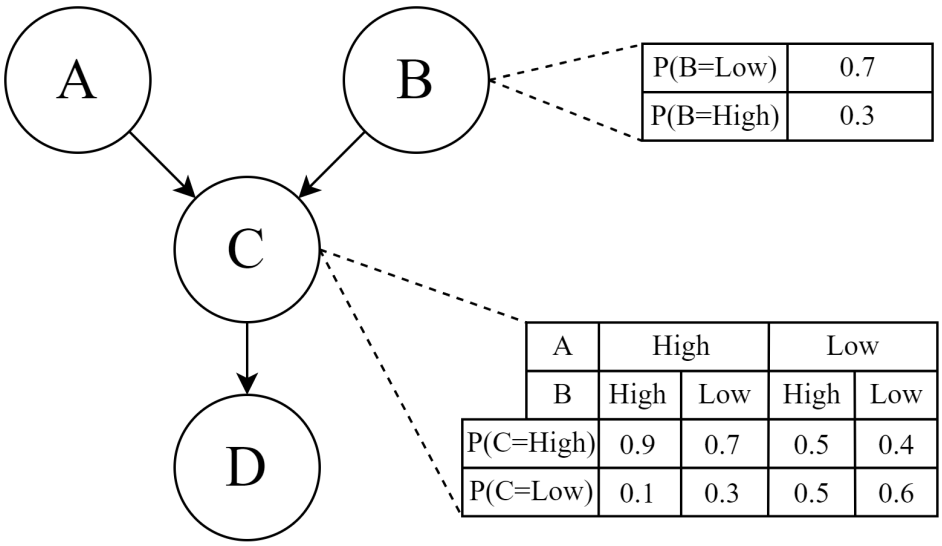

Figure 2.3: Example Bayesian network structure.

($Pr(e|H)$) [66]. The expression is then normalized by the overall probability that the effects are seen ($Pr(e)$).

$$Pr(H|e) = \frac{Pr(e|H)Pr(H)}{Pr(e)} \tag{2.1}$$

This concept can be extended to find the joint probability distribution of a node given its parent nodes. For a set of nodes $A_1, ..., A_n$, Equation 2.2 describes the joint distribution of *n* variables. This formula assumes conditional independence of X_j from all other variables besides its parents $pa(X_j)$.

$$P(A_1, ..., A_n) = \prod_j Pr(X_j|pa(X_j)) \tag{2.2}$$

To parameterize the network, marginal probabilities are assigned to the root nodes, while the conditional probabilities of the descendant nodes are assessed conditioned on the states of

their parent nodes according to Equation 2.2. These probabilities are stored in conditional probability tables (CPTs), which are typically assessed using data or expert judgment. Tables showing example marginal probabilities for node B and example conditional probabilities for node C are shown in Figure 2.3. The probability of node C taking on a given state is conditional on both the state of node A and node B. BNs are capable of incorporating a variety of data sources, and are easy to update when additional data is obtained. Furthermore, through setting evidence on a node, BNs allow for both inferential and deductive reasoning about the other node states in the network.

2.7.2 Application to Human Reliability

Bayesian networks are a good tool for developing HRA models because they provide a robust framework to address a number of shortcomings in existing HRA methods [67]. These shortcomings include oversimplifications of cognitive processes, lack of technical basis, and lack of causality. Furthermore, BNs allow for the expansion and adjustment of HRA models with the availability of more data.

A comprehensive review categorized previous applications of BNs to HRA into five main groups: modeling management and organizational factors, relationships between PIFs, adapting existing HRA methods to BN reasoning, assessing dependency between HFEs, and modeling operators' situational assessment [68]. All of the methods reviewed highlight different strengths of the BN method that are relevant to HRA, including the ability to incorporate diverse sources, reason about causes even with limited data, and model complex, multi-layer relationships [68]. Newer BN applications such as [48] or [46] as well as [6, 14, 45] highlight the ability to proba-

bilistically model cognitive processes or the influence of contextual factors.

While the application of BNs in HRA has been increasing in recent years, a number of issues have been found with their current use [68]. First, the studies surveyed in [68] were found to have an over-reliance on expert judgment to develop the BN structures. This resulted in limited emphasis on building traceable and credible models [69] due to the lack of documentation on how the expert interviews and questionnaires were structured. Second, the BNs are presented as complete works, without context and insights into the reasoning behind the nodes included as well as the causal connections between them, which is foundational to the representational power of the BN [70]. Third, large variations exist between analysts in the definitions of BN nodes' states.

In order to address some of the limitations in HRA modeling, a method for simplifying qualitative BN models was developed by Zwirglmaier, Straub, and Groth [14]. Through this process, dependencies between the CFMs (from the IDHEAS [2] method) and their causal factors were identified. A similar method is employed in this work to capture causal psychological paths leading to human failure events. In another work, Groth, Smith and Moradi developed a hybrid algorithm fusing data from multiple sources, existing HRA models, and cognitive literature [6]. Both works serve to enhance the traceability and scientific basis of HRA methods and provide a foundation for the current research.

Chapter 3: Causal Pathways Leading to Human Failure Events in Information-Gathering System Response Activities

This chapter is reproduced from [44], [25], and [71].

Camille S. Levine, Ahmad Al-Douri, and Katrina M. Groth. "Causal Pathways Leading to Human Failure Events in Information-Gathering System Response Activities." In: *13th Nuclear Plant Instrumentation, Control & Human-Machine Interface Technologies (NPIC&HMIT 2023)* (July 15–20, 2023). Knoxville, TN.

Camille S. Levine, Ahmad Al-Douri, Vincent Philip Paglioni, Michelle Bensi, and Katrina M. Groth. "Identifying human failure events for human reliability analysis: A review of gaps and research opportunities." In: *Reliability Engineering & System Safety* 245 (May 2024), p. 109967. ISSN: 0951-8320. DOI: 10.1016/j.ress.2024.109967.

[Submitted, in press]: Camille S. Levine and Katrina M. Groth. "Three Novel Causal Logic-Informed Models for Identifying and Quantifying Operator Errors, Part I: Information-Gathering." In: *Reliability Engineering & System Safety* (2025).

3.1 Introduction: Targeting Causal Pathways Leading to HFEs

One of the main goals of HRA is the identification and characterization of potential HFEs. HFEs arise from complex chains of causally related events, and represent the culmination of a human-machine team's failure to complete an objective [27, 31]. This objective is comprised of multiple high-level cognitive or physical actions, called major crew functions (MCFs) [31]. MCFs are system-specific instantiations of macrocognitive functions, such as information gathering, decision-making, or action execution. This chapter will examine the pathways by which information gathering failures occur.

A proportionally large share of accidents across various industries can be attributed to human error or contexts conducive to human error [16]. HRA is a useful tool for characterizing this human error in a wide variety of domains so that it can be acknowledged and mitigated. There is a need for cognitively realistic, yet quantifiable, models that account for the full range of performance contexts inherent to system operation [6–9]. While older HRA models frequently lack this cognitive realism, newer models may lack quantification schemes that are feasible in practice. Current research has sought to enhance this theoretical basis [2, 13, 14], but is yet incomplete in its lack of explicit acknowledgment of human error's causal chains.

Most HRA methods directly compute human error probabilities (HEPs) from multipliers derived from combinations of performance influencing factors (PIFs) relating to the event context. However, it is clear that there is more psychological depth to the HFE than current methods and models represent [32]. There is a present need for a cognitively based HRA method with a topography based in both organizational and psychological literature, and state-of-the-art reliability engineering methods [6]. These research opportunities include improved HFE characterization,

enhanced theoretical basis, and need for considering failure mechanisms [25]. It is important to fully describe all these pathways to failure so that mitigation strategies can be applied in a targeted manner. By substantiating each pathway through the manners by which PIFs exert their influence, model outputs will be rendered traceable back to theory. The remainder of this chapter documents the structure, quantitative capabilities, and validation of a human performance model for information-gathering that seeks to address these opportunities in the HRA field.

3.2 Method: Model Structure Development

3.2.1 Causal Factors

As shown in Figure 3.1, there are multiple contextual factors that tell the story of an HFE through a causal chain. This hierarchy of contextual factors underpins the development of the three causal BN models presented in this chapter and in Chapters 4 and 5. By labeling HRA variables within a multi-layer structure, distinctions can be made between the kinds of influence exerted by each variable type. Bayesian network model structures are also ideal because quantification techniques are compatible across and between each layer of the model.

In recent literature, the HFE has been defined as the result of a failed function-level HRA variable and the culmination of the human failure process [27]. Macrocognitive functions, such as information-gathering, decision-making, or action execution, have been defined as a higher-level cognitive process as outlined by the Information-Decision-Action (IDA) framework [19]. The major crew function (MCF) is a high-level action taken by a human-machine team and is a system-specific instantiation of a macrocognitive function [6, 31]. This could be tasks such as interpreting an alarm, deciding on a procedure to follow, or actually carrying out the procedure.

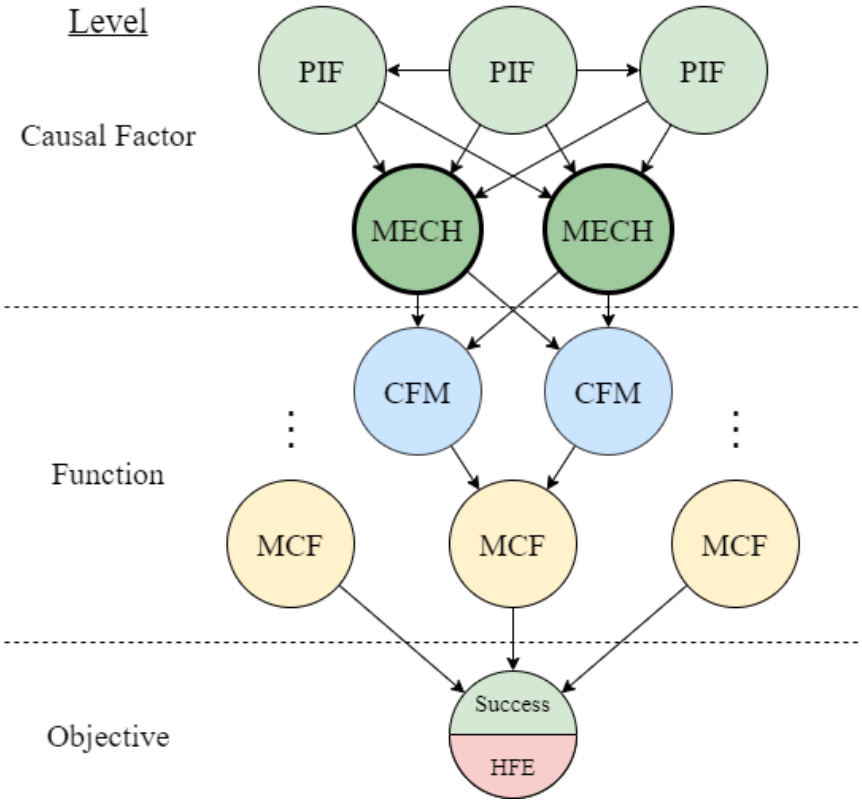

Figure 3.1: Hierarchy of HFE variables, shown in a Bayesian network framework.

Failure of an MCF can occur through various different crew failure modes (CFMs). For example, someone might fail to gain situational awareness of a system parameter because they misinterpreted data, or because they were looking at the incorrect data source. These errors are examples of observable modes through which failure can occur.

Then, PIFs serve as the context during which the error occurs. PIFs may be categorized into personal, team, machine, situational, and organizational factors as shown in the taxonomy developed by Groth & Mosleh [24] in Table 3.1. Some PIFs exert a more direct influence on CFMs and represent the mechanism of a cognitive or physical causal pathway, similar to failure mechanisms of mechanical systems and components. There is a trade-off involved between constructing complete, psychologically substantiated models, and the ease of conducting an HRA. There are

infinitely many cognitive pathways to a functional-level failure, as there are infinitely many responses to a situation. However, only the most relevant and potentially risk-inducing pathways should be modeled, and a manageable subset of the most commonly documented mechanisms of failure should be used. The HRA construct of mechanisms and their usage are described in more detail in [25].

3.2.2 Synthesis of Causal Modeling Literature and Methods

Alongside this HRA variable hierarchy, a method for creating qualitative BN models developed by Zwirglmaier, Straub, and Groth was implemented in this work [14]. This process identifies dependencies between the Phoenix HRA method's CFMs (using decision trees from the IDHEAS [2] method) and relevant causal factors based on cognitive psychology. For scientific accuracy, the model presented in [14] includes nodes that may be too abstract to collect data on, but remain integral to model completeness. In another work, Groth, Smith and Moradi developed a hybrid algorithm to fuse data from multiple sources, existing HRA models, and cognitive literature [6]. Both works serve to enhance the traceability and scientific basis of HRA methods and provide a foundation for the current research.

We defined an external flooding hazard scenario, examined in detail in Chapter 6, using the NUREG/CR-7256 report titled "Effects of Environmental Conditions on Manual Actions for Flood Protection and Mitigation" [60]. The report provides an approach to decompose manual actions into tasks, subtasks, and specific actions, and further enumerates performance demands upon the operators. Specifically, Section 6.3 of the report provides three task decomposition examples that were achieved through group discussion and consensus building by a research

team representing a wide array of expertise. We used these task decompositions to identify human failure events (HFEs) that may occur in external actions aimed at mitigating nuclear power plant flooding hazards and developed event tree and fault tree models to determine relevant CFMs to each task.

Building upon those models, our method in this work was to use the definitions for the nine information gathering CFMs from the Phoenix framework documentation [23] to determine the possible mechanisms by which a failure may occur. Phoenix enumerates a set of CFMs, shown in Table 3.2, which are intended to exhaustively describe all the potential ways failure could occur. The CFM descriptions included explanations to how each CFM might be brought about, which allowed for the immediate identification of relevant PIFs.

Then, we utilized Groth and Mosleh's PIF taxonomy [24] to identify further applicable PIFs, shown in Table 3.1. The taxonomy is organized into five main categories: machine-based, person-based, team-based, organization-based, and situation/stressor-based, with each containing a multi-level hierarchy of PIFs. The PIFs and their organization are particularly appropriate for this work as it provides a mechanism for integrating information from multiple sources. This division scheme allows the PIFs to be defined with respect to the appropriate category of the socio-technical system. The taxonomy's organization and PIF descriptions further enable the identification of many interrelationships between PIFs. In turn, this allows the analyst to identify the root cause of the human error and supports strengthening the causal basis of HRA.

We iteratively examined each PIF category for each CFM to determine the category's relevance. If it contained pertinent causes of the CFM, the PIFs contained within the category were considered for inclusion into the model. Certain PIFs were denoted as direct causes of CFMs; for example, poor human-system interface (HSI) directly causes CFM I6 (Reading Error). Other PIFs

Organization-based	Team-based	Person-based	Situation/stressor-based	Machine-based
Training program	Communication	Attention	External environment	HSI
Availability	Availability	To task	Conditioning events	Input
Quality	Quality	To surroundings	Task load	Output
Corrective action program	Direct supervision	Physical & psychological abilities	Time load	System response
Availability	Leadership	Alertness	Other loads	
Quality	Team coordination	Fatigue	Non-task	
Other programs	Team cohesion	Impairment	Passive information	
Availability	Role awareness	Sensory limits	Task complexity	
Quality		Physical attributes	Cognitive	
Safety culture		Other	Execution	
Management activities		Knowledge/experience	Stress	
Staffing		Skills	Perceived situation	
Scheduling		Bias	Severity	
Workplace adequacy		Familiarity with situation	Urgency	
Resources		Morale/motivation/attitude	Perceived decision	
Procedures			Responsibility	
Availability			Impact	
Quality			Personal	
Tools			Plant	
Availability			Society	
Quality				
Necessary information				
Availability				
Quality				

Table 3.1: Performance Influencing Factors taxonomy from [24].

Table 4-5
Set of CFMs.

ID	Crew Failure Modes in "I" phase	ID	Crew Failure Modes in "D" phase	ID	Crew Failure Modes in "A" phase
I1	Key alarm not responded to (intentional and unintentional)	**D1**	Plant/system state misdiagnosed	**A1**	Incorrect timing of action
I2	Data not obtained (intentional)	**D2**	Procedure misinterpreted	**A2**	Incorrect operation of component/object
I3	Data discounted	**D3**	Failure to adapt procedures to the situation	**A3**	Action on wrong component/object
I4	Decision to stop gathering data	**D4**	Procedure step omitted (intentional)		
I5	Data incorrectly processed	**D5**	Inappropriate transfer to a different procedure		
I6	Reading error	**D6**	Decision to delay action		
I7	Information miscommunicated	**D7**	Inappropriate strategy chosen		
I8	Wrong data source attended to				
I9	Data not checked with appropriate frequency				

Table 3.2: Crew Failure Modes (CFMs) proposed by the Phoenix HRA Framework [10].

exerted more indirect influences. For example, poor training availability may cause knowledge gaps that impact coordination, causing CFM I7, (Information Miscommunicated), but miscommunication errors are not a direct effect of poor training.

Next, we consulted a set of draft Bayesian networks developed by Groth & Hendrickson [72]. The drafted causal structure for information-gathering failures documents causes of seven of the nine I-phase CFMs, with arcs illustrating some causal relationships among the PIFs. The draft BNs did not adhere to a strict set of PIFs, but they did identify a layer of cognitive psychological mechanisms by which the PIFs may potentially influence human error. These mechanisms, such as "AttentionDistraction" or "ExpectationBias" represented the proximate causes of error. This construct has been modified in our models to render the variable more quantifiable. A fuller

discussion of mechanisms in this context is available in [33].

The next step was to substantiate these relationships through cognitive literature to solidify the causal basis of these BN models. The literature review contained in NUREG-2114, "Cognitive Basis for Human Reliability Analysis" [13, 32] comprises a strong body of psychological and human factors literature which was used to substantiate the BN models. Additional literature related to the works in NUREG-2114 were also consulted. The references, shown in Table 3.3, were used to substantiate each directed arc, justifying all of the choices made during model development. Some sources, upon further inspection, cited studies that justified the addition or removal of an arc from the BN. The model structures were refined as necessary to more closely align with the psychological research. Finally, Paglioni's HRA dependency idioms [35] were employed to document the nature of the causal relationships between nodes. These idioms were also applied during the quantification process, described later in 3.4. The synthesis of these sources is summarized in Figure 3.2.

Literature concerning the psychological and organizational factors leading to error was used to develop the theoretical basis of this work, building upon the literature review detailed in [32]. In particular, Endsley's work on situational awareness described the effects of many person-based PIFs, such as stress, human-system interface, and loads, as well as several error mechanisms, including bias and attention [73]. Endsley poses a general model of situational awareness and reviews sources documenting the pathways by which information synthesis errors may occur. Letsky et al.'s macrocognitive model of team collaboration was instrumental in developing the relevant organizational- and team-based factors [74]. The complete list of literature used in the substantiation of this Bayesian network is documented in Table 3.3.

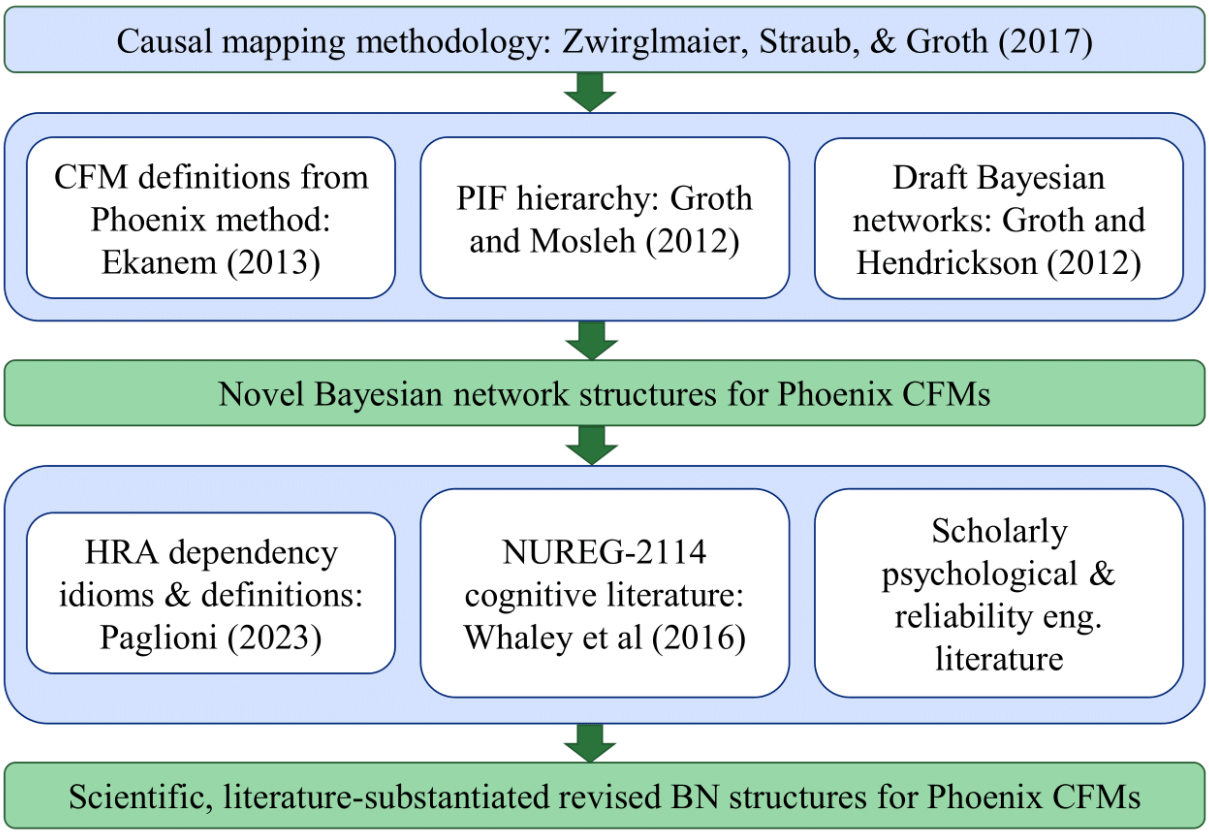

Figure 3.2: Sources used in the development of human performance Bayesian networks.

3.3 Result: Model Structure

3.3.1 Model Architecture

The full Bayesian network structure is shown in Figure 3.3. Each arc in the structure is substantiated in existing psychological literature or dependency idioms as described here, and the source is indicated on the arc. Some of the structures are logically derived idioms from [35], such as the nature of the Time Load, Task Load, and Non-Task Load construction comprising All Loads, or the state of Task Complexity defining the state of Task Load. Some PIFs can be observed to deterministically affect their child nodes; for example, with inadequate time available,

Ref on BN	Ref Number	Citation	Area on BN
a)	[75]	Broadbent, D. (1958).	Sensory limits
b)	[73]	Endsley, M. R. (1995).	Person-based PIFs, attention, bias, knowledge, training
c)	[76]	Eriksen, C.W., & St. James, J.D. (1986).	Sensory limits, attention
d)	[77]	Jones, D. G., & Endsley, M. R. (1996)	Loads
e)	[78]	Klein, G. A. (1993).	Person-based PIFs for misinterpretation
f)	[79]	Klein, G., & Moon, B. (2006).	Discounting of data
g)	[74]	Letsky, M. P., Warner, N. W., Fiore, S. M., & Smith, C. A. P. (2008).	Collaboration and organization based PIFs
h)	[80]	Lipshitz, R. (1993).	Training, perceived impact
i)	[81]	Orasanu, J., & Martin, L. (1998).	Bias, knowledge, training, goal prioritization
j)	[82]	Roth, E.M. (1997).	Safety culture, knowledge and training for misinterpretation
k)	[83]	Wickens, C. D., Lee, J. D., Liu, Y. & Becker, S. E. G. (2004).	Stress, fatigue, knowledge, training, perceived impact
l)	[35]	Paglioni, V. P. & Groth, K. M. (2024).	Deterministic dependencies

Table 3.3: References used to substantiate Information-Gathering network model arcs.

loads are certain to be very high. This is an instance of the Definitional dependency idiom, which is present throughout the model. For these kinds of structures, no other literature substantiations were needed due to the logical nature of these particular constructs.

During the development of these models, clusters of PIFs relating to different cognitive mechanisms repeatedly emerged. Like the cognitive proximate causes introduced in the draft BNs [72] or the causal pathways identified in Zwirglmaier's method [14], these clusters each represented the influence of a different psychological or physical mechanism [33]. Similar to how mechanical failure mechanisms describe mechanical failure modes, these variables describe *how* crew failure modes are brought about. These human failure mechanisms may or may not be ob-

servable. For example, Coordination is a mechanism that may be affected by several team-based and person-based PIFs, and in turn can give rise to several CFMs. Although Coordination may not be directly observable or quantifiable in all situations, it is still a psychological mechanism that influences the likelihood of various failure modes, such as miscommunication or data being checked with an inappropriate frequency. A fuller discussion of human failure mechanisms and guidelines for their identification can be found in [33].

These mechanisms are not deterministic in their causation: they are probabilistic, increasing the probability of occurrence of a CFM. Thus, multiple clusters can contribute to a particular CFM, with differing degrees of influence on the causation of that CFM. For example, operators may be biased against checking a particular data source, or they may simply fail to give it the necessary continuous attention. These causes are both heavily error-inducing, but to differing degrees. Both mechanisms may give rise to the CFM I9 (Data Not Checked With Appropriate Frequency). Although neither mechanism is absolutely necessary to cause the error, either would provide a sufficient cause.

3.3.1.1 Model Architecture: Crew Failure Modes

In the first causal layer of the model, the information-gathering Phoenix CFMs are shown in light blue. These CFMs all define the state of a failed Information-gathering MCF, shown at the bottom in light yellow. The MCF cannot be in a failed state without the presence of at least one CFM, and if one of the CFMs occurs, the MCF is failed. This is deterministically defined as such. The CFMs are brought about by the presence of human failure mechanisms, described in more detail in the next section. There is significant overlap between the direct and indirect

causes of each CFM due to underlying similarities in errors that may arise during the cognitive processes of information-gathering. CFMs may also be brought about by individual PIFs that are not mechanisms, such as HSI or external environmental conditions. A similar layer of CFMs leading into MCF failure also exists for the decision-making and action-taking models.

3.3.1.2 Model Architecture: Human Failure Mechanisms

In the second layer of the model, human failure mechanisms are shown in darker green. There are five key causal clusters of PIFs readily apparent in the structure: *Prioritization*, *Bias*, *Attention*, *Procedure Error*, and team *Coordination*. Each corresponds to a mechanistic PIF.

The first mechanism cluster, *prioritization*, directly causes two CFMs (I1 and I3). Prioritization errors are directly caused by three PIFs (perceived situation severity, urgency, & impact, safety culture, and morale, motivation & attitude (MMA)). These errors are also indirectly caused by loads (primarily task, non-task, and time loads) which can be compounded by three other PIFs: HSI, task complexity, and staffing. Prioritization can also be caused by the presence of bias, which is identified as another mechanistic PIF.

The second mechanism cluster, *bias*, directly causes five CFMs (I1, I2, I3, I4 and I9). Bias errors were found to be directly caused by two PIFs (perceived situation severity, urgency, & impact, and training/knowledge/experience/familiarity). These errors are also indirectly caused by high loads through perceived situation urgency. In addition to prioritization errors, bias is a mechanism that may cause attention errors.

The third mechanism cluster, *attention*, directly causes eight CFMs (all except I2), the most among the five clusters. Attention errors were found to be directly caused by four PIFs: sensory

limits, stress, loads, and alertness and fatigue. These errors are also indirectly caused by task complexity (through exceeding of sensory limits) and loads (through high stress), as well as the causes of loads enumerated above. The mechanism of Attention was found to be connected to the most PIFs and CFMs out of all of the mechanistic PIFs, indicating that it may be one of the most common mechanisms for information-gathering failures. This type of attention is also known in the literature as "selective attention," or one's ability to select relevant cues from the environment and detect information.

The fourth mechanism cluster, *procedural error*, directly causes three CFMs (I3, I5, and I8). Procedural errors were found to be directly caused by four PIFs: procedure availability and quality, training/knowledge/experience/familiarity, loads, and safety culture. This mechanism reflects a misapplication of the procedure to the situation.

The fifth mechanism cluster, *team coordination efficacy*, directly causes three CFMs (I1, I7 and I9). These errors were found to be directly caused by four PIFs: staffing, training/knowledge/ experience/familiarity, communication availability and quality, and team cohesion. Team coordination efficacy errors are also indirectly caused by the following PIFs: safety culture (through staffing); and training/knowledge/experience/familiarity, staffing, and supervision (through cohesion).

3.3.1.3 Model Architecture: Performance Influencing Factors

In the third and uppermost layer of the model, an interconnected layer of PIFs (shown in lighter green) capture the context leading up to the mechanisms and CFMs. In addition to their direct causes, all of the mechanism clusters are also affected by influences that propagate in from

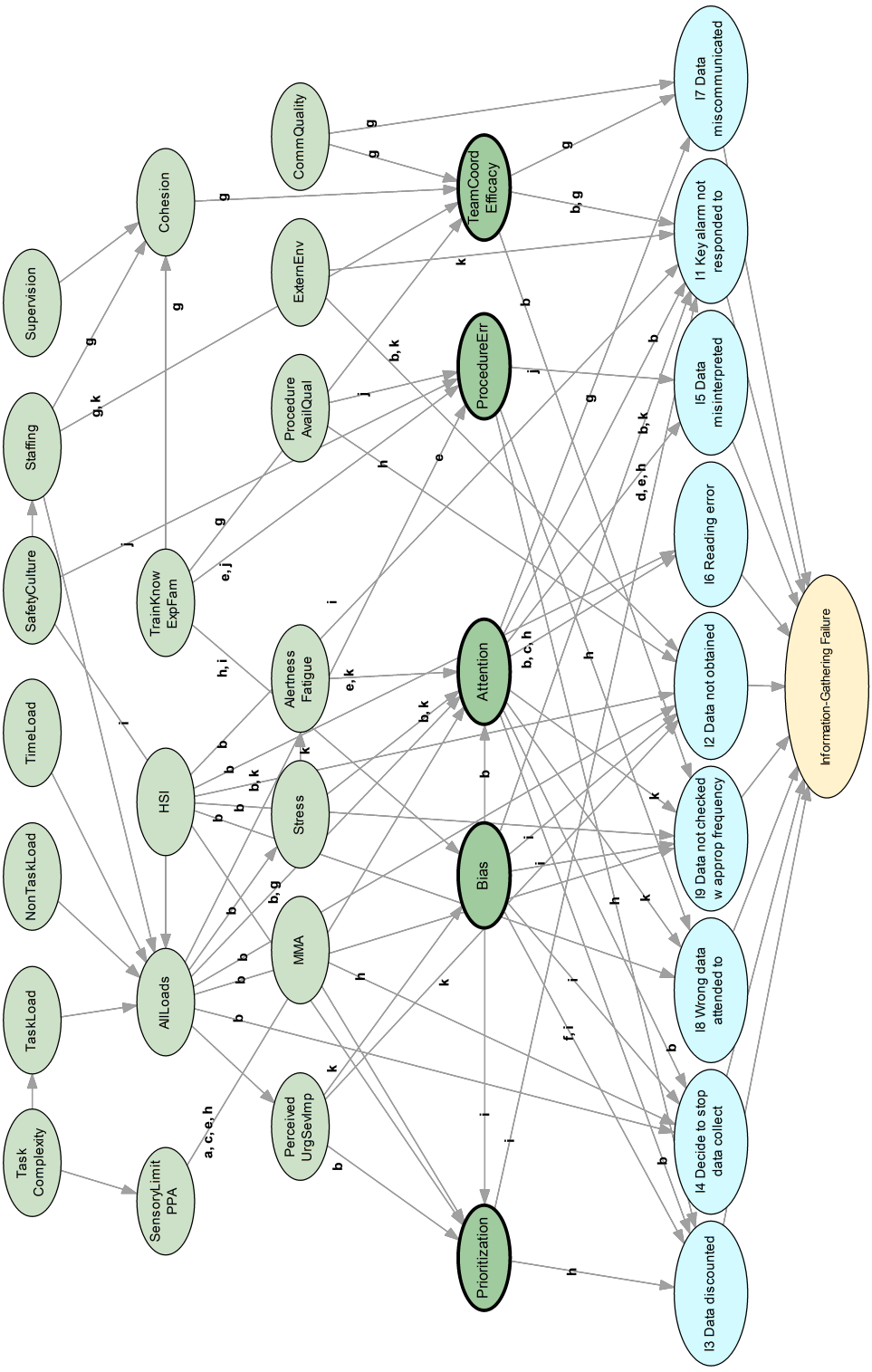

Figure 3.3: Full cognitive literature-substantiated Bayesian network structure for the Phoenix method's I-phase CFMs.

further-out PIFs. For example, staffing can be considered as one of these higher-level influences because it affects the overall load (All Loads), which may in turn induce stress on the operator.

Several notable PIF structures emerged due to the interrelated nature of the factors. The combination of high task load, non-task load, time load, and task complexity, along with the presence of degraded staffing and HSI will generally add up to overloading. This is represented through the structure of "All Loads" and its parent PIFs. Overloading may then also cause stress to the operator, which in turn wears them down, causing fatigue.

Another emergent PIF structure represents factors that lead to the mechanism of coordination issues. A safety culture that does not prioritize the system's workforce may lead to staffing issues, in turn affecting cohesion along with supervision quality and individuals' awareness of team roles. Staffing issues may also directly result in poor coordination due to an insufficient number of team members. In addition to poor team cohesion and staffing choices, coordination may also be affected by the quality of communication between team members as well as their individual knowledge and experience levels.

Finally, the model includes a PIF comprising the highly related factors of training, knowledge/experience, and familiarity. A suitable training program should give the operator the necessary knowledge needed to handle various scenarios, and repeated training will result in high familiarity. Likewise, a knowledgeable and experienced operator will likely have high familiarity with a variety of situations due to their rich knowledge base. The effects of each facet of an operator's knowledge base are not easily separable in the data and causal literature, therefore this construct was represented by a single PIF of training/knowledge/experience/familiarity.

The resulting model inherently captures a wide range of additional indirect factors that meaningfully influence multiple CFMs resulting in structures with a rigorous, complete causal ba-

sis. This approach captures the dependencies between individual and mechanistic PIFs, whereas other existing models do not consider the interactions between the different HRA variables.

3.4 Method and Data: Model Parameterization

One of the longest-standing gaps in HRA practice is the trade-off analysts must make between having a robust causal basis and a robust quantitative basis while selecting a method [25]. In HRA methods with a strong psychological focus, there tends to be little to no guidance for quantitative analysis, or the quantification scheme is difficult to implement. In practice, this means that such methods are not used in favor of methods with more straightforward HEP quantification [18]. However, HRA methods that are easy to quantify tend to be lacking in their scientific and causal basis. The HEPs provided or calculated by these methods are typically derived from expert judgment or aggregated human error data that ignores scenario context. For methods that assign PIF multipliers to a base HEP for a limited palette of PIFs, the assumption that contextual factors can be fully modeled through this treatment is non-trivial [84]. This method of quantification ignores the interactions between PIFs as well as the multitude of ways that PIFs may exert causal influence on human performance.

It is important that HRA models be based in real psychological and physical processes so that the analysis results reflect operational reality. Model structures can be developed through cognitive literature (e.g. [2, 44]) or through data-driven methods (e.g. [45]) to embody real-world observations. Parameterization augments the previously developed literature-based structures by adding quantification capabilities to the models. This process involved leveraging several existing HRA data sources and extracting data from the psychological literature to find conditional HEPs

under each given error context.

3.4.1 Data Sources

Human reliability is a particularly data-sparse field. Collecting accurate, generalizable, and detailed human reliability data is time-consuming and costly, so there are not many data sources available with all of these attributes. With HEPs on the order of 10^{-2} for each MCF, a strong class imbalance is present in the data, so it becomes necessary to collect around 10^4 data points or more to have a complete dataset [85]. Difficulties in using HRA data also tend to arise due to the lack of standardization of HFE definitions, the level of specificity, and the values that variables are allowed to take on (e.g., the available PIF levels). This necessitates mapping variables between methods, which introduces model uncertainty.

Over the years, many data resources have been developed to support the ever-increasing quantification capabilities of HRA methods. Earlier efforts include CORE-DATA, which synthesized simulator and incident data from 11 different industries [86, 87], Operator Reliability Experiments (ORE), which collected time-reliability data from simulator experiments [88], and Human Event Repository and Analysis (HERA), which collected detailed narratives and contextual attributes from nuclear power plant (NPP) event reports and simulator studies [89]. NUREG/CR-6949, a feasibility study on empirical data and Bayesian methods in HRA, builds upon the HERA database by expanding into other mathematical techniques for HRA quantification [84]. NUREG/CR-6949 also provides discrete expert-elicited Bayesian prior estimates for several PIFs and PIF states, which were incorporated into this work alongside the empirical data.

One of the largest ongoing efforts has been the U.S. NRC's SACADA database, which

collects detailed data from training simulator exercises [87]. SACADA's data taxonomy was developed with a strong basis in cognitive science, considering four main macrocognitive functions of operation and a wide array of situational factors, as well as error recovery and task dependencies. This data was indispensable towards model parameterization efforts for the wide variety of PIFs considered and for the significant number of data points in each macrocognitive function.

Other recent quantification efforts include IDHEAS-DATA, developed by the U.S. NRC to support the IDHEAS-G and IDHEAS-ECA methods [90]. This repository aggregates data from a multitude of scientific studies to quantify the probability of error under each PIF in the IDHEAS framework as well as several other relevant HRA parameters like minimum HEPs. The data was collected from a large-scale literature review, and then classified and generalized to support the IDHEAS methods. Further data from the IDHEAS project was made available by collaborators at Pacific Northwest National Laboratory. This resource categorized a large amount of psychological and organizational literature according to the IDHEAS PIFs, and included some additional resources related to the minimum HEPs for certain error types and the interactions between PIFs. The additional data proved extremely useful as a starting point to find further relevant quantitative psychological literature.

As a final data resource, other HRA methods were consulted to parameterize areas of the models not covered by any other source. Probabilities were extracted from the Phoenix method documentation [10], as well as from CREAM [16] and SPAR-H [1]. The latter two methods apply multipliers to the error probability to account for PIF effects, which were leveraged in a limited number of cases. The Phoenix method provides factors for the CFMs which correspond to the failure mode's HEP given that no PIFs are in a degraded state. Phoenix handles the uncertainty inherent to failure-causing factors using the Leaky NoisyOR function, which considers failure

causes as probabilistically affecting CFMs, and incorporates a leak factor to account for the likelihood that the CFM arises without any particular cause being present [10]. This base probability for the CFMs could then be modified by SACADA, IDHEAS, literature, or HRA method-derived multipliers to obtain conditional probabilities, in that order of preference.

3.4.2 Synthesis of Data

Bayesian networks are a versatile modeling tool that allow for the integration of many disparate data sources, making them ideal for HRA applications that require the synthesis of sources with varying levels of completeness and confidence. As the model's nodes are all discrete, Bayesian updating with a multinomial-Dirichlet conjugate distribution provided the most balanced solution to combining probabilities when the number of observations was known or could be estimated. This method is applied in a similar manner in NUREG-6949 to demonstrate the utility of the Bayesian prior estimates alongside data [84].

The Dirichlet distribution is a generalization of the beta distribution to accept categorical variables with an arbitrary number of possible outcomes. Like the beta distribution, it can flexibly assume a wide variety of shapes. The variables in these models may take on 2, 3, or 4 different values, which the Dirichlet distribution accommodates. The distribution is parameterized by a vector $\boldsymbol{\alpha}$, which may be interpreted as "pseudocounts" of how frequently the variable takes on each of its states. When categorical counts become available in the data, the distribution's parameter values can be updated simply by summing the data counts and the α parameter for each category according to Equation 3.1. α_i refers to the Dirichlet prior parameter for category i, n_i is the number of observed occurrences of category i in the data, and α'_i is the Dirichlet posterior

parameter for category i. These can then be divided by the total number of observations in order to directly obtain probabilities.

$$\alpha_i' = \alpha_i + n_i \tag{3.1}$$

In NUREG-6949, no data confidence or counts are provided; only the expert-elicited probabilities are given [84]. As the priors all fall between zero and one, they become negligible when combined with datasets as large as SACADA. To combat this, we assumed a confidence of 120 observations for the prior to account for the experts' significant experience in the nuclear field.

After combining data such as NUREG-6949 and SACADA that covered similar areas of the network, it became necessary to combine these probabilities with other sources to fully parameterize some of the larger conditional probability tables. Error multipliers provided in the IDHEAS-ECA method [17] were the primary source for this procedure. Each multiplier, covering one state of one PIF, doubled the amount of contexts that the BN was capable of quantifying. Additional multipliers were derived from IDHEAS-DATA and associated organizational and psychological studies [90]. Finally, in the few cases where no conditional probabilities could be found in the data or literature, PIF multipliers from SPAR-H [1] and CREAM [16] were used.

Some of the network's structures were parameterized with the use of HRA idioms, which are akin to building blocks for the dependencies between HRA variables [35]. One idiom in particular, *definitional* dependency, encodes node relationships through Boolean logic. This logic underpins the causal nature of CFMs and the failure of information-gathering MCFs. If any CFM occurs, then the function may be deterministically defined as being in a failed state. Likewise, if no CFM occurs, then there is no reason for the function to be marked as failed. Therefore,

the conditional probability table for the MCF node is essentially a Boolean OR gate. Similar logic was also applied to some areas of the All Loads node's conditional probability table. When Time Load is found to be Inadequate, it is deterministic that All Loads should be Very High. Conversely, when all of the PIFs that comprise All Loads are in a Nominal or better state, it is deterministic that All Loads should be Nominal, which results in no negative effect to the task's outcome.

One final adjustment to the networks' parameterization was necessary in cases of multiple negative PIFs or mechanisms. Many of the mechanisms, when present, had extreme negative effects on the likelihood of CFMs occurring. For example, a poor crew morale, motivation, and attitude resulted in an effective multiplier of 380 times the probability of CFM I4 (Decide to Stop Data Collection). Degraded attention and the presence of bias were also found to have strong effective multipliers of nearly 49 times and 23 times the probability of CFM I4, respectively. In the presence of all of these causal factors, the probability of CFM I4 occurring is almost certain. However, it becomes necessary to apply a correction factor so that the probability is not greater than one, like the correction applied in SPAR-H when three or more PIFs are in a negative state [1]. To prevent a deterministic view of failure in the face of these negative influences, this correction shown in Equations 3.2 and 3.3 allows for a miraculous success that becomes vanishingly less likely as the negative influences become stronger. $Pr(\overline{A})$ refers to the probability that the adverse outcome is avoided; $Pr(A)$ refers to the more likely outcome that the adverse state occurs. The variable x refers to the product of the nominal probability of an adverse state and the various negative multipliers causing it to be greater than one- as such, it is not itself a probability.

$$Pr(\overline{A}) = \frac{1}{x + 10} \tag{3.2}$$

$$Pr(A) = 1 - Pr(\overline{A}) \tag{3.3}$$

This correction factor was implemented to ensure that the probability of a success becomes smaller and smaller as more negative factors become present. The factor asymptotically approaches $\frac{1}{x}$ as the product of the negative factors become greater. This causes a more severe effect than the correction factor applied in SPAR-H, which asymptotically approaches $\frac{100}{x}$ for diagnosis tasks and $\frac{1000}{x}$ for action tasks. Especially in the presence of multiple failure mechanisms, the probability of a CFM occuring is almost certain, and the correction reflects this reality.

3.4.3 Parameterization with the SACADA Database

The U.S. NRC's SACADA database [87] provided a basis for parameterizing much of the network. To extract these probabilities, a similar method to Paglioni's BN model construction method using SACADA data [91] was applied.

The SACADA database stores training scenario data at the task level, with associated contextual factors for each task. These tasks are described by training objective elements (TOEs), and the operational context is characterized through a set of situational factors (SFs). SACADA employs the concept of macrocognitive functions, sorting TOEs into the functions of Detecting, Understanding, Deciding, Executing actions, and Supervising. Some SFs are applicable to only one of the macrocognitive functions, such as "Status of Alarm Board" only being applicable to the Detection function, while some of the SFs are applicable to tasks regardless of macrocognitive function and are denoted as "Overarching," such as "Workload" [87].

This structure can be compared to the HFE variables employed in this work. TOEs are analogous to MCFs, while SFs are analogous to PIFs. In addition to these variables, SACADA also introduces "error modes," which are CFMs. Similarly, SACADA also describes "error causes," which are analogous to human failure mechanisms. Each of these data fields was mapped to a variable in the Bayesian networks.

Finally, the outcome of each task is described by a performance variable that can take on the states UNSAT (unsatisfactory) or SAT (satisfactory), as well as the additional states of SATΔ (successful but deficient) and SAT+ (exemplary). For the purposes of this work, UNSAT and SATΔ are counted as failures, and SAT and SAT+ are counted as successes, following the method in [91]. Extracting the percentage of successes and failures for each cognitive phase and task context enables the calculation of HEPs.

3.4.3.1 SACADA Data Mapping and Cleaning

First, SACADA SF states were mapped to states of the BN nodes. The SFs of the SACADA database do not correspond exactly to the PIFs used in the networks, sometimes with multiple SFs describing a single PIF and some PIFs not covered at all. For many of the TOEs, no cognitive phase was indicated, or the TOE was marked as being in the Coordination phase. We retrospectively categorized these TOEs that were not marked either information-gathering, decision-making, or action execution based on keywords in the task description including those enumerated in Table 3.4. The remainder of the uncategorized TOEs were then either categorized according to analyst judgment or discarded.

The data provided in SACADA itself is relatively sparse. Failures are few compared to

Phrase	Cognitive Phase
Acknowledges the alarm, Report alarms	Information
Determines/Verifies/Diagnoses that [event is happening]	Information
Monitors, identifies, checks [parameter]	Information
Requests/receives [information]	Information
Dispatches [person] to get [information]; Inspection ordered	Information
Makes appropriate notifications; Notifies co-owners	Decision
Enters procedure; Transitions to procedure; Continues through procedure	Decision
Refers to/References/Responds/Direct actions/Briefs/Declares per [procedure]	Decision
Reviews procedure, Performs read-through	Decision
Trigger step/Calculates/Evaluates options/Prioritize	Decision
Determines action (is/is not) required; Determine procedure (does/does not) apply	Decision
Performs immediate actions from memory	Decision
Authorizes/Supervises	Decision
Directs/Performs/Continues through actions of procedure	Action
Tries/Attempts/Directs/Performs [plant actions]	Action
Dispatches [person] to [perform action]	Action
Raises, responds, closes, stops, continues, returns, starts, initiate, informs, contact, lowers, places, energizes, isolates, supervises, makes, removes, swaps, depressurizes, recloses, swaps, reduces, adjusts, deselects, takes, borates, blocks, commences, actuates [plant parameter or control]	Action

Table 3.4: Keywords and key phrases that were used to categorize TOEs into I-D-A phases.

successes, with an overall error rate close to 1.8%. Many of the SFs are marked as being not relevant to a given TOE, denoted "NULL." This introduces challenges to interpreting the data as it is often not clear whether the SF state is irrelevant or simply in a nominal state.

Previous work [85] applied downselection techniques to the dataset's SFs for ease of modeling. For the purposes of this work, it was preferable to remove as few SFs as possible, so that as much information as possible can be gained from the database. The only SACADA SF not considered here was "Manipulation Recoverability," as there was no node in the Action network to accept this information. We assumed that SFs were only marked as relevant when they were indeed relevant, that is, to take the information at face value and not to discard any variables for their sparsity.

Finally, the dataset was re-dimensioned according to the process previously detailed in [91]. Each row, corresponding to an MCF completed under a specific set of contextual factors, also includes the number of times this task was performed by the crews (n). Every row was duplicated and appended to the dataset $n-1$ times so that each row corresponded to only one instance of a function being performed. This step streamlined the process of obtaining value counts for both SFs and failures.

3.4.3.2 SACADA Data Extraction

The SACADA database was used for characterizing the bulk of the probabilistic causal relationships in the networks. To extract these conditional probabilities, an iterative counting algorithm was used, introduced in [85]. For a node *B*, the probability of taking on state *b* given that its parents $pa(B)$ are in the state $pa(B)_i$ can be computed as follows: the number of instances where the node is in state *b and* the parents are in state $pa(B)_i$, divided by the total number of instances where the parents are in state $pa(B)_i$. This is summarized mathematically in Equation 3.4. This computation can then be iterated over each set of node parent states and each node in the network.

$$Pr(B = b|pa(B) = pa(B)_i) = \frac{N_{B=b,pa(B)=pa(B)_i}}{N_{pa(B)=pa(B)_i}} \tag{3.4}$$

These conditional probabilities were then inserted into the network using the GeNIe Modeler program's Python wrapper, PySMILE [92]. PySMILE allows the user to directly populate the network's conditional probability tables for each combination of node, node state, and parent node states using a Python script. PySMILE also enables the user to set evidence on the network

and update beliefs, which automated the validation process later in Chapters 4 and 5.

3.4.4 Parameterization with IDHEAS-DATA, Scientific Literature, and HRA Methods

Although the majority of the networks' conditional probability tables could be populated through SACADA, it was necessary to consult other data sources to ensure that the networks were fully parameterized. As mentioned above, multipliers were derived and applied to the existing network probabilities.

IDHEAS-DATA and IDHEAS-ECA provide many of these multipliers outright [17, 90]. For example, IDHEAS-ECA gives an error multiplier of 1.5 for poor HSI. The value is averaged from several sources, including a study on NPP alarm integration, and a meta-analysis of aviation displays on pilot errors [93, 94]. This was applied to the networks to account for the effect of poor HSI on the probability of CFMs I1 (Key Alarm Not Responded To), I2 (Data Not Obtained), I5 (Data Misinterpreted), I8 (Wrong Data Attended To), and I9 (Data Not Checked with Appropriate Frequency). IDHEAS enabled the parameterization of many network arcs in this manner.

In addition to the main IDHEAS reports, the IDHEAS project also contains an extensive psychological and reliability engineering literature repository categorized by PIF. Some of these studies are already incorporated into the IDHEAS method, but many have not yet been. We extracted ratios between error rates and correlations between variables to find additional multipliers not already presented. For example, a study which focused on visual searches for rare targets found that attentional error rates jump from about 0.104 to 0.38 when the participant applies heuristic biases, a 3.65 times increase [95]. This was applied to the conditional probability

$Pr(Attention = Degraded|Bias = Yes)$. The IDHEAS-DATA repository also provided a starting point for literature search terms to find further psychological literature relevant to the remaining unparameterized arcs in the network. Many arcs were parameterized in this manner.

As a final source, when no further empirical data was available, multipliers from the CREAM and SPAR-H methods [1, 16] were applied to the last remaining arcs. This was not preferred, as these methods cannot trace the values back to any particular source. However, it was necessary to ensure that the networks were completely parameterized. One example of the application of CREAM multipliers is to account for the effects of good HSI, as neither SACADA nor IDHEAS provides data for this potential scenario context. A multiplier of 0.5 was applied to obtain the conditional probability of CFMs I1, I2, I5, I8, and I9, given that there is a particularly supportive HSI.

3.5 Result: Model Quantification Capability

After parameterizing each conditional probability table in the information-gathering model, it is possible to quantify the probability of information-gathering errors given a scenario context. Bayesian networks allow the analyst to set evidence on nodes representing their state of knowledge of the situation. Here, this can be as much or as little evidence as is available: it is possible to set evidence on all of the PIF nodes, or none at all. The network can also be used by setting evidence on the mechanisms or CFMs and reasoning backwards to their most likely causes.

The fully parameterized network is shown in Figure 3.4. Because the GeNIe software displays probabilities to a precision of 1%, the probability of each CFM occurring with no evidence set on the network is displayed in Table 3.5. Most of the PIFs, as expected, are predicted to be in

a nominal state for the majority of the time. Some PIFs, such as Task Complexity and All Loads, are more likely to be in a degraded state. The CFMs have probabilities on the order of 10^{-2} or 10^{-3}, which is in line with conventional HRA knowledge.

CFM	**Probability of Occurrence**
I1: Key Alarm Not Responded To	$2.2 \times 10^{-}2$
I2: Data Not Obtained	$2.1 \times 10^{-}2$
I3: Data Discounted	$2.4 \times 10^{-}2$
I4: Decide to Stop Data Collection	$1.3 \times 10^{-}2$
I5: Data Misinterpreted	$4.7 \times 10^{-}2$
I6: Reading Error	$1.1 \times 10^{-}3$
I7: Data Miscommunicated	$7.9 \times 10^{-}3$
I8: Wrong Data Attended To	$2.2 \times 10^{-}3$
I9: Data Not Checked with Appropriate Frequency	$3.3 \times 10^{-}2$

Table 3.5: Probability of each information-gathering CFM occurring when beliefs are propagated through the network with no evidence set.

Setting evidence on the network allows the analyst to update their beliefs about the likelihood of factors that may have caused the evidence (backwards inference), or the likelihood of the evidence's downstream effects (forward inference). As an example of forward inference, we examine a scenario where the analyst sets evidence on several PIFs. Perhaps the task at hand is very complex with a poor HSI layout, and they would like to assess the probability of an HFE under these conditions. This evidence has been set on the network, as shown in Figure 3.5.

Several downstream effects of this context can immediately be observed. First, a very high task complexity is certain to lead to the operator's senses being overloaded, and increases the likelihood of high task load and moderately high overall loads. In turn, this increases the likelihood of extreme or high stress levels. The combination of overloaded senses, high stress, high loads, and degraded alertness then greatly increases the likelihood of degraded attention. The probability of degraded attention is now roughly quintupled from its likelihood without observing

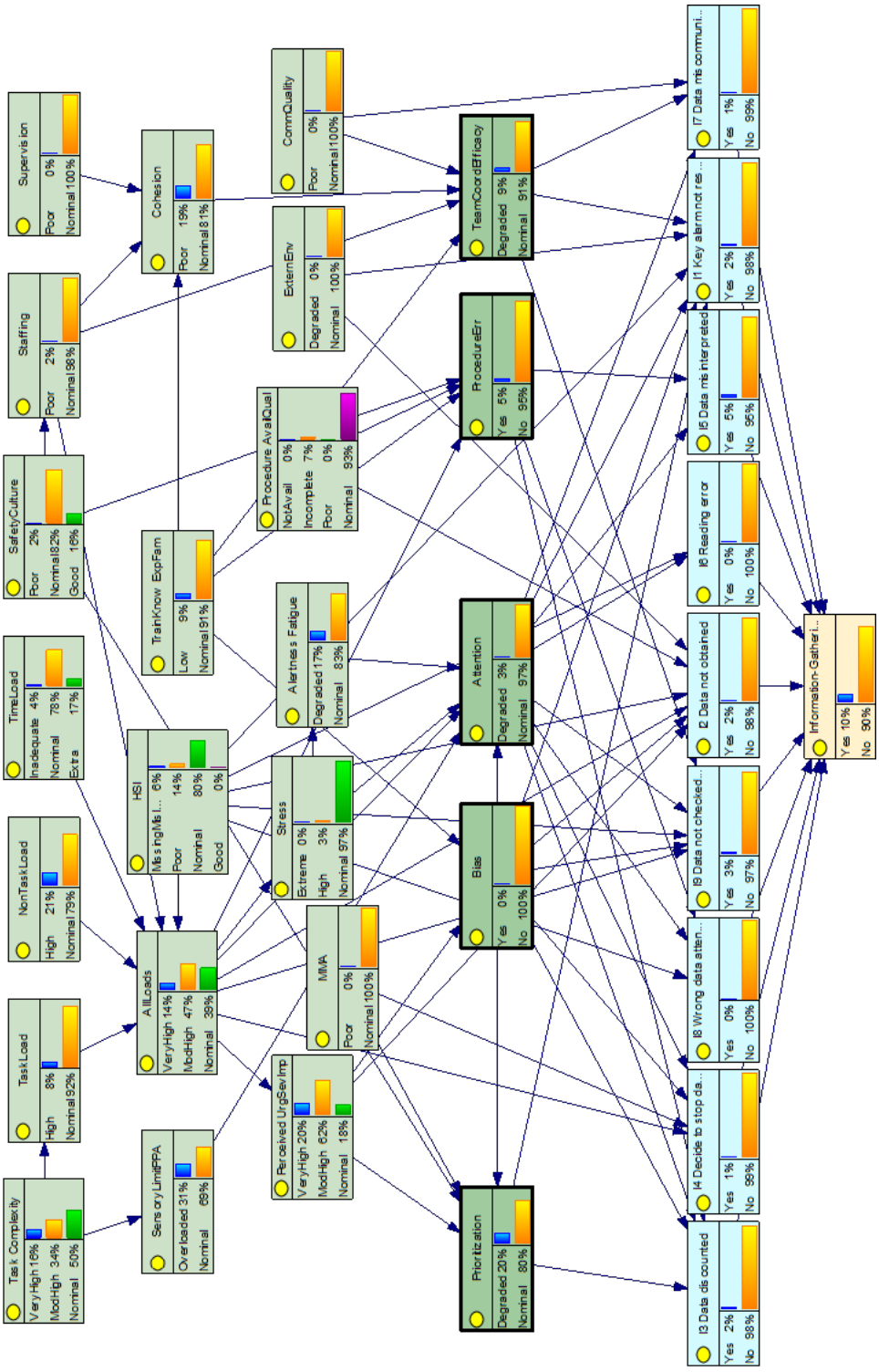

Figure 3.4: Fully parameterized Bayesian network model for information-gathering human failure events, with more detailed CFM probabilities in Table 3.5.

this evidence, increasing from 3% to 15%. The other mechanisms are largely unaffected, because no evidence has been set on the PIFs that are their main causes.

It is also apparent that the likelihood of several CFMs has greatly increased, especially for I1, I3, I4, I5, and I9. These failure modes are all strongly linked to the mechanism of attention, as well as HSI and loads to the operator. These attentional and overloading issues mean that the operator may unintentionally be pulled away from the correct information sources, or they may discount or misinterpret what they do see. The other CFM probabilities remain at roughly the same values, with the probability of I2 (Data Not Obtained) even decreasing a bit. The presence of bias is the single greatest predictor of CFM I2, as this is the crew's conscious decision to stop seeking more information. A slight decrease in the probability of very high perceived urgency and severity causes a decrease in the probability of bias, and thus a decrease in the probability of CFM I2 occurs.

Just as one can apply forward inference to examine how evidence on contextual factors shapes up into later effects, one can also apply backwards inference to discern what recipe of factors results in a particular effect. For example, the analyst may want to examine which factors are most likely to lead to the operator's error in deciding to stop data collection (CFM I4). Figure 3.6 shows the result of setting this evidence on the network.

Because degraded attention and the presence of bias are the main mechanisms that cause the CFM, the probability of these two nodes being in an adverse state increases significantly. The probability of degraded attention jumps from 0.032 to 0.63, and the probability of bias being present significantly increases from 0.0032 to 0.041. The PIF state of poor MMA is also a main cause of this error, and it increases from a probability of 0.00035 to 0.021. An increase to the probability of degraded attention similarly increases the probability of several explanations,

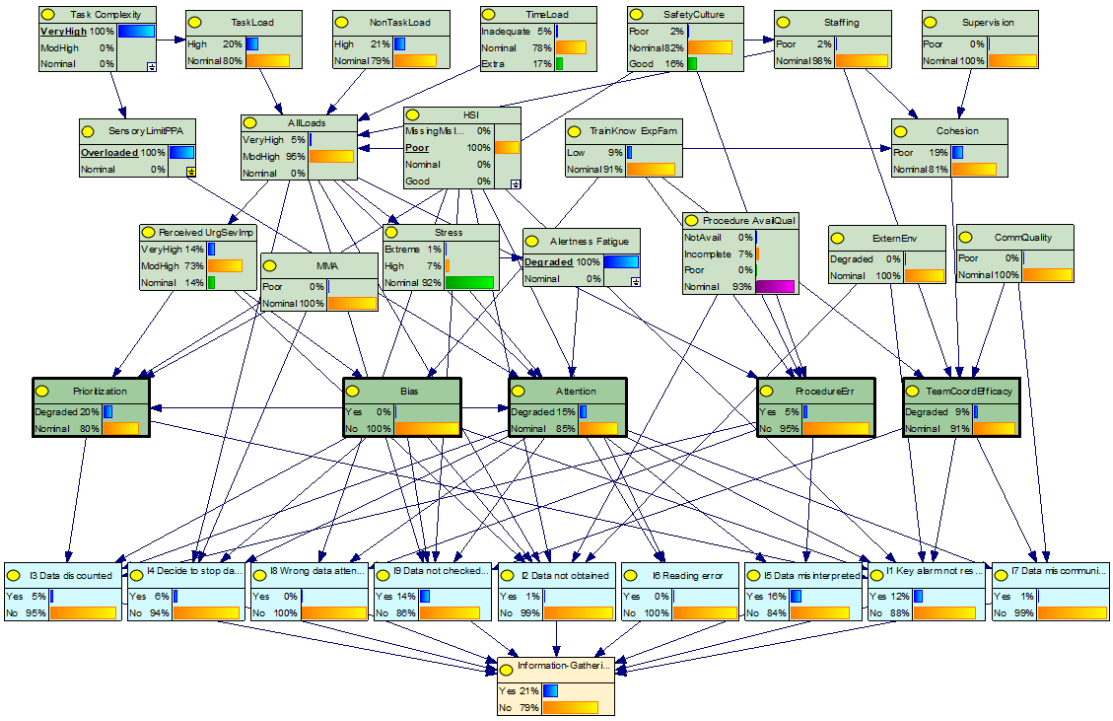

Figure 3.5: Information-gathering BN with evidence set: Task Complexity = Very High, HSI = Poor, and AlertnessFatigue = Degraded.

including high stress, very or moderately high loads, and overloaded senses. To prevent an error of this sort, these are the main factors an organization should strive to mitigate.

Another BN property of observing evidence on an effect is that beliefs can be indirectly updated about other effects. The CFM nodes are conditionally independent if no evidence is observed on their causes. However, if evidence is observed on a CFM, it is possible to show its comorbidity with other CFMs by updating the likelihood of their mutual causes. We can observe that by setting evidence on CFM I4 occurring, probabilities of the CFMs I1, I5, and I9 are greatly increased, with probabilities of the other CFMs also increasing to a lesser extent. This indicates that mitigating the causes of CFM I4 will also lead to a reduction in these types of errors, as they are causally related.

It is important to note that not all CFMs may be relevant to a particular situation. CFM

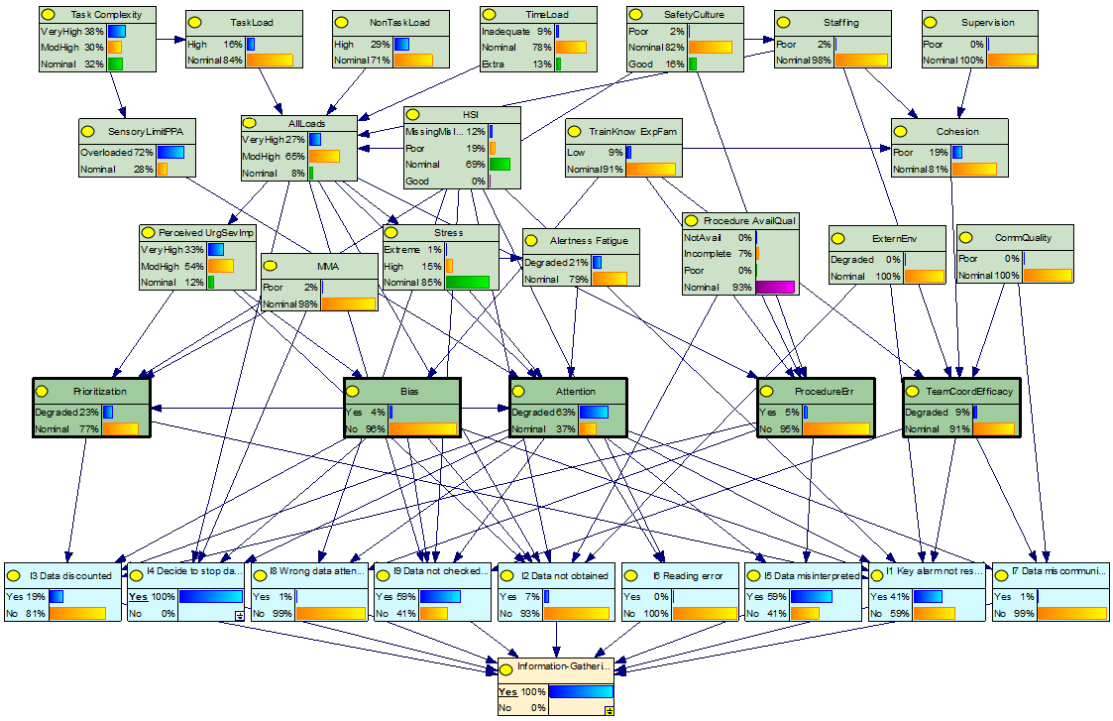

Figure 3.6: Information-gathering BN with evidence set: CFM I4 (Decide to Stop Data Collection) = Yes.

I1 (Key Alarm Not Responded To) cannot occur for information-gathering tasks that do not involve responding to an alarm. Likewise, CFM I7 (Data Miscommunicated) cannot occur for information-gathering tasks that do not involve communication. Guidance for deciding which CFMs are relevant to the tasks in a particular event is outside the scope of the present work. However, this guidance is provided in the Phoenix method in the form of fault trees taking into account various scenario characteristics. Different CFMs might apply depending on whether the crew is following operational procedures or their own knowledge, or whether the information-gathering process is active or passive. This can be implemented in the BN models by externally controlling the values of the irrelevant CFMs that they have not happened.

3.6 Validation: Expert Discussions and ATHEANA Narratives

Expert feedback sessions were held in December 2023 for an initial validation of the models' structures. These sessions were intended to ensure model completeness, accuracy, and parsimony using the feedback of experienced analysts. We conducted two structured sessions with several eminent experts in the HRA field whose expertise totaled 92 years. Some persuasive arguments were made for the addition and removal of arcs from the network, and many of these changes were implemented in the finalized structures.

The second stage of validation involved leveraging operational experience narratives to validate the progression of the models' causal chains through real-world events. We were directed toward the Korean Institute for Nuclear Safety (KINS) OPIS database [96], as well as pre-analyzed narratives in the appendices of the ATHEANA method's technical basis [97]. We chose to use the ATHEANA narratives towards the validation of several causal pathways leading to failure.

3.6.1 ATHEANA Case Study Narrative: Three Mile Island

The technical basis of the US NRC's ATHEANA HRA method [97] contains a retrospective analysis of several selected nuclear power plant operational events that resulted in accidents or near-misses. Each event analysis contains a summary of the event, a timeline of the unsafe actions and recoveries, and the key mismatches between operators' mental models of the situation and the reality. Crucially, the analysis also characterizes the relevant performance shaping factors and categorizes the unsafe actions into failures of specific macrocognitive functions. This makes the narrative analyses relevant for validating the causal chains of PIFs that lead to the failure modes

as represented in our BN models. Five of the ATHEANA analyses were used in model validation, covering a wide variety of CFMs and their causes. The analysis of the 1978 events of Three Mile Island was most applicable in validating the information-gathering network structure.

The incident's initiating event was a loss of feedwater to the steam generator which caused a reactor trip. After the reactor tripped, the emergency feedwater (EFW) system automatically activated as expected, but due to recent maintenance, the EFW inlet block valves were closed, so the water could not reach the steam generator. The control room panel indicated that the valves were closed, but the position indicators were inappropriately hidden with a maintenance tag. As the pressure continued to increase in the steam generator, the pressurizer relief valves automatically cycled. However, the pressurizer emergency relief valve became stuck open, which caused a continual loss of coolant. There was no direct indicator of the valve being stuck open, because its position indicator showed the "demanded" position as being closed. Another indicator could have shown that the valve was open due to high temperature in the line, but the operators discounted this information as the valve frequently leaked.

Due to the operators' Navy training background, they believed that the steam bubble in the pressurizer was shrinking, which was one of the worst things that could happen. They believed that this could cause loss of pressure control and a loss of coolant accident (LOCA), while not realizing that a LOCA was indeed occurring in the plant. Due to this inaccuracy, they incorrectly shut off the high-pressure injection pumps. After some time, this action allowed two-phase flow in the primary loop that caused potentially damaging vibrations to the coolant pumps, so they were shut off as well. This series of actions caused the water level to drop, exposing the core, which led to overheating and core meltdown [97, 98].

The ATHEANA analysis identifies several main drivers of error present in the situational

context. Several informational mismatches contributed to the operators' incorrect mental models of the scenario and their eventual adoption of an inappropriate strategy. The human-system interface also proved problematic, as many necessary indications were inaccessible or faulty. The operators' training and system knowledge base limited the effectiveness of their response. In particular, they were not familiar with the situation dynamics of this type of LOCA, and their prior training caused them to improperly prioritize the available information. Lastly, the available procedures did not provide useful guidance on this type of scenario so they were forced to rely on only their knowledge.

3.6.2 ATHEANA Validation of Network Structure

In addition to the explicitly identified PIFs in the ATHEANA analysis, several implicit PIFs can be gathered from the narrative. First, the operators were under a considerable amount of stress as they knew that a serious accident was possible. The crew members were also not at their most alert, as the initiating events and responses occurred between the hours of 4 to 5 a.m. The tasks at hand were highly complex as well due to conflicting and obscured system information. Perhaps most important of all, several narrative details point to a deficient safety culture at the plant. Instead of addressing the leakiness of the pressurizer emergency relief valve, the operators compensated by habitually ignoring important indicators. In addition, the EFW inlet block valves should not both have been left closed, as this leaves the plant vulnerable and without the option to activate secondary cooling. A more systemic safety culture oversight was also present in the lack of procedures and training to cover the appropriate actions during small-break LOCA events.

These factors culminated in two main information-gathering failures. The first was that the

operators decided to discount information indicating that the pressurizer emergency relief valve was stuck open, which corresponds to an instance of the Phoenix CFM I3: Data Discounted. The second was that the operators then failed to collect the correct data, as several instruments could have revealed clues that a LOCA was in fact occurring. They instead focused on the increasing pressurizer water level to monitor what they thought was a shrinking steam bubble. This was an instance of the Phoenix CFM I8: Wrong Data Attended To.

Figures 3.7 and 3.8 highlight the partial Markov blankets of both CFM I3 and I8, demonstrating the causal pathways leading from degraded PIFs to the resultant information-gathering failures. For CFM I3 (Data Discounted), it is clear that the data was discounted due to a poor *prioritization* of cues, misplaced *attention* due to *biasing* heuristics, and a misapplied *procedural understanding* of the valve indicator. Poor prioritization was primarily driven by a lacking *safety culture*, and secondarily driven by a spurious *HSI* and high *task complexity* driving *loads* that caused the crew to underestimate their *perceived urgency and severity* of the indicator. Furthermore, their *training and knowledge* combined with the severity underestimate caused them to become biased against this indicator, directing their attention elsewhere. The presence of *stress* and *fatigue* also prevented the crew from focusing their attention on this issue. A procedural misunderstanding also occurred as the operators did not understand that the "demanded" position need not necessarily reflect the valve's actual position in the plant. Because this type of stuck-open failure was not anticipated, they were unable to compensate with either their training or procedures.

CFM I8 (Wrong Data Attended To) can be traced to similar causal origins, as shown in 3.8. This error was primarily caused by the operators' *biases* towards collecting more data about the shrinking steam bubble. This caused them to direct their *attention* away from indicators that

would have helped them. Even though this instrumentation existed, a generally poor *HSI* made it unlikely to be accessed. Like the error resulting in CFM I3, a *procedural misunderstanding* also drove this decision to continue monitoring the pressurizer water levels instead of seeking out this new information. This misunderstanding stemmed from a lack of *procedures* and *training* to cover the scenario, and a poor *safety culture* causing a lack of concern about these particular event dynamics. The operators' biases stemmed from their *training* having been on a different system architecture and their high *perceived urgency and severity* of the pressurizer water levels over the other indicators. Lastly, the high *task complexity*, high *stress*, and high levels of *fatigue* diverted attentional resources away from attending to the correct data.

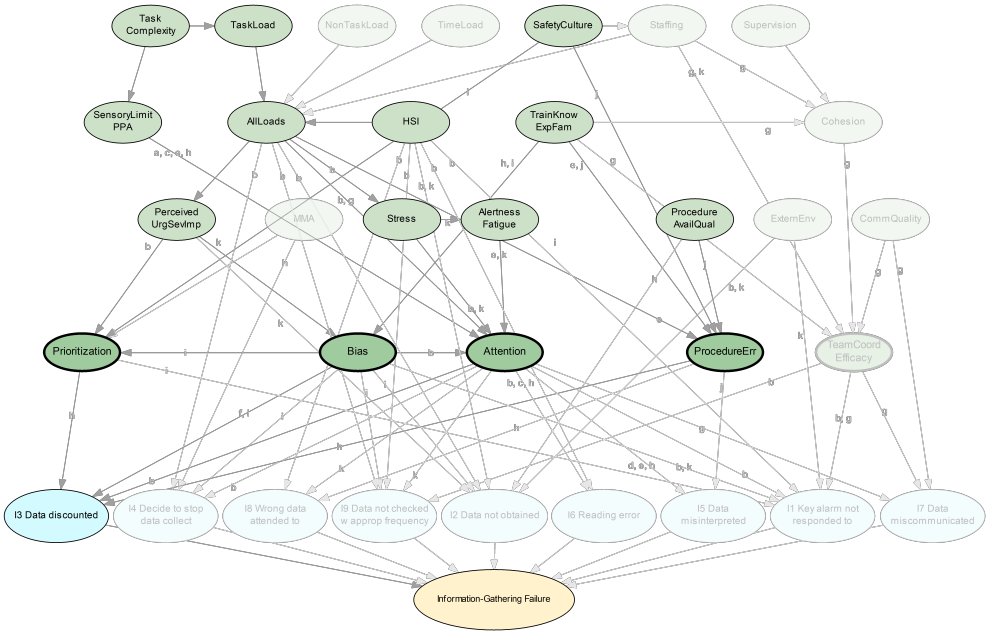

Figure 3.7: Partial Markov blanket highlighting the causal structure of factors leading to CFM I3: Data Discounted for the information-gathering failure that occurred at Three Mile Island.

It is possible to set evidence on these PIFs in the information-gathering network to observe how the probabilities of CFM I3 and I8 increase under these conditions. After setting evidence on the PIFs (with no evidence observed on the mechanisms), the probability of CFM I3 jumps from

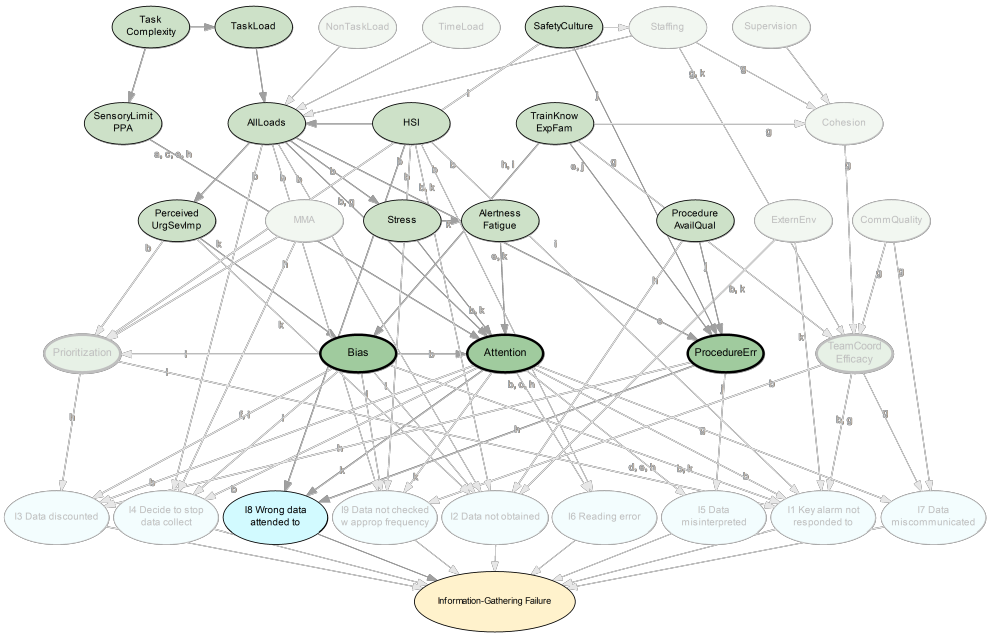

Figure 3.8: Partial Markov blanket highlighting the causal structure of factors leading to CFM I8: Wrong Data Attended To for the information-gathering failure that occurred at Three Mile Island.

0.024 to 0.065. The probability of CFM I8 occurring also jumps by nearly an order of magnitude from 0.0022 to 0.021. When further evidence is set on the mechanisms of bias, prioritization, attention, and procedure error as being degraded, CFM I3 is found to be nearly certain to occur.

3.7 Discussion and Conclusion

In this work, we developed a Bayesian network for nine Information-phase CFMs, fully substantiated with cognitive literature and parameterized through empirical data. This model will lead to a more valid, justifiable HEP estimation by eliminating uncertainties related to the origins of the causal relationships. Rather than assuming a given task is particularly prone to failure, a root cause can be identified so that an effective and targeted mitigation strategy can be developed. The modular nature of the network structures allowed for expedited development of decision-making and action-taking Bayesian networks, which are detailed in Chapters 4 and 5.

The development of this BN model is a step forward in enhancing the causal basis of HRA by adding theoretical foundations to the connecting arcs and causal relationships between variables. Additionally, the completed models present an opportunity to apply this method to the external actions domain [99], which we intend to complete in future work. We uncovered a wealth of psychological/cognitive literature, while well-known to human factors professionals, that had not been previously utilized in HRA to build a BN's causal connections. This improves upon current models which rely primarily on expert opinions and judgments. Several important insights were gained during this process.

We identified a set of five central PIFs that act as cognitive mechanisms through which CFMs may occur. These are similar to mechanisms of mechanical failure such as corrosion, fatigue, or material overstress. In turn, these mechanisms can be influenced by one or more layers of PIFs, depending on logically derived relationships and the documentation found in cognitive literature. A more comprehensive discussion of these mechanisms and their place in HRA is presented in [33].

Furthermore, the mechanism of coordination was found to be the one with one of the most complex causal origins. While the other mechanisms model a single operator's internal cognitive processes, any mechanism relating to coordination or team performance necessarily involves multiple operators' internal and external cognitive processes. These processes are not only affected by person-, situation-, organization-, and machine-based PIFs, but also team-based PIFs.

Incorporating all of the data sources to parameterize the networks was a uniquely challenging task. The sources are presented in disparate formats, without standardized variable names or states. Each source had distinct strengths and weaknesses in this application which required judgment to overcome. Nevertheless, Bayesian models contain built-in methods for handling

these challenges and incorporating data from many separate sources.

Insights from this work can also be used to guide future data collection efforts. Many data sources did not consider specific enough PIFs, or did not consider certain areas of PIFs. For example, SACADA would benefit from considering more organizational and personal PIFs, such as safety culture, quality of procedures, and individual abilities and motivation. Mechanisms, though not always observable, could be appended to current failure databases through retrospective interviews with individual operators.

These models will soon be ready for regulatory applications in the nuclear domain, and ongoing efforts have been focusing on integration into the Phoenix method. It is our belief that the models can be further generalized to other complex engineering systems, as the model structures and parameters are derived from cognitive and reliability engineering literature from a variety of applications. A more robust causal understanding of the factors leading to HFEs is integral to ensuring safe operations and supporting the energy security benefits offered by nuclear power.

Chapter 4: Causal Pathways Leading to Human Failure Events in Decision-Making System Response Activities

This chapter is reproduced from [100] and [101].

Camille S. Levine, Ahmad Al-Douri, and Katrina M. Groth. "Causal Pathways Leading to Decision and Action Human Failure Events: Structures and Validation." In: *Proceedings of the 17th Probabilistic Safety Assessment and Management Conference (PSAM17) & Asian Symposium on Risk Assessment and Management (ASRAM2024)* (Oct. 7–11, 2024). Sendai, Japan.

[Submitted, in press]: Camille S. Levine and Katrina M. Groth. "Three Novel Causal Logic-Informed Models for Identifying and Quantifying Operator Errors, Part II: Decision-Making." In: *Reliability Engineering & System Safety* (2025).

4.1 Introduction

Decision-making is one of the more intricate functions carried out by human-machine teams in complex engineering systems. For a decision-making error to occur, the team must first have correctly gathered and understood information relating to the system's status, but have insufficient awareness of the broader situation at hand [23]. There are many ways that teams can

make incorrect decisions, and many unobservable cognitive processes that go into response selection. Frequently, decision-making errors manifest as the team's poor adherence to procedure, but sometimes the incorrect decision remains unnoticed until an erroneous action has already been performed. This stage of operations is as crucial to team success or failure as the other functions of information-gathering or action execution [19].

Several HRA methods in use refer to all cognitive activities as decision-making, which may introduce ambiguity or bias through lack of specificity, and ignores the functionally distinct role of information-gathering. Furthermore, some methods give the analyst the option to not model cognition at all after the crew's initial diagnosis [12]. Although these simplifications reduce the complexity of the analysis, the accuracy of the results is compromised, as decision-making tasks present distinct opportunities for serious error. To ensure completeness of the model suite, equal importance was placed on the decision-making model as the information-gathering or action execution models.

Decision-making in complex engineering systems, particularly nuclear power plants, is marked by dynamic event progression which the operators need to react to in real-time [2]. Adverse events may be accompanied by abnormal plant behavior or spurious indications that do not match the expected states that operating procedures assume as prerequisites [102]. In other extreme task conditions that cause high stress and high workload, operators are especially prone to decision-making errors [81]. Latent issues with safety culture and training can also lead to incorrect decisions through different causal pathways [103].

Like the model developed and presented in Chapter 3 for information-gathering tasks, there is a need for similarly rigorous, literature-based models for decision-making errors. The rest of this chapter details the structure, parameterization, and empirical validation of a Bayesian

network modeling the causal pathways that lead to decision-making HFEs. Two related causal models for information-gathering and action execution errors are also included in this validation exercise.

4.2 Method: Model Structure Development and Parameterization

4.2.1 Model Structure and Literature Synthesis Method

To build the causal model's structure, we synthesized methods, variables, and data from a wide variety of HRA resources. Development of three Bayesian networks for modeling information, decision, and action errors is intended to match the three cognitive phases of the Information-Decision-Action in Crew Context (IDAC) framework [19]. IDAC defines these high-level crew functions as macrocognitive phases. Another foundational document was the Phoenix HRA method, which defines seven decision-making crew failure modes (CFMs) that describe how operator decision errors may manifest [10]. These CFMs, shown in 3.2, are designed to exhaustively cover all of the potential failure modes that may occur in the macrocognitive phase of decision-making without overlapping. Major crew functions (MCFs) are high-level actions taken by the human-machine team, and correspond to a specific instance of a macrocognitive function. An MCF is rendered failed if a CFM occurs during the task.

In previous work [22], we applied task decompositions from NUREG-7256 [60] to identify potential human failure events that may occur during external flood hazard mitigation actions. During the identification of potential human failure events and their requisite failure modes, an eighth decision-making CFM was posited (D8: "Premature Termination of Procedure"). An example of this CFM might occur during flood mitigation measures such as building a sandbag

barrier. The crew may correctly assess that a barrier needs to be built, but incorrectly decide to stop placing sandbags too soon, resulting in an ineffective barrier.

Groth and Mosleh's performance influencing factor (PIF) taxonomy [24] is used to characterize which aspects of the situational context are relevant to each CFM. The taxonomy, previously shown in Table 3.1, categorizes PIFs into machine-, person-, team-, organization-, and situation/stressor-based factors. Our method leveraged the Phoenix method's CFM definitions, provided in [23], to identify broader mechanisms of failure, then PIF categories, and then specific PIFs relevant to each failure mode. The exact method of model development is described in further detail in Chapter 3 and summarized here in Figure 4.1.

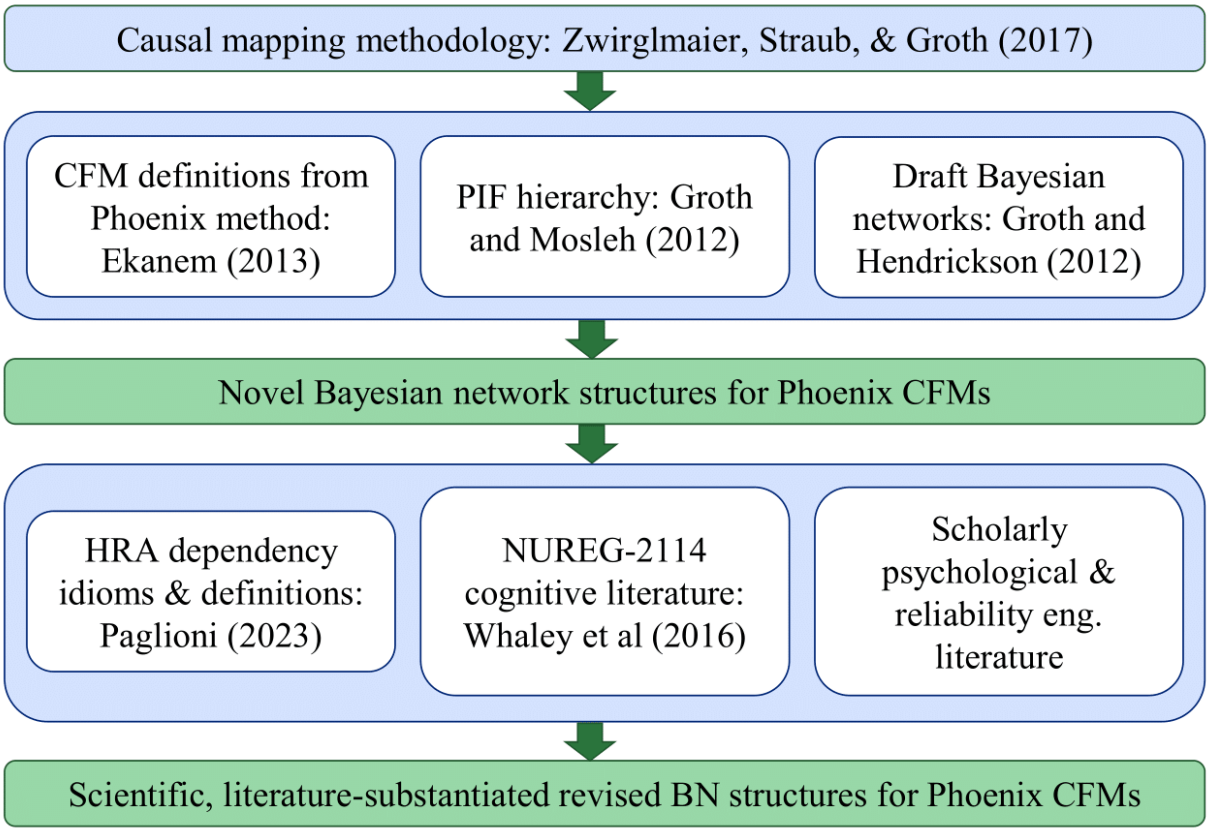

Figure 4.1: Sources used in the development of human performance Bayesian networks.

After developing the preliminary model structure, literature from the U.S. NRC's Cognitive

Basis for HRA [32] was used as a starting point to find additional research focused on naturalistic decision-making. After thorough examination of the literature, additional arcs were introduced and a few arcs were deleted from the model to better reflect real-world cognitive processes. Then, each arc in the network was substantiated with an excerpt from the literature.

Specific literature used to substantiate the decision-making Bayesian network model is documented in Table 4.1. The letter associated with each source is included directly on the network to indicate its use in substantiating a particular arc. Endsley and Jones' work on situational awareness [104] and Wickens et al.'s work in human factors [83] were instrumental in substantiating a majority of the arcs. Bye's work on future needs of human reliability analysis was useful in finding relevant Halden Reactor Project reports that examine effects of coordination and human-system interface (HSI) [105]. Lipshitz's and Letsky et al.'s work on naturalistic and team decision-making were also used in developing and substantiating this network for knowledge base and communication, respectively [74, 80, 106].

4.2.2 Model Parameterization Method

Upon substantiating and finalizing the BN model structure, we parameterized the model with data to enable HEP quantification. Selecting among the HRA methods in current use typically means that the analyst must make trade-offs between a robust cognitive basis and ease of quantification [25]. As such, it was essential to implement a transparent, usable quantification process to render the causal structures effective analytical tools. Further details about the parameterization and data are available in Chapter 3.

Several data sources were synthesized to parameterize the model, including the U.S. NRC's

Ref on BN	Ref Number	Citation	Area on BN
a)	[107]	Nystad E., Kaarstad, M., & McDonald, R. (2019).	Human-system interface
b)	[108]	Tversky, A. & Kahneman, D. (1974).	Familiarity and bias
c)	[109]	Siddle, B. (2008).	Stress
d)	[81]	Orasanu, J. & Martin, L. (1998).	Factors relating to mental model, overloading
e)	[73]	Endsley, M. (1995).	Effects of knowledge base and information
f)	[102]	Massaiu, S. & Holmgren, L. (2014).	Procedure Quality
g)	[80]	Lipshitz, R. (1993).	Knowledge base
h)	[110]	Braarud, P. & Johansson, B. (2010).	Complexity and coordination
i)	[74]	Letsky, M. (2008).	Team communication
j)	[111]	Hildebrandt, M. & McDonald, R. (2020).	Training
k)	[112]	Andersen, E., Kozine, I., & Maier, A. (2019).	Knowledge, information, communication
l)	[106]	Lipshitz, R. & ben Shaul, O. (1997).	Knowledge base
m)	[104]	Endsley, M. & Jones, D. (2016).	Situation and stressor based PIFs, organizational PIFs, PPA, prioritization
n)	[103]	Reason, J. (2000).	Safety culture, procedures
o)	[113]	Greitzer, F. et al. (2010).	Mental model, skills
p)	[83]	Wickens, C. et al. (2004).	Perceived severity, complexity, mental model, procedures
q)	[114]	Wickens, C. et al. (2012).	Perceived severity
r)	[35]	Paglioni, V. P. & Groth, K. M. (2024).	Deterministic dependencies

Table 4.1: References used to substantiate Decision-making network model arcs.

SACADA database [87], NUREG/CR-6949 Bayesian priors [84], IDHEAS-DATA and IDHEAS-ECA multipliers [17, 90], other HRA methods including Phoenix, CREAM, and SPAR-H [1, 10, 16], and additional psychological and organizational literature. Bayesian networks are able to flexibly accept and integrate dissimilar data sources such as these, which ensures that all data is used in some capacity. The SACADA database was particularly useful for this decision-making model, as nearly all the relevant PIFs and CFMs are covered.

Bayesian updating was applied to combine data sources that covered the same nodes in the network. The model is comprised only of discrete nodes, so a multinomial-Dirichlet conjugate distribution was assumed. In some situations, it was necessary to derive and apply multipliers from the data in order to fully parameterize the larger conditional probability tables. Several nodes were quantified using the Definitional HRA dependency idiom [35], which applies Boolean logic to deterministically define the state of a child node.

The model was created using the GeNIe software and its Python wrapper, PySMILE [92]. GeNIe is a powerful visualization and construction tool for Bayesian networks. PySMILE enabled automation of the parameterization process through a simple Python script, and also allowed the validation exercises to be more easily automated.

4.3 Result: Model Structure

4.3.1 Model Architecture

The full Bayesian network structure is shown in Figure 4.2. A source substantiating each arc in psychological literature is indicated on the arc. Some of the structures are substantiated through the presence of definitional idioms from [35] rather than through literature sources. For example, inadequate *time available* deterministically defines that *all loads* will be very high, as the operator must rush to arrive at the correct conclusion. This may in turn induce errors in the form of an incomplete mental model. A logical relationship exists between these two constructs whereby time loads are certain to influence the total loads experienced by the operator.

In the lower layer of the BN, there are eight decision-making CFMs, which are shown in light blue. The occurrence of any CFM deterministically defines the MCF, shown in light yellow,

as failed.

There are five key causal clusters apparent in the structure of the decision-making human performance model. Each cluster corresponds to a mechanistic PIF, denoted by nodes in darker green with thicker borders. These mechanisms are *Perceived Severity*, *Skills*, *Bias*, *Prioritization*, and *Mental Model*. These clusters are probabilistic in their causation of error, increasing the probability of occurrence of a CFM. Multiple mechanisms can contribute to the likelihood of a particular CFM with differing degrees of influence.

In addition to the mechanisms, there is the presence of an "Information Failure" node, shown in darker yellow. This encodes the potential for information-gathering failures to manifest through the failure of a decision-making MCF. Although faulty or delayed information acquisition is the result of an information-gathering human failure event, the actual failure may transpire as a decision-making error, or even later downstream in operations during the action execution stage. If the crew takes incorrect data as fact, their situational awareness is degraded, leading to an incorrect assessment of the plant state and potentially choosing an incorrect course of action. Similarly, with delays in gathering information, or conflicting data, the crew may decide to delay acting on their current knowledge of the system. The Information Failure node takes into account situation dynamics, and the sequential nature of information-gathering occurring before decision-making.

4.3.1.1 Model Architecture: Human Failure Mechanisms

As described above, there are five main human failure mechanisms that significantly increase the likelihood of decision-making CFMs. These mechanisms arise from the combination

and interaction of multiple PIFs, which may in turn be indirectly influenced by other PIFs.

The first mechanism cluster, *perceived urgency/severity*, directly causes two CFMs (D1 and D8). Perceived urgency/severity is directly caused by three PIFs (perceived decision weight, overloading, and task complexity). The indirect causes of the mechanism are the PIFs comprising loads to the operator, including time load, task load, non-task load, and task complexity. A more challenging or high-stakes situation will increase their perceived urgency and change how the task is approached.

The second mechanism cluster, *skills*, also directly causes two CFMs (D2 and D7). Skill errors are directly caused by four PIFs (personal and physical abilities (PPA), training/knowledge/ experience/familiarity, use of information, and stress). Skill errors may also be indirectly caused by issues with an individual's morale, motivation, and attitude (MMA) causing poor information use, overloading which causes stress, and poor safety culture which may lead to a poor training program. The individual may lack the necessary situational knowledge or ability to act upon this knowledge, or they may experience skill degradation due to stress.

The third mechanism cluster, *bias*, directly causes only one CFM (D4). Bias errors are directly caused by four PIFs (procedure quality, perceived decision weight, training/knowledge/ experience/familiarity, and perceived urgency). Bias errors may also be caused indirectly by overloading which causes high perceived urgency, or by issues with safety culture that lead to a lack of quality procedures or training. Biases are typically induced by the perceived situation characteristics, and may be worsened by poor procedural guidance.

The fourth mechanism cluster, *prioritization*, directly causes two CFMs (D4 and D5). Prioritization errors may also be caused by the mechanism of bias. It is caused by five PIFs: perceived decision weight, safety culture, perceived urgency/severity, MMA, and stress. Similar

to bias errors, prioritization errors are also indirectly caused by overloading caused by various demands to the operator, as well as the PIFs that cause bias errors through the paths previously identified. Prioritization, like bias, can be linked to perceived situation characteristics, but may also represent a problem endemic to the organization and its work processes.

The fifth mechanism cluster, *mental model*, directly causes a majority of the CFMs (D1, D3, D5, D6, D7, and D8). Poor mental models may be caused by four PIFs: training/knowledge/ experience/familiarity, overloading, communication quality, and information use. Poor mental models may be indirectly caused by issues with MMA causing poor information use behaviors, highly complex tasks exacerbating communication problems, safety culture leading to a lack of training program, and the demands to the operator that comprise overloading. Mental models are influenced by the operator's internal ability and capacity to integrate the available information into their knowledge base.

The complete network shown in Figure 4.2 captures the direct causal influences on CFMs and mechanisms as well as the indirect influences that propagate through secondary and tertiary contextual factors. This captures a wide range of factors that meaningfully influence human failure, documented through a strong, complete causal basis. Additionally, the resultant model shows the dependencies between PIFs, mechanisms, and CFMs, which represents a novel consideration of the interactions between different HRA variables.

4.3.1.2 Model Architecture: Performance Influencing Factors

In addition to the crew failure modes and human failure mechanisms, there is also an interconnected layer of PIFs (shown in lighter green) that acknowledge other influences of the

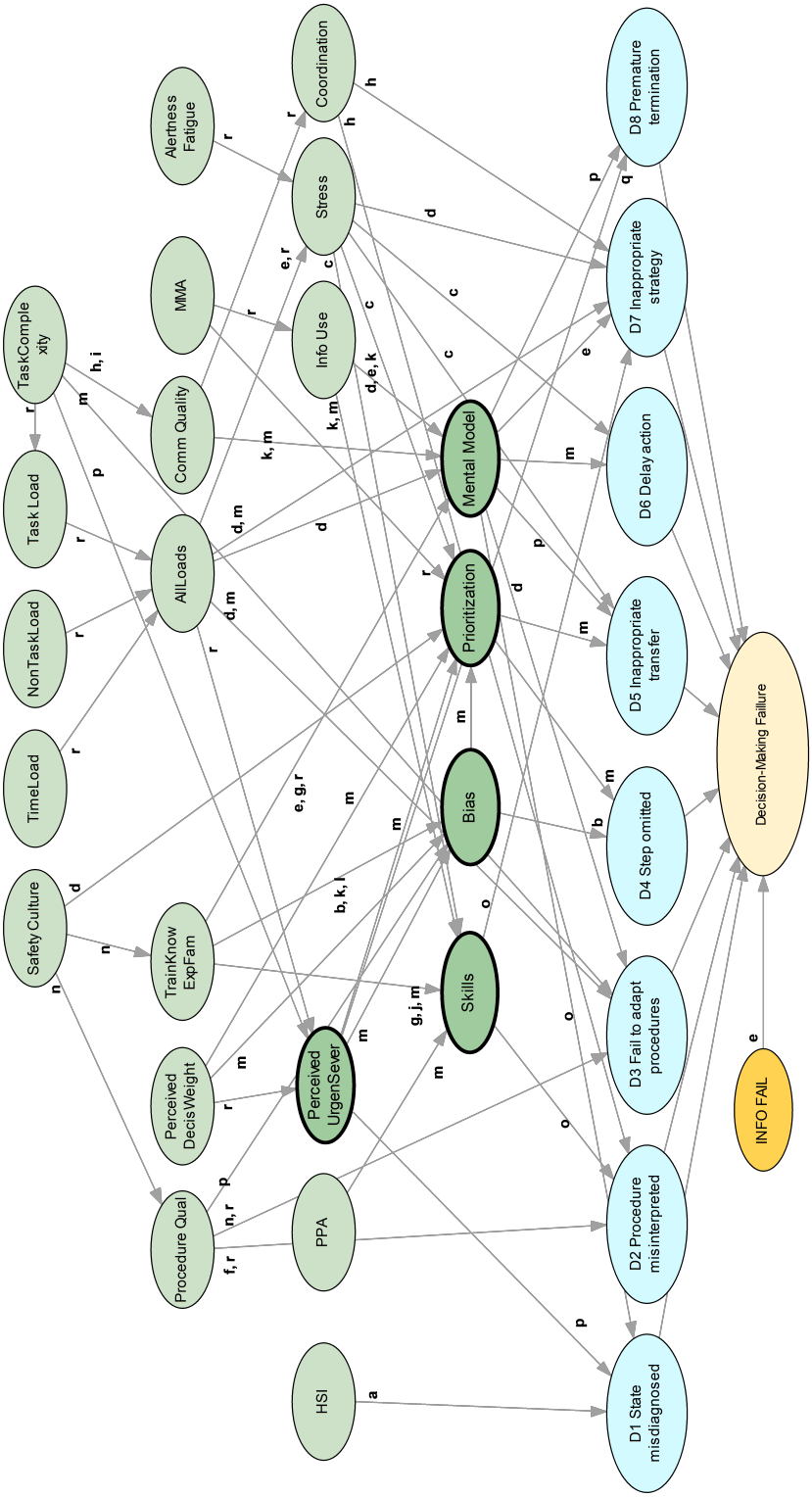

Figure 4.2: Full cognitive literature-substantiated Bayesian network structure for the Phoenix method's D-phase CFMs.

scenario context. Some of these PIFs influence the CFMs directly. For example, Procedure Quality directly causes the CFMs D2, "Procedure Misinterpreted" and D3, "Failure to Adapt Procedures." Beyond the mechanism clusters, there are several notable clusters of interrelated PIFs corresponding to the influences exerted among contextual factors.

First, the model employs a structure comprising All Loads to the operator. This includes the task load, non-task load, time load, and task complexity. Decision-making loads are not affected by the Staffing or HSI PIFs as they would be for information-gathering tasks, due to the internal nature of decision-making processes. When an operator becomes overloaded, these demands may then affect their perceived urgency of the situation and raise their stress levels.

Another area of PIFs is related to the organization's safety culture and the quality of its work processes, which may affect the operator's knowledge base and associated resources at their disposal. Safety culture may lead to a deficient training program or set of procedures, and also directly affects the mechanism of Prioritization. In an organization with poor safety culture, operators may prioritize productivity or output over safe outcomes. The quality of the operator's training directly affects their knowledge and experience level, which in turn affects their familiarity with the scenario. These nodes were included separately in the original model, but were condensed for ease of quantification as separating the effects of each facet of knowledge base in the data was intractable.

Finally, there is another cluster of PIFs related to the team's information dissemination practices and the individuals' use of that information. MMA may affect whether the operator chooses (or is able) to use the available information to its fullest extent. Poor integration of information can then lead to an incomplete or erroneous mental model. An incorrect mental model may also result from difficulties in communication. Good communication is necessary for

good teamwork, so team coordination issues tend to arise as well when communication is poor. A highly complex task may also render communication more difficult or enable misunderstandings. Coordination does not affect the mental model, but is directly essential for tasks such as selecting a strategy or interpreting procedure logic.

4.4 Result: Model Quantification Capability

The fully parameterized network is shown in Figure 4.3. Evidence has been set on the "Information Failure" node to indicate that a prior information-gathering failure has not occurred, so that the probabilities of pure decision-making failures can be examined more closely. The GeNIe software displays probabilities as a bar chart to a precision of 1%. This is insufficient for failure probabilities on the order of 10^{-2} or 10^{-3}, so the probabilities of each CFM (with no evidence set on the network) are shown in Table 4.2 to better illustrate the nominal failure probabilities.

Additional evidence can be set on any combination of nodes in the network to represent the analyst's state of knowledge about a past or future situation. It is possible to set evidence on PIF nodes and reason forward to the requisite changes to the probabilities of their effects, and it is also possible to set evidence on the mechanisms or CFMs to reason backwards to the most significant causes.

To demonstrate the quantitative explanatory power of the model, an HRA analyst may want to document the potential causes and effects of an operator's incorrect mental model. Evidence can be set on the "Mental Model" node that it is in a degraded state, and then the network's beliefs can be updated to access the new probabilities of its causes and effects. The result of this

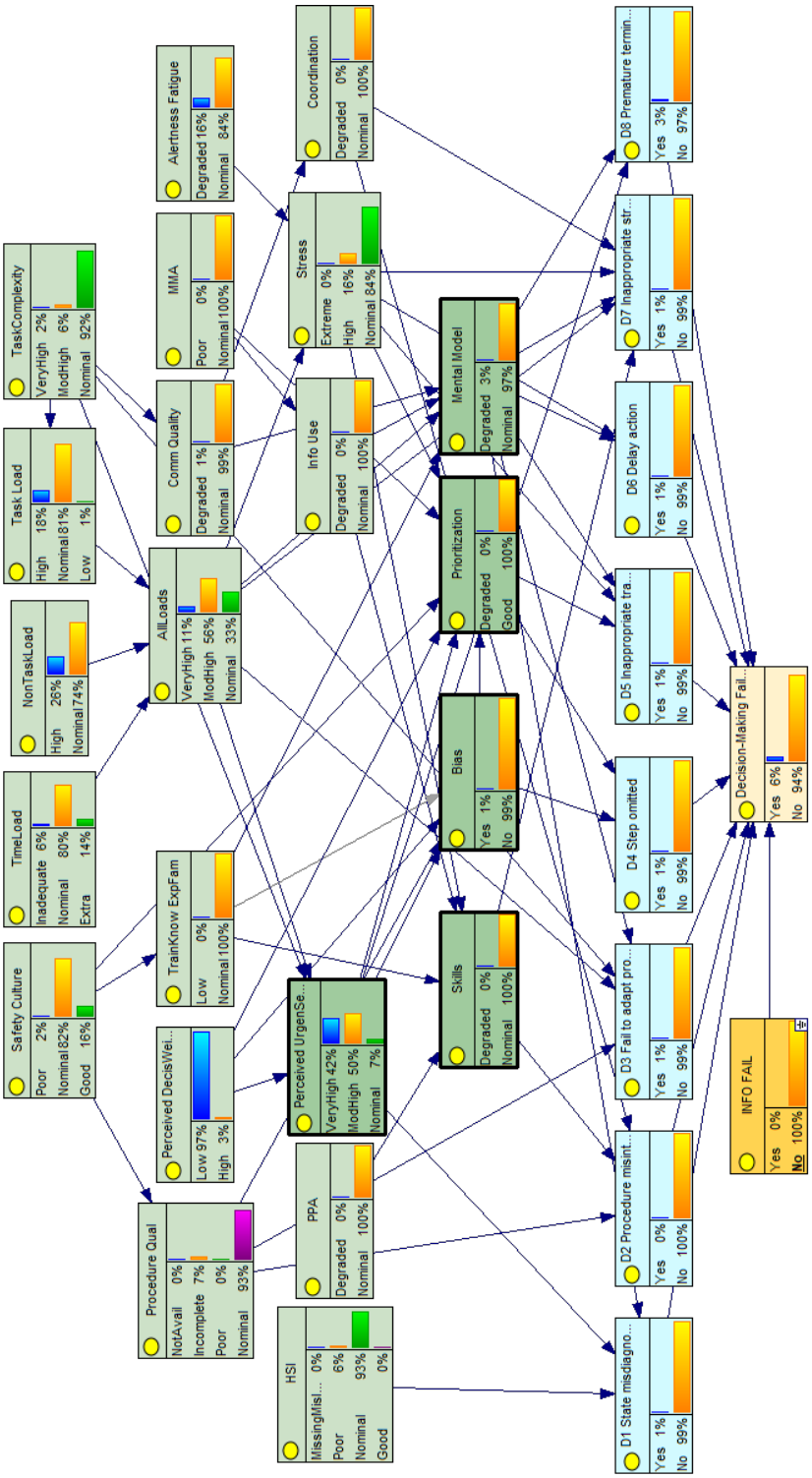

Figure 4.3: Fully parameterized Bayesian network model for decision-making failure events, with more detailed CFM probabilities in Table 4.2.

CFM	Probability of Occurrence
D1: State Misdiagnosed	$7.5 \times 10^{-}3$
D2: Procedure Misinterpreted	$4.4 \times 10^{-}3$
D3: Failure to Adapt Procedures to Scenario	$5.2 \times 10^{-}3$
D4: Step Omitted	$5.8 \times 10^{-}3$
D5: Inappropriate Transfer to Another Procedure	$7.7 \times 10^{-}3$
D6: Decision to Delay Action	$5.8 \times 10^{-}3$
D7: Inappropriate Strategy	$7.7 \times 10^{-}3$
D8: Premature Termination of Actions	$2.7 \times 10^{-}2$

Table 4.2: Probability of each decision-making CFM occurring when beliefs are propagated through the network with no evidence set.

procedure is shown in Figure 4.4.

As discussed above, the human failure mechanism of degraded mental model has a strong direct influence on many of the CFMs. It can be seen that the probabilities of CFMs D1, D3, D5, D6, D7, and D8 occurring greatly increase by several orders of magnitude to between 13% and 18%. This is a consequence of using the SACADA database; many of the data points marked as having a poor mental model are also marked as failure events. This renders mental model a significant predictor of failure, increasing the overall probability of failing a decision-making MCF from 6% to 63%. Because this mechanism is not a direct cause of CFMs D2 and D4, the probabilities of these CFMs are largely unaffected.

It can also be observed that the root causes of a degraded mental model also increase in probability. The likelihood of degraded communication increases to 16%, and poor information use increases to 6%. Information use, in turn, can be explained partially by poor MMA, which also increases in probability. The likelihood of combined loads being very high increases to 74%, a majority of the time. This increases the likelihood of time load, task load, non-task load, and task complexity being high. Lastly, the probability of a poor knowledge base (low training/knowledge/experience/familiarity) remains low, but increases from 7.8×10^{-4} to $3.1 \times$

10^{-3}. To reduce the probability of this mechanism occurring, these PIFs should be the first causes addressed in any HRA corrective action program.

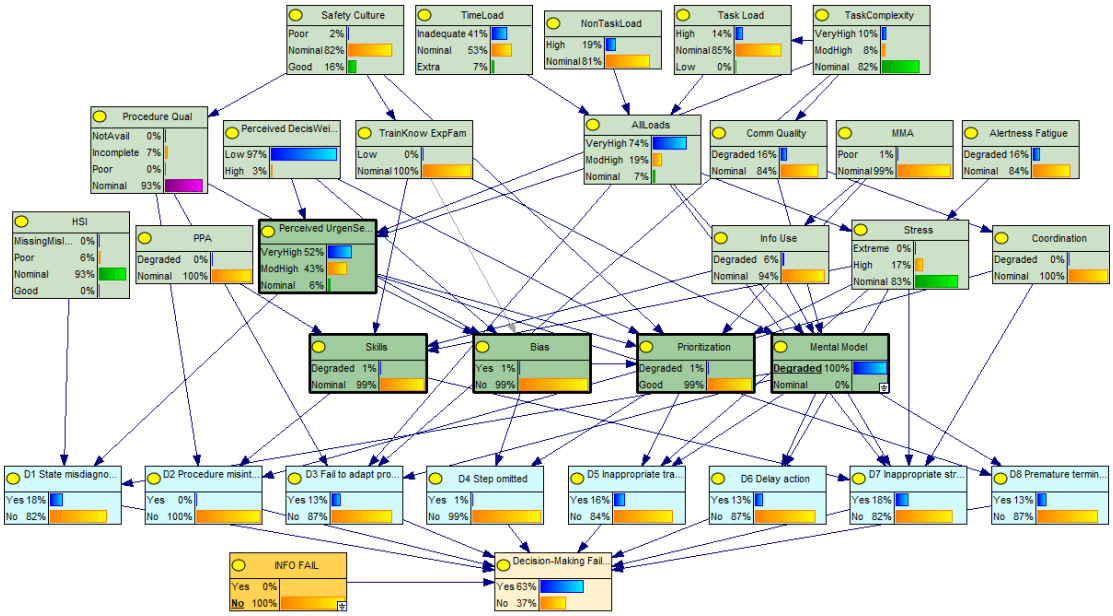

Figure 4.4: Decision-making BN with evidence set: Mental Model = Degraded.

4.5 Validation: Empirical Data from Operational Experience

This section describes the process by which the suite of BN models (information, decision, and action) were validated and refined against operational data. This validation task was essential in determining the models' quantification capabilities for common operational scenarios that may arise.

4.5.1 Available Data: German Nuclear Power Plant Operating Experience

Operational experience of nuclear power plants in Germany provided empirical data for use in model validation. The project was conducted by Gesellschaft für Anlagen- und Reaktor-

sicherheit (GRS) using reports from the German licensee event report system, and consists of 67 data items covering a variety of control room tasks [115, 116]. Error occurrences were directly extracted from the licensee event report text.

The authors of the study, Preischl and Hellmich, noted some challenges and listed the requisite solutions with extracting human reliability data from licensee event reports. First, not every human error that occurs necessarily meets the reporting threshold, and so they could not include any error that did not lead to detectable adverse consequences. Second, the reports do not include how frequently any given task is performed, which would provide the denominator of the error rate. To address this, they estimated the frequency of performance from the plant's operational history and regulatory standards. Third, the relevant contextual factors are not always fully reported, so investigations and interviews were conducted with plant personnel to gain further insight into the error conditions as necessary.

The data points are presented as an m_i/n_i error fraction, with m_i being the number of errors occurring for a given task *i*, and n_i being the number of times task *i* was attempted. For each sample point, the generic task and the associated error is described, e.g. "Operating a key control-wrong control selected." Then, the authors determined the relevant performance shaping factors, with a focus on those factors that might increase the likelihood of error, e.g. "Moderately high level of stress" or "Similar buttons nearby." Some factors that would be expected to decrease the error probability are included, but these are fewer. In addition to the error description and failure probability, they apply Bayesian updating methods to obtain a 90% uncertainty interval. The probability of error is assumed to follow a binomial distribution, while the binomial parameter is assumed to follow a conjugate prior beta distribution. They use the Jeffreys prior for the beta distribution parameters and update with the observed data to obtain a 5^{th} and 95^{th} quantile for

the estimated HEP. This follows Equations 4.1 and 4.2, where α' and β' are the Jeffreys prior estimates, and α and β are the posterior parameter estimates for the beta distribution. The values m_i and n_i are the number of errors and the number of task attempts, respectively.

$$\alpha = \alpha' + m_i \tag{4.1}$$

$$\beta = \beta' + n_i - m_i \tag{4.2}$$

The resulting database covers a variety of human errors in the control room, mainly corresponding to the incorrect or inaccurate manual actuation of controls while following operating procedures. With thorough and rigorous data collection methods despite the retrospective nature of the analysis, this data is ecologically valid to the real-world operation of nuclear power plants. However, significant mapping and preprocessing was necessary before it could be applied to our HRA framework, which will be described in the following section.

4.5.2 Data Mapping and Cleaning

As the GRS data collection method did not follow the Information-Decision-Action framework [19], the Phoenix HRA method's CFMs [10], or Groth's 2012 PIF taxonomy [24], it was necessary to map the GRS error descriptions to these methods so that the data could be directly compared to the Bayesian networks' outputs. A similar mapping had previously been performed on the GRS data to quantify the Phoenix method, with the errors mapped to Phoenix CFMs and the situational description mapped to a single high-level Phoenix PIF [23]. The existing mapping has been augmented here to incorporate GRS data published since the initial development

of Phoenix, and the PIF identification has been refined to allow the identification of multiple PIFs from the more granular 2012 taxonomy [24].

Several values were screened from the analysis. In particular, some rarely performed task samples were deemed too small to be of statistical significance. Tasks for which there was a zero-failure rate or a 100% failure rate were also not considered here. The models do not support a deterministic view of the CFMs; there are no conditions under which a CFM's probability of occurrence is precisely zero or one. Lastly, several samples with dynamic PIF effects (such as poor communications) were excluded from this analysis. It cannot be determined whether the PIF was in effect for all of the task attempts, or merely the few task attempts that resulted in errors. Each assumption could respectively result in a significant overestimate or underestimate of the task's error probability. In total, there were 53 data points retained for consideration.

4.5.3 Validation Process and Results

To validate the BN models' quantitative output, model estimates were generated for each data point under the given error conditions, then compared to the provided empirical values and their uncertainty bounds. This was automated using a Python script with the PySMILE package [92] which enables users to directly interface with Bayesian models through the GeNIe software. For each data sample, the relevant PIF states were set as evidence, an updating procedure was conducted, and the requisite CFM probability was extracted.

The resulting CFM probabilities were compared against the GRS empirical probabilities on a log-scale, shown in Figure 4.5. As the values span multiple orders of magnitude ranging from 10^{-1} to 10^{-5}, the log-scale was the most effective way to visualize the spread of values and the

relative differences between the model outputs and the data. In all cases where the BN model's value was greater than the GRS value, this difference was less than an order of magnitude. Similarly, for all cases where the BN value was less than the GRS value, this difference was less than two orders of magnitude. Generally, the BN predictions tended to be lower probabilities than the GRS values.

Next, the GRS-provided uncertainty bounds were considered as a benchmark for the BN models' performance. Overall, most of the BN estimates fell within or close to the bounds. The GRS probabilities and associated uncertainty bounds are plotted alongside the BN estimates in Figure 4.5. Many of the decision-making and action execution BN estimates have the same numerical value, as the same PIF/CFM combinations were found to be applicable for certain circumstances. Furthermore, it can be seen that the information-gathering and decision-making errors are observed and estimated at a higher error rate than action execution errors. This corroborates widely held HRA assumptions that cognitive tasks are more susceptible to error than manual action tasks, reflected in SPAR-H's base HEPs of 0.01 for Diagnosis and 0.001 for Action.

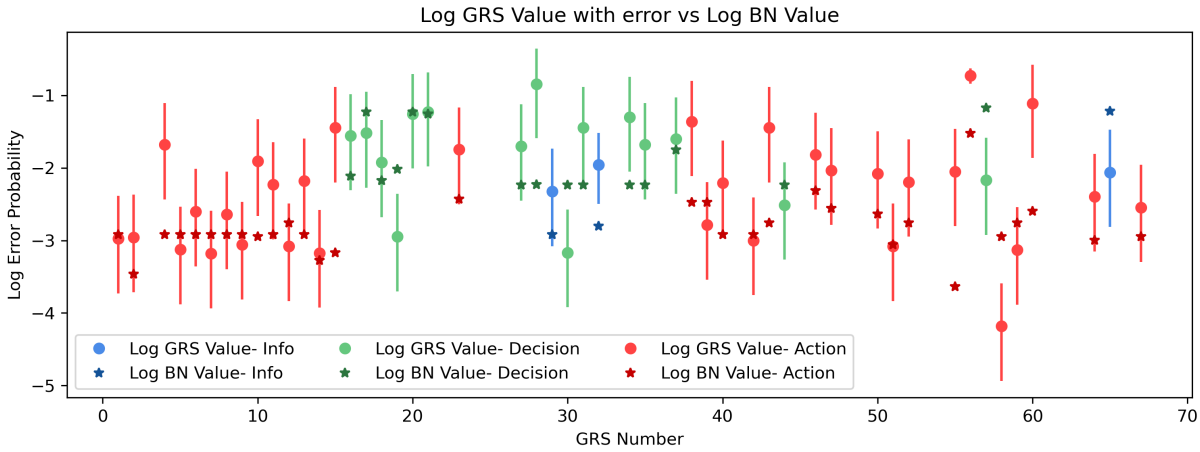

Figure 4.5: GRS data values with provided error bars from [115, 116] compared against BN models' outputs.

4.6 Discussion and Conclusion

This human performance model for operator decision-making is founded in causal logic, reliability engineering, and psychological theory. Using literature documented in NUREG-2114 [32] as a starting point, additional literature was sought out to identify and justify causally related variables. This improves upon current HRA models, which primarily rely on expert opinion to justify relationships between factors, and sometimes do not explicitly declare these relationships at all.

During the development of this network, five human failure mechanisms were identified that can lead to decision-making failure events. Analogous to mechanical failure mechanisms, these human failure mechanisms answer the question of *why* human failure events occur. Crew failure modes describe the observable manner in which (*how*) failure occurs, while mechanisms describe the underlying reasons for failure. Quantitatively, failure mechanisms were found to be by far the strongest causes of the failure modes. Setting evidence on the networks that a mechanism had occurred can increase the overall MCF failure probability by an order of magnitude or more. A detailed discussion of human failure mechanisms is available in [33].

The mechanisms of *Prioritization*, *Bias*, and *Mental Model* were found to be particularly complex in their causation. Prioritization and Bias are colored by the individual's perception of the situation, which is in turn caused by a mixture of personal, situational, and organizational PIFs. Similarly, Mental Model quality is impacted by the individual's knowledge base, the complexity of the scenario at hand, and their willingness to comprehend the situation. The composite origins of these mechanisms necessitates incorporating a wide variety of contextual factors into the model. These mechanisms are distinct from the information-gathering or action execution

mechanisms, as decision-making is a more internal cognitive process and the other phases require interfacing with the external world. Some similarities do exist however in the causal origins of the information-gathering bias and prioritization failure mechanisms.

Model quantitative outputs agree well with conventional HRA knowledge and the empirical GRS data. Decision-making and information-gathering failures are estimated to be more likely than action execution failures, which corroborates the common HRA assumption that cognitive tasks are more error prone. This also reflects the data-captured reality that action execution tasks can typically happen more times than decision-making tasks without errors occurring.

Several recommendations can be made for future development of comprehensive human error databases such as SACADA. SACADA was used extensively during the model parameterization process, due to the multitude of relevant decision-making variables provided. Many information-gathering and action execution parameters were collected, but the selection of decision making PIFs and CFMs was more complete to these BN models. It should also be acknowledged that SACADA is a failure-biased dataset, as the training scenarios are intended to be difficult, and the data is collected during or before the team's retraining.

Additional data collection in SACADA could benefit from a more extensive use of overarching PIFs that cover factors common to all macrocognitive phases. PIFs such as safety culture, personal and physical abilities, or training/knowledge/experience/familiarity would also be beneficial to assess. SACADA takes an adversarial view of these training scenarios, assuming a static competent crew pitted against tasks of varying difficulty. SACADA has many granular factors representing ways the tasks can be complicated, but very few factors relating to inherent issues with the team. When team-based and person-based factors are collected, they are categorized as error modes and causes, which leaves little room to study the defective yet mostly successful

teams which are an unfortunate operational reality.

Incorporating the collection of organization-based factors may be challenging for several reasons. First, if the training simulator staff have agreed to engage in data collection efforts, it may be an unwanted addition to their duties, and result in lower overall data quality. Second, organizations may not be willing to admit their own weaknesses in this area. Third, they may not be able to collect data on these factors due to being unaware of these deficiencies. However, these remain important issues to resolve. Model predictive capabilities could be greatly improved by collecting data on this greater breadth of PIFs.

In conclusion, the HRA models presented in this work are traceable and comprehensive, providing a full causal picture of the factors which may lead to HFEs. Bayesian networks provide a flexible model structure that can accommodate additional data as it becomes available, and a multipurpose tool that can be leveraged for many different types of analysis. These attributes are all integral to improving the next generation of HRA methods [6]. The methods presented here have distinct advantages over those in current use to meet regulatory requirements. It is our intent to incorporate the models into the Phoenix HRA method in the near future, and this combined model will soon be available for regulatory use in the nuclear power domain.

Chapter 5: Causal Pathways Leading to Human Failure Events in Action Execution System Response Activities

This chapter is reproduced from [100] and [117].

Camille S. Levine, Ahmad Al-Douri, and Katrina M. Groth. "Causal Pathways Leading to Decision and Action Human Failure Events: Structures and Validation." In: *Proceedings of the 17th Probabilistic Safety Assessment and Management Conference (PSAM17) & Asian Symposium on Risk Assessment and Management (ASRAM2024)* (Oct. 7–11, 2024). Sendai, Japan.

[Submitted, in press]: Camille S. Levine and Katrina M. Groth. "Three Novel Causal Logic-Informed Models for Identifying and Quantifying Operator Errors, Part III: Action Execution." In: *Reliability Engineering & System Safety* (2025).

5.1 Introduction

HRA seeks to characterize the causes and quantify the likelihood of HFEs. In HRA practice, there is a present need for cognitively realistic methods that acknowledge the multitude of causal pathways leading to HFEs. Equally important is the implementation of a quantification scheme rooted in data and observation that is feasible for analysts to conduct. In this work, we

present the development and parameterization of a human performance model for action execution tasks that seeks to address these needs.

Action execution is the third and last macrocognitive function of the Information-Decision-Action framework, and a crucial phase of operations. Even if the operator team has perfect situational awareness and correct goals, an error during action execution can result in a human failure event without the possibility for recovery. However, unlike the functions of information-gathering and decision-making, there are relatively few ways that action failures may manifest. Although there are an infinite number of wrong actions that an operator can feasibly take, there are limited types of action errors. Traditionally, human errors are classified into two categories: errors of omission (EOO) and errors of commission (EOC), corresponding respectively to no action, or incorrect action. EOC may be further categorized, for example, into selection, time, sequence, or qualitative errors [15].

In other HRA models, action errors are typically either modeled as flat probabilities for given task types (e.g., THERP [15]) or as modified probabilities given a set of performance influencing factor (PIF) states (e.g., SPAR-H [1] or CBDT [118]). The former modeling choice benefits from being easy to quantify by identifying task characteristics, and the latter modeling choice allows the analyst to take contextual factors into consideration. However, neither model style paints a complete picture of the causes of errors in action execution. This work employs the Information-Decision-Action in Crew Context (IDAC) [19] and Phoenix HRA methods [10] alongside Groth and Mosleh's 2012 PIF taxonomy [24] to build novel causal models for action execution that describe causal pathways to action HFEs through multiple layers of influence.

A wide variety of PIFs, such as tool availability and quality, human-system interface (HSI), and external environmental conditions are integral to the successful execution of operator actions.

For a human-machine team to function properly, first and foremost the physical environment must be suitable. Hazardous environmental condition and poor ergonomics present significant barriers to human action. In addition to a conducive work environment, the operators must be effective team members and actors in their own right. Unclear team roles and inability to focus on carrying out the task at hand may lead to improper or delayed actions [83, 119].

While action execution is simpler than the information-gathering and decision-making functions, there remains a need for rigorous, literature-based causal models for action-taking similar to those presented previously in Chapters 3 and 4. The rest of this chapter details the development, structure, parameterization, and validation of a Bayesian network modeling causal pathways leading to action execution errors.

5.2 Method: Model Structure Development

We created three Bayesian networks for information-gathering, decision-making, and action execution, intended to match the three cognitive phases of the IDAC framework [19]. IDAC defines these high-level functions as macrocognitve phases. Major crew functions, or MCFs, are specific instantiations of these high-level functions that only occur within a single cognitive phase. The Phoenix HRA method then defines three action execution Crew Failure Modes (CFMs) that describe how operator action errors may manifest [10]. These CFMs are designed to exhaustively cover all of the potential failure modes that may occur in the macrocognitive phase of action execution- both errors of comission and omission. The complete method of model development is presented in Chapter 3 and summarized in 5.1.

In work preceding the development of these models [22], and expanded upon in Chapter

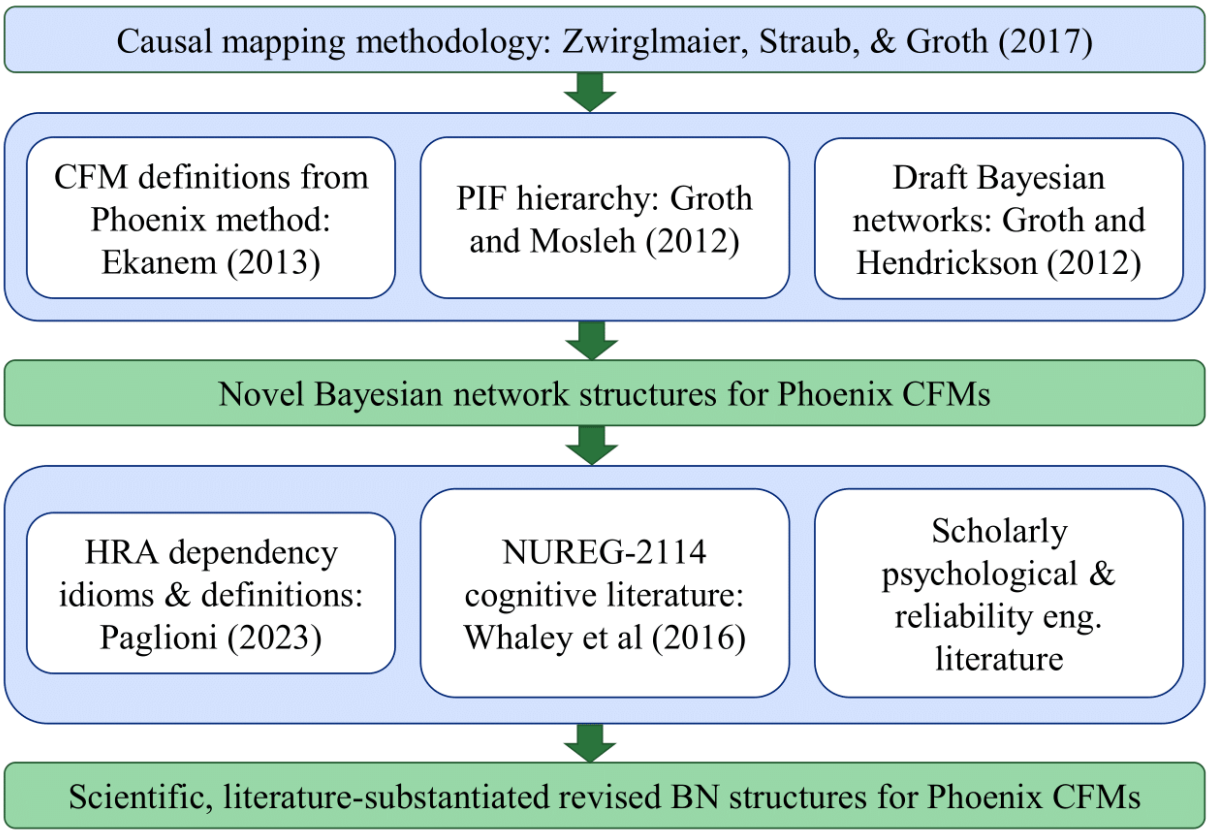

Figure 5.1: Sources used in the development of human performance Bayesian networks.

6, we used task decompositions from NUREG-7256 [60] to identify potential human failure events that may occur during external flood hazard mitigation actions. During the identification of potential human failure events and their requisite failure modes, a fourth action-taking CFM was posited: A4, "No Action Taken."

The CFM A4, "No Action Taken", was introduced to fill a gap identified after applying the Phoenix HRA method to ex-control room actions. Although the Phoenix CFM A1 "Incorrect Timing of Action" also covers failure scenarios where crews forget altogether to take the required action (i.e., infinite delay time), it is useful to make a distinction between a delay and an action that will not be taken. This could be due to lack of prepared tools available, or the mechanisms of attention [114] and coordination [119]. "No Action Taken" refers to a scenario in which the

correct information is gathered and the correct decision is made, but the actions are simply not carried out. For example, a crew might go to pick up equipment from an external location, but the vehicle is out of fuel (lack of resource availability), or the task is not properly assigned to an individual so it is erroneously assumed to be done, but never ends up being completed (poor coordination). Additionally, the individual operator may simply forget that the task was assigned due to extreme loads and stressful circumstances surrounding the situation (attentional error).

Groth and Mosleh's PIF taxonomy [24] allowed for identification of PIFs that are relevant to each CFM. The full taxonomy is shown in Table 3.1. It categorizes PIFs into machine-, person-, team-, organization-, and situation/stressor-based factors, and also organizes these PIFs into hierarchical levels of specificity. Then, we applied the Phoenix method's CFM definitions [23] to identify broader mechanisms of failure, or the cognitive/physical reasons for the CFM's occurrence. These mechanisms were frequently a high-level PIF from the taxonomy, or representative of a cluster of PIFs. Finally, we selected specific PIFs relevant to each failure mode, and inferred causality between the factors and mechanisms according to the method described in [14].

After developing the preliminary model structure, literature extracted from [32] was used as a starting point to find additional human factors literature focused on action execution. After thorough examination of the literature, the model structure was slightly revised to better reflect the cognitive and physical processes underlying operator actions. Then, each arc in the network was substantiated with an excerpt from the literature resources. Some of the arcs are deterministic, or definitional, in nature, such as task complexity's direct effect on subsequent task load. This represents an instance of the Definitional HRA idiom, further described in [35].

The literature used to substantiate the action execution Bayesian network is documented

in Table 5.1. Huey & Wickens' work in human factors was used to substantiate a wide variety of connections, particularly the connections leading directly into the CFMs, as well as the arcs leading into and out of the *Attention* mechanism [120]. Like the information-phase and decision-phase networks, Letsky et al.'s work [74] in team-based factors was instrumental in substantiating the arcs leading into the *Coordination* mechanism, and Reason's work was used to substantiate the effects of safety culture [103].

5.3 Result: Model Structure

The full Bayesian network structure is shown in Figure 5.2. A source substantiating each arc in psychological literature is indicated by a corresponding letter on the arc matching that in Table 5.1. Some of the structures are substantiated through the presence of definitional idioms from Paglioni and Groth [35]. For example, if the operator is under extreme time load, with insufficient time to complete actions, they are by definition overloaded. Therefore, one can say that a degraded state in the parent node of Time Load defines the presence of a degraded state in the child node of All Loads.

In addition to the CFMs comprising the lower layer of the network, a failed action execution MCF may also be the result of an incorrect decision. A prior decision-making failure may go unnoticed until it has manifested as the incorrect actions. This is encoded as a deterministic relationship between the "DECISION FAIL" node and the "Action Execution Failure" node: if the crew has made the wrong decision, it is certain that their actions will also be wrong in some fashion.

Ref on BN	Ref Number	Citation	Area on BN
a)	[121]	Proctor, R. & Van Zandt, T. (2017).	PPA, environment, loads, HSI
b)	[83]	Wickens, C. et al. (2004).	Attention, coordination
c)	[74]	Letsky, M. et al. (2008).	Coordination
d)	[122]	Hsieh, S., Tsai, C., & Tsai, L. (2009).	Fatigue
e)	[123]	Altmann, E. & Trafton, J. G. (2002).	Knowledge base
f)	[73]	Endsley, M. (1995).	Training
g)	[81]	Orasanu, J. & Martin, L. (1996).	Training
h)	[103]	Reason, J. (2000).	Safety culture
i)	[114]	Wickens, C. et al. (2012).	Attention, HSI, environment
j)	[120]	Huey, B. & Wickens, C. (1993).	Attention, procedure error, skills, task-related PIFs
k)	[119]	Poozhithara, J. et al. (2023).	Coordination
l)	[35]	Paglioni, V. P. & Groth, K. M. (2024).	Deterministic dependencies

Table 5.1: References used to substantiate Action-taking network model arcs.

5.3.1 Model Structure: Human Failure Mechanisms

There are four key causal clusters apparent in the human performance model for action execution. Each cluster corresponds to a mechanistic PIF, denoted by the darker green nodes with thicker borders. These clusters are *Coordination*, *Procedure Error*, *Skills*, and *Attention*. It is important to note that this type of attention draws upon different mental resources than the attention relevant to information-gathering. Rather than detecting informational cues in the environment, attention in the action execution phase refers to the operator's ability to focus on carrying out the task at hand. The mechanisms documented here are probabilistic in their causation of error, and increase the probability of occurrence of a CFM. Multiple mechanisms can contribute to the likelihood of a particular CFM with differing degrees of influence.

The first mechanism cluster, *coordination*, directly causes two CFMs (A1 and A4). Co-

ordination errors are directly caused by four PIFs (cohesion, role awareness, communication availability and quality, and safety culture). The mechanism is not otherwise related to any additional factors. Despite the limited number of PIFs comprising it, coordination is a complicated mechanism to model because it relies on the individuals' ability to harmonize their actions as well as the organization's conduciveness to teamwork.

The second mechanism cluster, *procedure error*, directly causes two CFMs (A1 and A2). Procedure error is directly caused by three PIFs (procedure quality, training/knowledge/experience/familiarity, and HSI). Safety culture may indirectly cause this mechanism through issues with the training program. This failure mechanism is particularly likely to happen when a poorly trained operator does not have supportive procedures to fall back on, or when procedures are insufficient and the operator has not received supplemental training on the situation.

The third mechanism cluster, *skills*, directly causes two CFMs (A1 and A2). Skill errors are directly caused by two PIFs (training/knowledge/experience/familiarity, and personal and physical abilities [PPA]). Similar to the causes of procedure error, a poor safety culture may lead to a poor training program. Issues arise with skills when the operator has incomplete knowledge of what needs to be done, or they are physically unable to carry out the actions.

The fourth and last mechanism cluster, *attention*, directly causes all four of the CFMs. Attention errors are directly caused by five PIFs (external environmental conditions, alertness & fatigue, stress, overloading, and perceived urgency). Indirectly, problems with the external environment may cause stress to the operator. In addition, overloading can be caused by a mixture of factors, including high task load, non-task load, time load, and task complexity. These factors can cause lapses in the operator's executive attention towards completing the task at hand.

Some of the structures in this network represent an effect-modulating dependency between

factors, as defined in [35]. For example, attentional errors are caused by the presence of degraded environmental conditions, stress, and high loads, and when more than one of these conditions is present, the resulting increase in the likelihood of error is greater than the linear sum of the PIFs' effects. This captures the notion that environmental stressors may exacerbate the condition of an operator who is already stressed or overloaded from other causes.

The network shown in Figure 5.2 captures both the direct and indirect causal influences on the Action CFMs. Due to the more direct nature of action execution, fewer indirect causes are present in this network than in the information-gathering and decision-making models. However, it is essential that each causal pathway leading to human failure events is identified. Documenting and substantiating each modeling choice in human factors literature ensures a causally robust model that captures a wide range of factors influencing human performance. Additionally, the model captures several different kinds of dependency between PIFs, mechanisms, and CFMs, which were previously not addressed in the context of HRA modeling.

5.3.2 Model Structure: Performance Influencing Factors

A layer of PIFs causing the human failure mechanisms is shown in lighter green, capturing additional complexities of the scenario context. Some PIFs have definitional relationships with the CFMs, such as a lack of tool availability causing CFM A1, "Incorrect timing," and poor tool quality causing CFM A2, "Incorrect operation." Other PIFs exert a direct causal influence on the CFMs, such as stress affecting the likelihood of CFM A2 occurring.

Like the other two BN models for information and decision, there is a structure in this network comprising all loads to the operator. This includes task load (affected by task complexity),

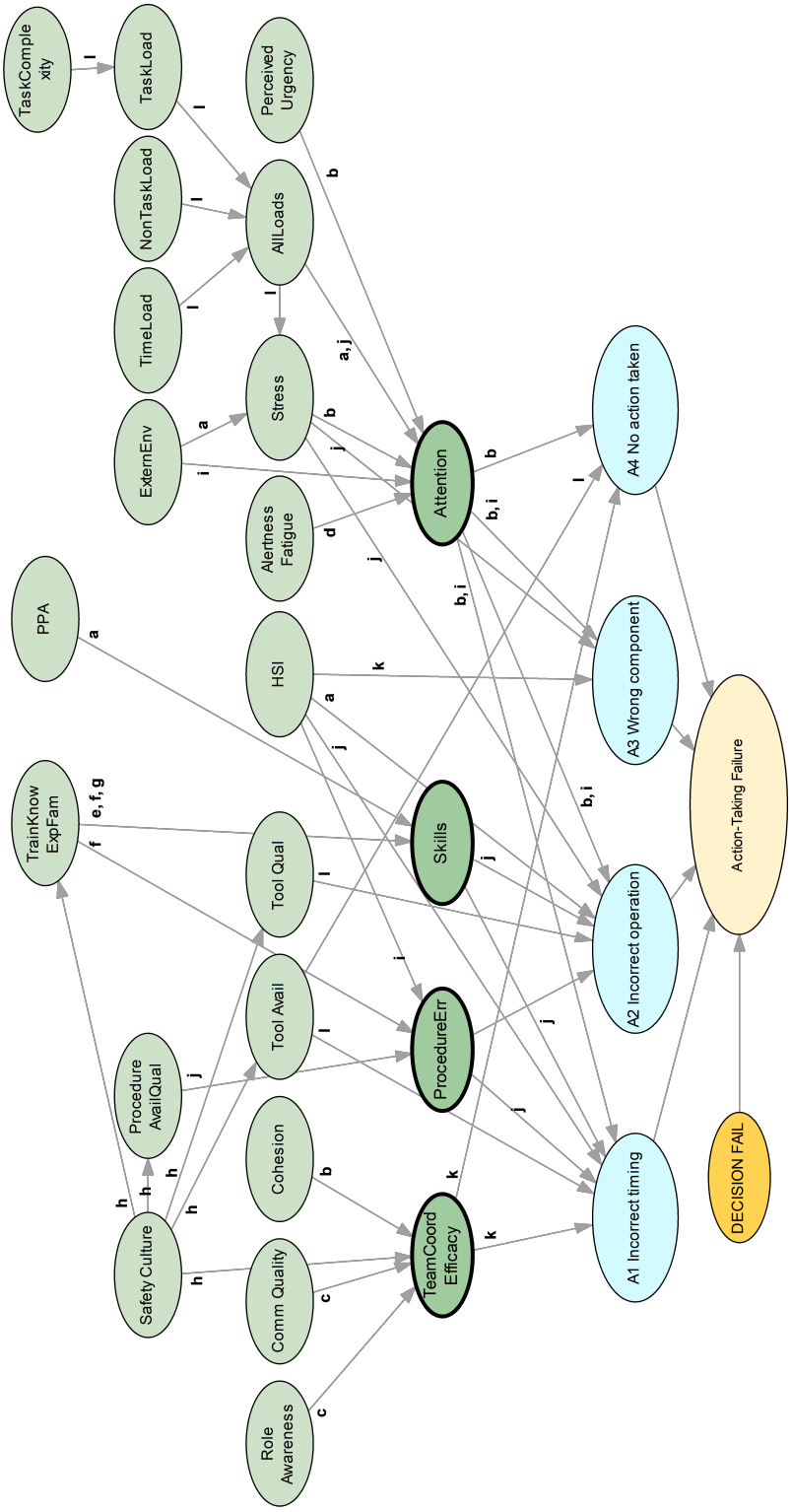

Figure 5.2: Full cognitive literature-substantiated Bayesian network structure for the Phoenix method's A-phase CFMs.

non-task load, and time load. Unlike the information-gathering network, HSI and staffing do not present loads to the operator during action execution.

Parallel to the effects of their various loads, the operator's perceived urgency of the situation may impact their ability to apply executive attention towards the task. Their levels of fatigue may also impact their ability to pay attention. Degraded environmental conditions, particularly during external hazard scenarios, may pose significant loads and extreme stress to operators executing actions.

Another large PIF structure relates to the organization's safety culture as well as various team-based factors. A poor safety culture may result in a poor training program, lack of tool availability or tool quality, and overall poor coordination. These broad effects underscore the importance of maintaining a good safety culture to create an environment conducive to operator actions. In addition to safety culture, poor communication, cohesion, and team role awareness may result in poor coordination. Coordination requires balanced team dynamics and a definitive leadership structure: without either of these conditions, the team will be ineffective at implementing the necessary actions.

5.4 Method: Model Parameterization

After the BN model structure was finalized and all arcs were substantiated with psychological and organizational literature, we parameterized the model so that it could be applied to quantitative analyses. Many HRA methods in current use do not employ both a strong theoretical basis and a straightforward quantification method [25]. Additional long-standing opportunities in the HRA field are the effective treatment of errors of commission and ex-control room actions.

Because of these gaps, it was crucial to develop a transparent, usable quantification process alongside the model structures, so that the models could be effectively applied to quantitative analyses in these areas. A more detailed explanation of this parameterization method is presented in Chapter 3.

Bayesian networks were chosen for their unique capacity to combine many distinct data sources with varying degrees of confidence. Different data sources may be combined to parameterize a single node. The main data sources used in this network were the U.S. NRC's SACADA database [87], NUREG/CR-6949 Bayesian priors [84], and IDHEAS-DATA [90] and IDHEAS-ECA multipliers [17]. IDHEAS-DATA was used as a starting point to find additional quantitative psychological literature. Lastly, other HRA methods such as Phoenix [10], CREAM [16], and SPAR-H [1] were applied to nodes that could not be parameterized by any other source.

To combine sources that covered the same nodes in the network, we employed Bayesian updating with a multinomial-Dirichlet conjugate distribution. The multinomial-Dirichlet is the multi-state generalization of the binomial-beta conjugate distribution, which is frequently applied to binary success/failure data. Applying Bayesian updating effectively accounted for the confidence in each data source, placing more importance on those sources with more data points. The network is comprised entirely of discrete nodes with 2, 3, or 4 states, so the multinomial distribution accommodated the structures well.

In many cases, it was necessary to derive multipliers from the SACADA database and psychological literature, or apply the multipliers provided in IDHEAS or in the other HRA methods. The multipliers were mainly applied to the larger conditional probability tables, which had far more combinations of node states than were covered directly by the data. Lastly, several nodes were partially parameterized using the definitional HRA dependency idiom described in [35],

which deterministically defines the state of a child node in the network based on the state of the parent node.

The effect-modulating dependency idiom from [35] was applied to several PIF combinations that were found to be especially error-prone. In particular:

- Procedure error is almost certain to occur when the procedure quality is degraded and the team has no knowledge base to fall back upon, and vice versa.
- Degraded attention is almost certain to occur when the team is under simultaneous high stress and high loads.
- CFMs A1, A2, and A3 are more likely to occur when the HSI is poor and the attention is also degraded.

Multipliers were applied to the conditional probability tables to better reflect the reality of these compounding factor effects.

However, effect-modulating is not present to the degree that one might expect between certain variables. For example, one might expect a degraded physical environment and mental fatigue to synergistically compound to make an even worse condition for operators. Instead, one study found that both cold/heat and mental fatigue have significant negative effects on dexterity, but their combined effect is not worse than the effect of mental fatigue alone [124]. As such, the multipliers were judiciously applied to only those scenarios known to be more conducive to errors according to the operational data provided in [115, 116].

The model structure was developed in the GeNIe software and parameterized using the related Python wrapper, PySMILE [92]. A Python script was used to automate data cleaning,

extraction, and parameterization, and also streamlined the model validation exercises described later in this chapter.

5.5 Result: Model Quantification Capability

The fully parameterized network for action execution is shown in Figure 5.3. A decision-making failure node encodes the possibility that a prior cognitive failure manifests as an action execution MCF failure. Evidence has been set on this node that a prior decision-making failure has not occurred, so that the probabilities of pure action execution failures can be illustrated. Because the GeNIe software does not display probabilities below a precision of 1%, the probabilities of each CFM with no evidence on the network are shown in Table 5.2.

One benefit of using Bayesian networks is that any amount of evidence about a situation can be incorporated into the analysis. Evidence can be set on PIF nodes in order to reason forward about contextual effects on CFM probabilities, or itcan be set on CFM nodes to reason backward and diagnose causes. Similarly, evidence can also be set on mechanism nodes to determine their relative contribution to CFM probabilities and identify the causal PIF clusters that they represent. It is important to note that this analysis can be conducted retrospectively, to gain a fuller causal picture of events that have transpired, or prospectively, to analyze future events and to better understand potential risks and their causes.

Without evidence on the network, many of the CFM and failure mechanism probabilities are quite low, with CFM probabilities on the order of 10^{-3} or less. Likewise, the probabilities of degraded coordination, skills, or attention are 2.7×10^{-3}, 1.4×10^{-3}, and 4.4×10^{-3}, respectively. The highest likelihood of any mechanism is procedure error, at 1.7×10^{-2}. The PIFs are in a

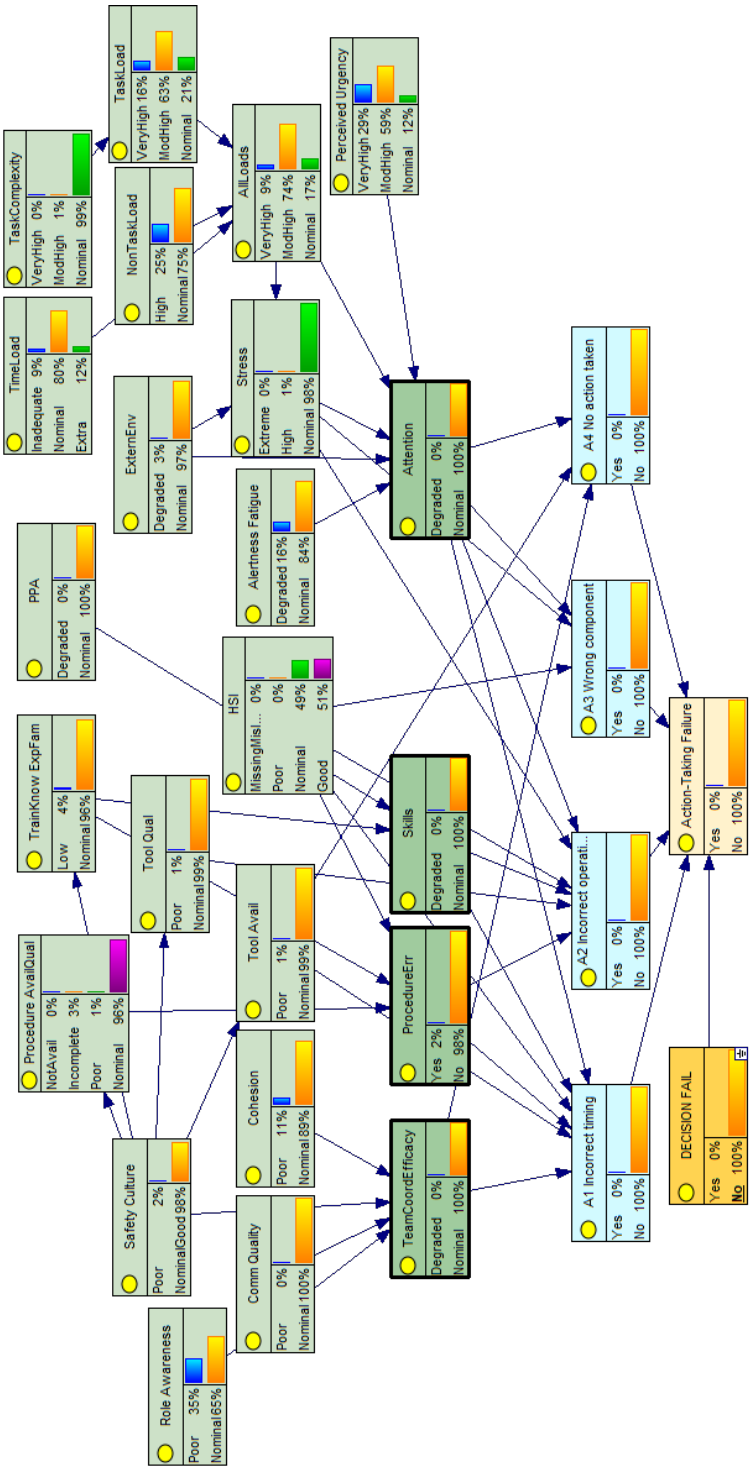

Figure 5.3: Fully parameterized Bayesian network model for action execution failure events, with more detailed CFM probabilities in Table 5.2.

CFM	Probability of Occurrence
A1: Incorrect Timing	5.3×10^{-4}
A2: Incorrect Operation	3.4×10^{-4}
A3: Wrong Component Selected	2.3×10^{-4}
A4*: No Action Taken	1.0×10^{-3}

Table 5.2: Probability of each action execution CFM occurring when beliefs are propagated through the network with no evidence set.

nominal or good state more often than not, particularly machine- and organization-based factors. Generally, there is no reason to believe that there is deficient HSI, tools, procedures, or safety culture until it has been demonstrated otherwise. However, certain PIFs, such as fatigue, role awareness, cohesion, and non-task load are somewhat likely to be degraded. This is an artifact of how large organizations such as nuclear power plants operate, with long overnight shifts and individual workers that are frequently unaware of their role in overall plant safety.

In Chapters 3 and 4, we showcase the models' predictive ability to obtain changes to CFM probabilities when evidence is set on PIFs, and the models' combined predictive and diagnostic capability to examine causes and effects when evidence is set on a failure mechanism. Here, we expand upon the models' diagnostic power by setting evidence that the CFM node "A2 Incorrect operation" has happened. The network with this evidence and the subsequently updated beliefs is shown in Figure 5.4.

Immediately, it can be observed that the failure mechanisms of procedure error and skills increase in likelihood to 0.25 and 0.16, respectively. In turn, because operators' knowledge bases strongly affects these mechanisms, the probability of poor training/knowledge/experience/ familiarity increases from 0.036 to 0.26. The probability of poor safety culture also skyrockets from 0.023 to 0.25, and the probability of procedures being in a less-than-adequate state increases from 0.043 to 0.27. The likelihood of tool availability being poor increases to 0.064, and poor

tool quality increases to 0.068. To a lesser extent, the mechanism of degraded attention is also affected, increasing to 0.019. This smaller effect is because the CFM A2 is primarily caused by poor skills and procedural understanding. Likewise, the PIF states causing poor attention also increase in likelihood by a small amount, such as extreme stress increasing from 0.0022 to 0.015 and high stress increasing from 0.014 to 0.022, but these factors are less affected than the PIFs which cause degraded skills and procedural error.

To develop a mitigation program for errors of incorrect operation, one would first discern which mechanism is the main cause. It is more likely that the error is caused by procedure error or insufficient skills, so in this case it would be important to ensure that the organization's safety culture is adequate or exemplary. Since poor safety culture is a main direct cause of procedure availability and quality, operator knowledge base, and tool availability, improving the safety culture should be the first step towards ensuring the operators have the necessary resources to complete their work. Poor or missing HSI elements may also lead to procedural errors, so this factor should be examined and improved if necessary. Additionally, poor PPA is a significant cause of skill errors. If the operators lack the necessary abilities to complete their job, one should hire additional individuals who are more able and reassign the previous operators. In the less likely case that the CFM is caused by attentional slips, it would be important to discern whether the operator is overloaded, fatigued, stressed, affected by external environmental conditions, or has an incorrect perception of the situation's urgency. The deficient PIF or PIFs causing issues with executive attention can then be addressed by the organization.

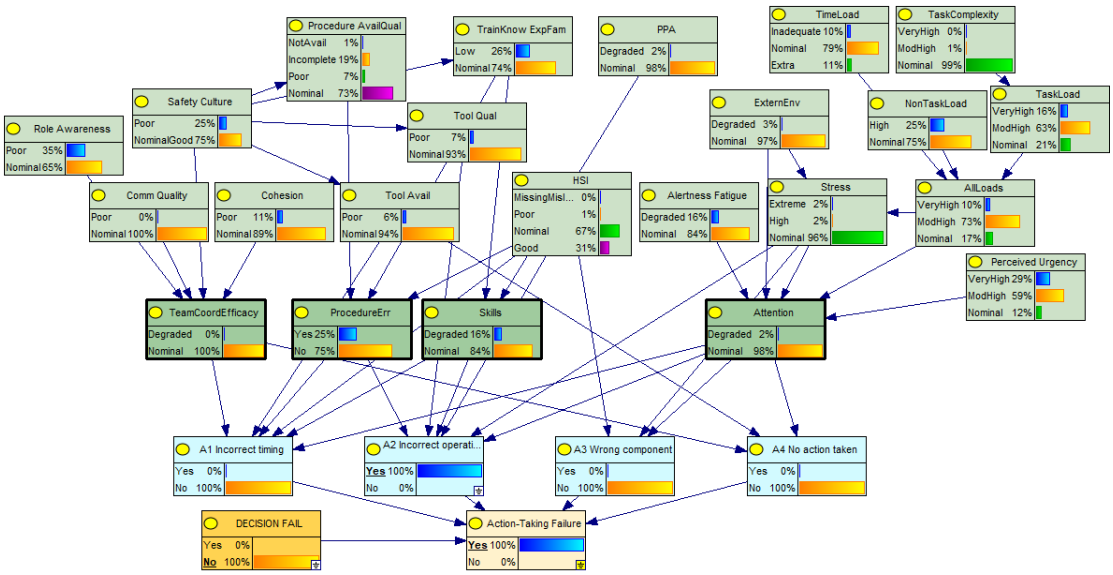

Figure 5.4: Action execution BN with evidence set: CFM A2 Incorrect Operation = Yes.

5.6 Validation: Application to U.S. HRA Empirical Study

This section details the data and scenarios from the U.S. HRA Empirical Study [125] and their application to a summative validation of the BN models. As an extension of previous validation of the model structures against pre-analyzed ATHEANA narratives in Chapter 3 and the model quantification output against German operational experience data [115, 116] in Chapter 4, this validation exercise demonstrates the models' applicability to several scenarios and compares the models' outputs to other HRA methods.

5.6.1 Empirical Study Scenarios and Data Collection

The U.S. HRA Empirical Study collected performance data on four crews completing several pre-determined scenarios in a PWR training simulator. These scenarios included 1) a loss of feedwater (LOFW) followed by a steam generator tube rupture (SGTR), 2) a loss of compo-

nent cooling water (CCW) and reactor coolant pump (RCP) sealwater, and 3) a simple SGTR event. Data was collected on the crew's performance and the timing of their actions through several channels, including experimental observers, simulator logs, video recordings, debriefing interviews, and a questionnaire to assess important contextual factors [125]. For each scenario, a timeline of the main actions is provided, as well as an assessment of the relevant PIFs and their effects on the crews.

The first scenario, a LOFW followed by a SGTR, begins with all feedwater pumps tripping while all but one auxiliary feedwater pump are unavailable. The crew should manually trip the reactor, but it will automatically trip if they do not act. Then, an incorrectly positioned valve prevents water from this auxiliary pump from reaching the steam generators. The crew must realize this and establish feed & bleed to remove decay heat. The event HFE 1A occurs if the crew cannot do this within 45 minutes, assuming they have manually tripped the reactor, and HFE 1B occurs if they cannot do this within 13 minutes, if the reactor trip was automatic. If the crew succeeds, the simulation continues with a SGTR. The crew must isolate the ruptured steam generator and maintain its pressure below a setpoint. If they fail to do this, HFE 1C has occurred.

The second, and most challenging, scenario consists of a loss of CCW and RCP sealwater. The scenario starts with a failed distribution panel, which forces the crew to establish manual control of several channels. Then, the reactor trips due to high levels in one of the steam generators. Faults elsewhere in the CCW system will render all pumps out of service. The crew must realize this has happened and trip the pumps. They must then start the positive displacement pump (PDP) before the rising RCP seal temperatures render this task too dangerous. If they fail to trip the RCPs within one minute of losing CCW, or fail to start the PDP within about seven to nine minutes, then HFE 2A has occurred.

Finally, the third scenario consists of a standard SGTR event. The crew must isolate the ruptured steam generator and maintain its pressure below a setpoint, similar to the functions in HFE 1C. No other mechanical or electrical failures occur that might complicate or delay their understanding of the event. If they fail to do this, HFE 3A occurs.

5.6.2 Mapping of Scenario Progression and Context

To assess the probability associated with each HFE occurring, it was first necessary to break the HFEs down into their constituent MCFs. The study provides timelines of the most critical actions for the scenarios, such as "Enter procedure" or "Adjust valve." These tasks each represent a specific instance of one macrocognitive function, which if failed, would cause the HFE to occur without the opportunity for recovery. Each critical task was categorized into information-gathering, decision-making, or action execution.

Implicit information-gathering MCFs were appended to the task breakdowns to ensure completeness. For the crew to make decisions such as transferring to an emergency procedure, they must first gather information about plant parameters to ensure that the correct conditions are in place. They may also need to monitor a parameter in order to control it within a certain range. Two information-gathering tasks for obtaining and monitoring parameters were added to the list of MCFs for HFEs 1A and 1C, respectively.

After identifying and classifying the scenarios' MCFs, relevant CFMs were selected. Generally, most of the CFMs were relevant to each MCF. The crew could fail all of the action execution tasks by poor timing, incorrect operation, operating the wrong component, or not taking any action, which are all of the action execution CFMs. Many of the decision-making tasks did

not involve repetitive actions, so the CFM D8 (Premature Termination of Actions) was deemed not relevant. Likewise, many of the information-gathering tasks did not involve communication or responding to an alarm, so CFMs I1 (Key Alarm not Responded To) and I7 (Information Miscommunicated) were not considered for these tasks. For one information-gathering task that did involve detecting an alarm, CFM I1 was the only relevant CFM.

The BN models consider all CFMs by default. However, some methods provide the analyst with guidance on narrowing the failure modes considered. For example, the Phoenix method provides an extensive fault tree that can be used to select the relevant subset of CFMs based on the event's context. Different information-gathering and decision-making CFMs are relevant depending on whether the operators are relying on procedures or on their own knowledge. For the tasks in these validation scenarios, we considered all of the Action CFMs and most of the Information and Decision CFMs to be relevant. Irrelevant CFMs were removed by GeNIe's Control Value method, which mathematically separates the node from the rest of the model by removing all of the arcs from its parent nodes.

Lastly, evidence was observed on a selected set of relevant PIFs. The empirical study considers PIFs selected from and defined by the NRC's Good Practices for HRA [126] with several additional PIFs to more thoroughly explain crew behavior. These PIFs are enumerated in Table 5.3 along with their mapping to nodes in the BN models. The empirical study evaluates these PIFs in the context of crew operations, as positive, neutral, negative, or strongly negative influences, and these evaluations are reflected in the mapped node states.

The validation was conducted using the GeNIe software and its Python wrapper, PySMILE [92]. We instantiated BN structures for each HFE sequence using GeNIe's graphical interface, and set contextual evidence on the Information, Decision, and Action networks using PySMILE

to obtain error probabilities for each task. Then, these probabilities could be combined using BN structures to obtain an overall probability for the HFE.

PIF In U.S. HRA Empirical Study	**Node in BN Models (Relevant Cognitive Phase)**
Adequacy of time	Time Load (I, D, A)
Stress	Stress (I, D, A)
Scenario complexity	Task Complexity (I, D)
Indications of conditions	HSI (I, D)
Execution complexity	Task Complexity (A)
Training/Experience	Training/Knowledge/Experience/Familiarity (I, D, A)
Procedural guidance	Procedure Availability/Quality (I, D, A)
Human-machine interface	HSI (A)
Work Processes	Safety Culture (I, D, A)

Table 5.3: Mapping of empirical study PIFs to nodes in BN models.

5.6.3 Validation Results

This validation process was applied to all five HFEs in the three scenarios. Tables 5.4 to 5.7 show the HFE task progression of MCFs, cognitive phase of each MCF, and relevant PIF states that were set as evidence in the BNs. We iteratively obtained error probabilities for each task-context pair.

Then, the task error probabilities were combined into a total HEP value using a simple BN as an OR-gate, shown in Figure 5.5. In the BN, an Information, Decision, or Action MCF corresponding to the task and context is instantiated with a failure probability calculated by the appropriate BN model. The failure probability was generally higher for the Information and Decision tasks. This was in part because the PIFs were more often degraded for the cognitive portion of actions and in part because these tasks are typically more error-prone than action execution tasks. Each MCF node in the BN (pale yellow) is the direct parent of the HFE node

(darker yellow). The child HFE node has a simple and deterministic conditional probability table with a 1 (certain to happen) if any MCF is failed and a 0 (certain not to happen) if all the MCFs are successful.

MCF Name	Cog. Phase	CFMs Omitted	1A Error Prob.	1B Error Prob.
PIFs and States Applied: TimeLoad = Extra (I, D, A)*; Task Complexity = Moderately High (I, D); Task Complexity = Nominal (A); HSI = Poor (I, D); HSI = Nominal (A); TrainKnow-ExpFam = Nominal (I, D, A); ProcedureQual = Incomplete (A) *In HFE 1B, all PIFs identical except TimeLoad = Inadequate for I, D, and A functions.				
Start procedure ES-01	D	D4, D8	3.55×10^{-2}	2.18×10^{-1}
Detect that steam generator levels remain low	I	I1, I7	7.77×10^{-2}	1.64×10^{-1}
Start procedure FR-H1	D	D8	4.11×10^{-2}	2.23×10^{-1}
Actuate safety injection	A	–	1.92×10^{-3}	7.26×10^{-3}
Open power-operated relief valve to start feed & bleed	A	–	1.92×10^{-3}	7.26×10^{-3}
Close recirculation valve	A	–	1.92×10^{-3}	7.26×10^{-3}
Cross-connect auxiliary feedwater	A	–	1.92×10^{-3}	7.26×10^{-3}
Total Predicted HEP for HFE 1A			1.54×10^{-1}	–
Total Predicted HEP for HFE 1B			–	5.06×10^{-1}

Table 5.4: Task progression for HFE 1A & 1B and PIF states. Error probabilities for individual tasks and overall error probability included.

MCF Name	Cog. Phase	CFMs Omitted	Error Probability
PIFs and States Applied: TimeLoad = Inadequate (I, D, A); Task Complexity = Very High (I, D); Task Complexity = Nominal (A); TaskLoad = High (I, D); TaskLoad = Very High (A); HSI = Poor (I, D); HSI = Good (A); TrainKnowExpFam = Nominal (I, D, A); ProcedureQual = Poor (I, D); ProcedureQual = Nominal (A)			
Isolate ruptured steam generator	I	I1, I7, I9	3.67×10^{-1}
Stop auxiliary feedwater to ruptured steam generator	A	–	6.48×10^{-3}
Transfer from procedure FR-H1 to FRP1	D	D8	5.50×10^{-1}
Start procedure E-30	D	D8	5.50×10^{-1}
Adjust steam generator power operated relief valve setpoint	A	–	6.48×10^{-3}
Monitor steam generator pressure	I	I7, I1	3.81×10^{-3}
Stop safety injection pump	A	–	6.48×10^{-3}
Total Predicted HEP for HFE 1C			8.74×10^{-1}

Table 5.5: Task progression for HFE 1C and PIF states. Error probabilities for individual tasks and overall error probability included.

MCF Name	Cog. Phase	CFMs Omitted	Error Probability
PIFs and States Applied: Task Complexity = Very High (I, D); HSI = Poor (I, D); HSI = Good (A); TrainKnowExpFam = Low (I, D, A); ProcedureQual = Poor (I, D, A); ProcedureErr = Yes (I)			
Start procedure ES-01	D	D4, D6, D8	3.91×10^{-1}
Detect lack of component cooling water and sealwater	I	I1, I7	4.72×10^{-1}
Start procedure POP4-RCP	D	D4, D6, D8	3.91×10^{-1}
Stop all reactor coolant pumps	A	–	3.98×10^{-3}
Start positive displacement pump	A	–	3.98×10^{-3}
Total Predicted HEP for HFE 2A			8.06×10^{-1}

Table 5.6: Task progression for HFE 2A and PIF states. Error probabilities for individual tasks and overall error probability included.

MCF Name	Cog. Phase	CFMs Omitted	Error Probability
PIFs and States Applied: TimeLoad = Extra (I, D, A); Coordination = Nominal (I, D, A), HSI = Good (I, D, A); TrainKnowExpFam = Nominal (I, D, A); ProcedureQual = Nominal (I, D, A)			
Detect radiation alarms	I	(All but I1)	4.54×10^{-3}
Trip reactor	A	–	1.20×10^{-3}
Put auxiliary feedwater pump in pull-to-lock	A	–	1.20×10^{-3}
Start procedure E-30	D	D3, D8	2.07×10^{-2}
Adjust power operated relief valve setpoint	A	–	1.20×10^{-3}
Close main steam isolation valve and bypass	A	–	1.20×10^{-3}
Close auxiliary feedwater outside containment isolation valve	A	–	1.20×10^{-3}
Total Predicted HEP for HFE 3A			3.10×10^{-2}

Table 5.7: Task progression for HFE 3A and PIF states. Error probabilities for individual tasks and overall error probability included.

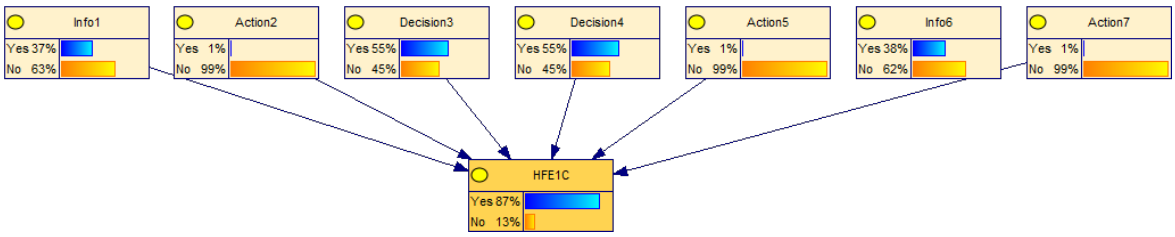

Figure 5.5: Probabilities combined for HFE 1C using a Bayesian network. The HFE 1C node is deterministically defined to happen if any of its parent nodes occur.

After obtaining the overall HEPs in this manner using the BN models, the resulting HEPs were compared to the HEP values obtained by experts using various established HRA methods as well as the empirical failure data from four crews. The U.S. HRA Empirical Study applies four

different HRA methods to the scenarios, including ASEP, SPAR-H, CBDT & HCR/ORE, and ATHEANA. Each method was conducted by two or three teams to allow for further intra-method comparison between analysts.

Four crews participated in the scenarios, which generated limited empirical failure data. All four of the crews failed HFE 2A, and three out of four failed HFE 1C. No failures were observed for HFEs 1A or 3A, and the scenario leading to HFE 1B did not occur. The crews qualitatively ranked each scenario by difficulty after completion of the study, with 2A and 1C being more difficult than 1A and 3A. The empirical study also performs a Bayesian updating procedure using the Jeffreys prior to obtain 5th and 95th percentile uncertainty bounds for the overall HEPs. These uncertainty bounds, as well as the HRA method results and the BN model results, are plotted in Figure 5.6.

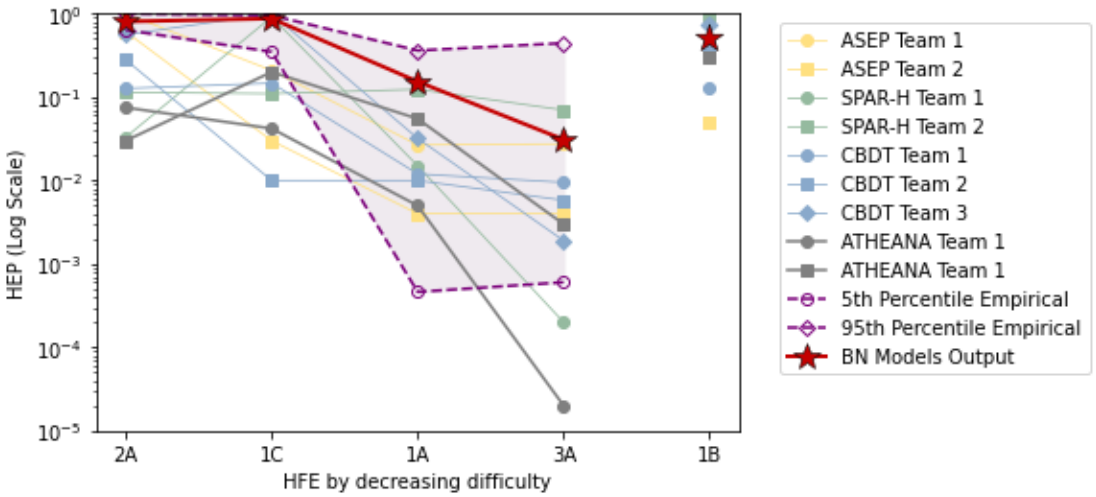

Figure 5.6: US Empirical Study method results, with ASEP, SPAR-H, CBDT & HCR/ORE, and ATHEANA methods considered. BN model outputs are marked with stars. Dashed lines indicate the uncertainty bounds obtained from the simulator data.

5.6.4 Discussion of Validation Results

As seen in Figure 5.6, all of the BN predictions fell within the empirically estimated 5th and 95th percentile uncertainty bounds. Most of the other HRA methods, with a few exceptions, under-predicted the HEP for HFEs 2A and 1C. These particular HFEs represent real, if unlikely, operational scenarios with catastrophic consequences. Furthermore, they present significantly greater challenges to operators than the steam generator tube rupture in HFE 3A, which is routinely trained upon. A commonly used HRA method should not predict an error probability of 0.01 or 0.03 for these scenarios when 4/4 operator teams (100%) or 3/4 teams (75%) are observed to fail. Although the BNs predict slightly higher HEPs than the other HRA methods, it is conservative in these situations where conservatism is desirable. For HFEs 1A and 3A, the predicted HEPs are high but remain well within the uncertainty bounds.

Another long-standing HRA issue highlighted in the empirical study is inter-method and inter-analyst variability. For example, nearly four orders of magnitude separate the analysts' estimates for the HEP of HFE 3A. The study also illustrates that different teams conducting analyses with the same method frequently estimate HEPs for the same task and context with one or more orders of magnitude difference. It is our hope that the BN method would resolve this issue; given a list of tasks and contextual factors, another analyst should be able to arrive at more similar conclusions.

This validation exercise demonstrates the BN models' holistic applicability to the analysis of real-world scenarios in nuclear power plants. Not only is it possible to apply the networks to any series of crew functions, whether inside or outside the control room, these models reflect the reality of crew actions without fully rejecting the wisdom and judgment of existing HRA

methods.

5.7 Discussion and Conclusion

Like the information-gathering and decision-making Bayesian networks documented in Chapters 3 and 4, this human performance model for operator action execution is founded in causal logic and reliability engineering theory. Using cognitive psychological literature documented in [32] as a starting point, additional human factors literature such as [83] and [114] was sought out to identify and justify causally related variables. This improves upon current HRA models, which primarily rely on expert opinion to justify relationships between factors, if they are explicitly declared at all. The parameterized structures are presented as a new, causally-based HRA method applicable to a wide variety of scenarios.

During model development, it was found that there are four human failure mechanisms that can lead to action execution human failure events. Analogous to mechanical failure mechanisms, these human failure mechanisms answer the question of *why* human failure events occur; the underlying reasons for failure. Crew failure modes can then describe the observable manner of how failure occurs. A fuller discussion of human failure mechanisms is available in [33].

The mechanisms in this model were found to be relatively unrelated in their modularity. The ancestor nodes of attention are entirely unrelated to the other three mechanisms, and the ancestor nodes of coordination are mostly unrelated to the other mechanisms with the exception of safety culture. However, the CFMs share a lot of parent nodes, forming several instances of common context idioms [35]. This may reflect the tendency of action execution error causes to be more direct: there are multiple factors that directly bring about the CFM, rather than intricate,

inter-connected causal chains representing cognitive processes.

We parameterized the model using a combination of data sources, including the SACADA database, expert-elicited Bayesian priors, IDHEAS-ECA and IDHEAS-DATA, other HRA methods, and psychological and organizational literature. The model outputs agree well with conventional HRA knowledge, with the probability of action execution failures on the order of 10^{-3} or 10^{-4}. It is possible to incorporate contextual evidence into the analysis for the purposes of diagnosing failure causes, or enumerating the direct and indirect effects of causal factors. In addition to previous structural and quantitative validation exercises, the models have been validated against nuclear training simulator data and other established HRA method analyses in the U.S. HRA Empirical Study. We found that the models' estimated HEP outputs consistently fell within the expected uncertainty bounds, more so than the other HRA methods, which suggests that they better reflect the error-prone reality of difficult emergency scenarios.

By comprehensively documenting the full causal origins of human failure events, the BN models described here provide a more valid and traceable HEP estimate than other HRA methods in current use. The model structures have a robust theoretical basis in reliability engineering, cognitive psychology, and HRA methodological literature. Furthermore, the model parameterization process is data-driven, incorporating all available data resources to derive the networks' conditional probabilities.

Future work can involve refining the error probabilities as additional data becomes available. No single data source exists that covers the full range of factors, mechanisms, and nodes contained in the network. BNs are able to flexibly accept further information without discarding existing data, and the models would benefit from further knowledge. In addition to refinement of the networks' parameterization, we intend to incorporate the models into the Phoenix HRA

method [10] for eventual regulatory use. Phoenix has many strengths in its thorough multi-layered decomposition of crew actions, its compatibility with existing probabilistic risk assessment tools, and its applicability to actions inside and outside the control room [22]. The BN models are fundamentally compatible with the Phoenix method through the use of similar PIFs and CFMs, which makes them a suitable addition to the method's framework. As the models are now complete and validated through several analysis cases, it is our intent to deploy the models for regulatory applications in nuclear power in the near future.

Chapter 6: Application of Bayesian Models to Case Study in Identifying Human Failure Events for External Hazards Scenarios

This chapter is partially reproduced from portions of [22] with additional contextualization, results, and conclusions.

Ahmad Al-Douri, Camille S. Levine, and Katrina M. Groth. "Identifying Human Failure Events (HFEs) for External Hazard Probabilistic Risk Assessment." In: *Reliability Engineering & System Safety* 235 (2023), p. 109236. DOI: https://doi.org/10.1016/j.ress.2023.109236.

6.1 Introduction

Nuclear power plant risk characterization is expressed through the metrics of core damage frequency (CDF), large early release frequency (LERF), and radiological consequences to the public and the environment. These values are determined using a variety of PRA tools and nuclear expert knowledge. Plant operating states are used to subdivide the plant operating cycle into unique states by examining operational characteristics (e.g. reactor power level, in-vessel temperature and pressure). Initiating events are the disruptions to the plant steady state that challenge plant control and safety systems, whose failure may lead to core damage or radioactive releases.

These initiating events have causes attributed to either internal or external hazards. This chapter focuses on external hazards, which include seismic events, high winds, or external flooding on coastal and river sites. The International Atomic Energy Agency (IAEA) has published general PRA requirements for external hazards and specific PRA requirements for each hazard type [127, 128].

Recently, there have been improvements in the capabilities of many NPP PRAs to reflect a greater understanding of external hazards, plant response, and sources of uncertainty [36, 129, 130]. However, several significant uncertainties related to the following issues remain: (1) complexity and diversity of natural phenomena, (2) impacts on structures, systems and components (SSCs), and (3) close coupling of the impacts of physical hazards, plant response, and human reliability of actions associated with plant response and recovery. Among the uncertainties of the third issue is the time (both required and available) to complete those actions under the hazards, as well as the ex-control room conditions under which the actions will be performed. Thus, human actions and human-plant interactions are key elements of successful prioritization of uncertainties within an external hazard PRA.

Human actions to mitigate external hazards frequently include tasks such as the construction of temporary barriers, checking to ensure that backup equipment is functional, and installing additional temporary equipment [131]. After the catastrophic seismic and tsunami events at the Fukushima Dai-ichi facility in 2011, the U.S. Nuclear Regulatory Commission conducted a review of existing plant safety in the face of external hazards and developed new safety requirements for nuclear power plants nationwide [131]. This led to the enhancement of plant capabilities and equipment, updating hazard assessment, and strengthening the degree of plant emergency preparedness. Under this review, licensees made modifications to procedures and identified al-

ternative mitigation strategies. However, human reliability analyses of these ex-control room actions are still not performed, as there is a lack of HRA methodologies equipped to address these scenarios [25, 99].

Within the wider PRA framework, HRA methods are used to qualitatively and quantitatively assess the human contributions to the overall risk of a system. This is done by building an understanding of potential human failure events (HFEs), estimating their probabilities and impacts, and proposing measures to reduce those errors. However, one of the shortcomings of most HRA methods, such as THERP [15] and SPAR-H [1], is that they were developed for internal events. The more recently developed second-generation HRA methods such as IDAC [19] and CREAM [16] are notable for their extensive grounding in psychological literature. However, the cognitive nature of these methods means that their primary application also lies in the analysis of control-room actions, where the physical nature of tasks is minimal and the bulk of errors occur in following procedures, reading signals, or completing other knowledge-based tasks. Starting from internal events PRA, licensed nuclear sites have attempted to account for the effects of external hazards by using human error probability HEP multipliers to account for the influence of PIFs, to capture the main driving factors influencing personnel behavior. The subjective nature of these HEP multiplying factors can impact the overall results and identification of dominant human errors [132].

The Phoenix method addresses shortcomings that are present in several earlier methods, which are enumerated in detail in Chapter 2. These include the lack of procedures to identify errors of commission and the lack of causal basis for linking operator error to context. These issues lead to insufficient traceability of qualitative and quantitative results and give rise to inter-analyst variability [10]. As such, the widespread use of Phoenix would improve the validity of

quantitative results used in regulatory risk-informed decision-making.

In this case study, Phoenix's application to ex-control room actions is rigorously addressed by: (1) evaluating the applicability of the cognitive-based Phoenix method to ex-control room actions, (2) modifying it to support the identification of human activities, causal factors leading to error, and associated uncertainties, and (3) quantifying the HEPs of these results by applying a novel, Bayesian network-based causal HRA method.

6.2 Literature Review

While significant efforts have been undertaken to model operator behavior for control room applications, modeling human actions in ex-control room environments have become increasingly important. In particular, humans play a critical role in performing flooding mitigation measures in power plants in the cases of hurricanes or tropical storms in coastal settings, or heavy rainfall in riverine settings. An analysis of 1,256 NPP events with safety significance found external events (seismic, flooding) to be the cause behind 66% of multi-unit incidents [133]. A major challenge in multi-unit PRA studies is the proper inclusion and accounting of human and organizational dependencies, which have been treated in a limited fashion for single-unit PRAs while being non-existent for multi-unit PRAs [50, 134–136]. Zoulis and Chang compared human error probability values from existing HRA methods (THERP, SPAR-H, and CBDT) for operators using portable pumps to mitigate spent fuel pool risks after a seismic event, and found the methods to be limited in addressing ex-control room actions [137].

A recent field study on actions outside the control room for a research-scale nuclear reactor identified three performance shaping factors as especially important in in external actions: com-

munication, situation awareness, and background knowledge [99]. Previous work from Groth and Mosleh presents a hierarchical set of performance influencing factors (PIFs) based on five categories (machine-based, person-based, team-based, organization-based, and situation/stressor-based) that can be expanded or collapsed based on the study objectives [24]. The PIFs identified in one research-scale reactor study [99] can be linked to Groth and Mosleh's taxonomy [24]: communication corresponds to team-based factors, situation awareness to situation/stressor-based factors, and background knowledge to person-based factors. These findings were used to develop and verify a framework for understanding the factors influencing human actions outside the control room. The five elements identified in the framework were: task identification and characterization, scenario characteristics, PIFs, testing environments, and methods for empirical data collection [20].

The IDAC method [19] was applied to model the conceptual component of error in this work. The method's cognitive framework considers three main stages of human response: Information, the process of gathering knowledge of the situation; Decision, the operator response phase of situation assessment, diagnosis, and planning; and Action, the process of executing the correct course of action. Each I-D-A phase can be decomposed into further I-D-A structures. For example, within the Information stage, the operator must first recognize the incoming information (I-in-I), decide how to process the information (D-in-I), and pass this information to their decision-making strategy (A-in-I). Crew interactions are also considered, characterized by formal communication, informal communication, and coordination, but due to their highly complex and dynamic nature, these are not modeled in IDAC beyond the crew members' primary responsibilities [19].

The Phoenix HRA methodology [10] expands upon IDAC through a layered qualitative

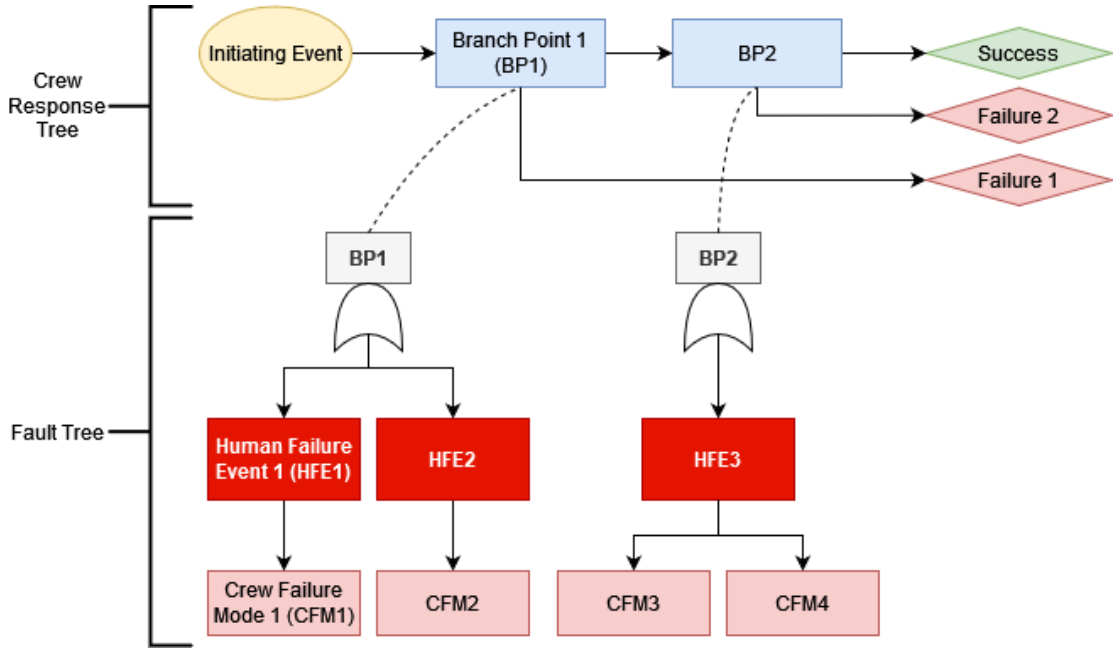

Figure 6.1: Multi-layered structure of the Phoenix HRA framework, adapted from [10].

analysis framework, as shown in Figure 6.1. The analyst must first gather PRA scenario information by developing, or obtaining, a detailed task analysis. This scenario information is used in concurrence with a guiding flowchart to develop a crew response tree (CRT) at the top layer. A CRT is a forward-branching event tree systematically covering all crew-system interaction scenarios, leading to the identification of human failure events [23]. In the next layer of decomposition, the human response model employs I-D-A to identify CFMs for the crew responses determined in the CRT.

Phoenix proposes a set of CFMs, shown in Table 3.2, to enumerate each possible form of failure that can occur at the Information, Decision, and Action stages. To determine which CFMs are applicable at each point in the scenario, a fault tree (FT) can be developed for each CRT branch point. Only a subset of the CFMs will be applicable depending on the I-D-A stage. After breaking down the failure context through an I-D-A lens, the lowest-level endpoints will consist of the HFE's relevant CFMs. The final layer of Phoenix involves selection of PIFs for each CFM

and encoding these factors into a quantitative causal structure, such as Bayesian networks. Instead of quantifying through the Phoenix method's PIFs and BN structure, we apply the structures shown in Chapters 3 to 5.

For non-NPP control room applications, the IDAC method has been used in the maritime industry to mitigate collision risks between autonomous surface ships. Similar flowcharts to those used in the Phoenix method were used to develop CRTs as event sequence diagrams. In one study, I-D-A was used to identify task phases in a hierarchical task analysis for supervising and assessing ship safety [138]. This was expanded upon through a concurrent task analysis, which re-describes tasks until they relate to only one of the I-D-A phases. Concurrent task analysis also identifies the interface tasks which exhibit dependency on another agent of the system [139].

Phoenix has also been used to identify HFEs arising in industry scenarios. For oil refinery and petrochemical plant operations, the method was been applied with modified CFMs to address fundamental differences between these facilities and nuclear power plants [140]. Phoenix has also been used to guide the development of a qualitative HRA framework for severe accident conditions at nuclear power plants [141]. This framework describes human error modes corresponding to the CFMs identified in Phoenix, as well as PIFs aggregated from a variety of existing HRA methods.

6.3 Methods

In this research, we applied the Phoenix framework to a set of task decompositions for ex-control room actions in order to create CRTs and FTs identifying human failure events. Through this approach, we assessed the suitability of the Phoenix method to ex-control room actions and

identified uncertainties that may arise in such an analysis.

6.3.1 Scenario Development Process

The qualitative Phoenix method is used here to guide the process of identifying HFEs and characterizing the precursor scenarios in the case of external hazards. This process begins by developing the PRA scenario(s) leading to the HFEs and gathering the needed information to support the construction of the CRT [10]. These scenarios can be developed in a variety of ways, including plant visits and operator interviews, tabletop walkdowns with a project team, reviews of plant procedures and training manuals, governmental reports and documents, and formal task analysis. Expert judgment may be necessary to synthesize the material from all of these sources, especially because there are frequently unwritten actions that may not be obvious to the analyst [60].

The objective of the CRT is to find the paths to predefined or new HFEs and identify possible opportunities for recovery. The CRT output is in the form of an event tree (ET) that can be used to find the failure and success paths of the function. This allows the analysts to perform a detailed assessment of the conditions that could lead crews to take inappropriate paths.

The main resource used to define the flooding scenario examined in this chapter was the NUREG/CR-7256 report titled "Effects of Environmental Conditions on Manual Actions for Flood Protection and Mitigation" [60]. The report provides an approach to decompose manual actions into tasks, subtasks, and specific actions, and further enumerates performance demands and presents a typology of these demands. Section 6.3 of [60] provides three flooding mitigation task decomposition examples that were achieved through group discussion and consensus

building by a research team representing a wide array of expertise.

In NUREG/CR-7256, three representative manual actions with available procedures were decomposed. These manual actions included the installation of a portable diesel pump, installation of flood barriers on exterior intake structure walls, and construction of a sandbag berm around the service water strainer pit. The authors chose one task from each manual action and then decomposed it further into subtasks and specific actions. Additional information was provided for each specific action, including the degree of sheltering, the location (fixed or variable), and supplementary comments. A task decomposition for fastening barriers presented in [60] is shown here in Table 6.1.

6.3.2 Development of Crew Response Trees

After obtaining the task analysis, CRTs were developed to identify HFEs corresponding to a given safety function. Since the Phoenix framework was developed for internal hazards in NPPs, safety functions are defined as the intended function of a specific plant system, a desired state of the plant in response to an upset event, or a combination of both. Using the modular approach proposed in the Phoenix qualitative framework, one CRT was developed to model HFEs corresponding to each identified safety function. The two main inputs to this process were the crew and plant context, and the procedures used to carry out the safety function. The supplementary comments for each specific action were of particular use in characterizing the CRT branching points and their context.

Specific Actions	Degree of Sheltering	Location	Comments
Subtask 1 - Procure a Forklift and Move it into Position to Lift the Steel Plate			
Walk from the current location of the individual performing the subtask to where the forklift is parked	Unsheltered	Variable	If a forklift is not already available from an earlier subtask in which material and equipment are staged at the intake structure, then a forklift must be procured.
Climb into the forklift	Unsheltered	Fixed	The forklift was assumed not to have an enclosed cab.
Operate the forklift from the forklift's parking location to drive to the appropriate exterior wall locations	Unsheltered	Variable	This action involves driving around the building to appropriate locations and requires a forklift operator.
Subtask 2 – Load Steel Plate onto Forklift			
Operate the forklift to pick up the steel plate	Unsheltered	Fixed	This task involves a skilled forklift operator using the forklift controls to lift the steel plate.
Subtask 3 - Place the Steel Plate Against the Opening Using the Forklift			
Operate the forklift to move the steel plate	Unsheltered	Variable	This task involves a skilled forklift operator using the forklift controls to move the steel plate to where it needs to be installed at openings.
Communicate non-electronically with other personnel to coordinate placement of steel plate on the exterior wall	Unsheltered	Variable	Involves non-electronic communication and coordination.
Operate the forklift to place the steel plates against the doors or openings on the exterior walls	Unsheltered	Semi-fixed	This task requires a forklift operator to drive the forklift and to manipulate the steel plate into position. It may also require the coordinated actions of additional personnel.
Move heavy materials manually (i.e., adjust the steel plate against the wall/door)	Unsheltered	Fixed	Requires the coordinated actions of several personnel.
Subtask 4 – Drill Holes into the Intake Structure and Secure Steel Plate Using Fasteners			
Use hand tools (i.e., drill holes in the exterior walls)	Unsheltered	Fixed	This task involves skilled personnel (i.e., craft labor) using hand tools to drill holes in intake structure walls for lag bolts or concrete anchors.
Use hand tools (i.e., to place lag bolts or concrete anchors in place)	Unsheltered	Fixed	This task involves skilled personnel (i.e., craft labor) using hand tools to place lag bolts or concrete anchors.

Table 6.1: Task Decomposition for Fastening Barriers to External Walls (reproduced from [60]).

6.3.3 Identification of Crew Failure Modes for CRT Branches

The Phoenix methodology presents a set of 19 CFMs, categorized by the Information (I), Decision (D), or Action (A) phase in which it occurs. These CFMs, shown in Table 3.2, are defined as the generic functional mode of failure of the crew in its interactions with the plant and represent the proximate cause of failure. Phoenix also defines CFMs as being mutually exclusive or orthogonal to avoid double counting crew failure scenarios during HEP estimation. After constructing the CRT event tree, fault trees were developed based on each of the branch points identified. In these fault trees, the HFEs are the top event and CFMs are the basic events [10].

During this exercise, we identified two additional CFMs not covered by the Phoenix method, though they were not identified during the task shown in this case study. These were CFMs D8 (Premature Termination of Actions) and A4 (No Action Taken), described further in Chapter 4 and Chapter 5. These CFMs refer respectively to situations where the operator prematurely decides to stop taking actions such that the end result is ineffective, and where the operator correctly gathers information and decides on a course of action but simply does not take that action. Neither of these new CFMs arose during analysis of the barrier installation task, but they were relevant to the pump and sandbag installation tasks.

6.4 Creation of CRTs and FTs

Following the methodology outlined above, we created CRTs and the corresponding FTs for each of the main subtasks in Table 6.1 for the task of fastening barriers to walls. CRTs and FTs were also developed for each of the subtasks for the pump installation and sandbag building tasks, omitted here for brevity. The CRT takes the form of an event tree, with failure and success

paths stemming from branch points of interest. Each BP in the CRT generally corresponds to a main specific action in the task decomposition. Local recovery opportunities, such as extra time, change of crew, or change of equipment, are shown leading from the failure paths back onto the main CRT. If the subtask is successful, the success path leading to the next subtask is followed. For example, if the operator(s) successfully complete all four actions in Subtask 3 (Place steel plate against opening using forklift), they will now move on to the next step, Subtask 4 (Drill holes and secure plate with fasteners). Based on the CRT layer for each subtask, a subsequent FT layer is developed to identify specific I-D-A CFMs for each BP.

These CRTs and FTs are shown here in Figures 6.2, 6.3, 6.4, and 6.5. Color coding for CRTs and FTs is as follows: yellow for subtasks, blue for branch points (sub-subtasks), light gray for branch point failure events, darker pink for HFEs, and lighter pink for specific CFMs. The failure path and success path boxes are light red and green, respectively. Dashed lines indicate recovery opportunities, though recovery is not considered in this analysis.

For all three major tasks (pump, barriers, and sandbags), it was found that most of the HFEs, and most of the identified CFMs, were in the Action macrocognitive phase. A significant portion of the relevant CFMs identified for the barrier-fastening task shown above were in the Information phase due to the needs for coordination and continual monitoring of the forklift controls. Since the tasks span all three cognitive phases, it was necessary to apply all three BN models during quantification.

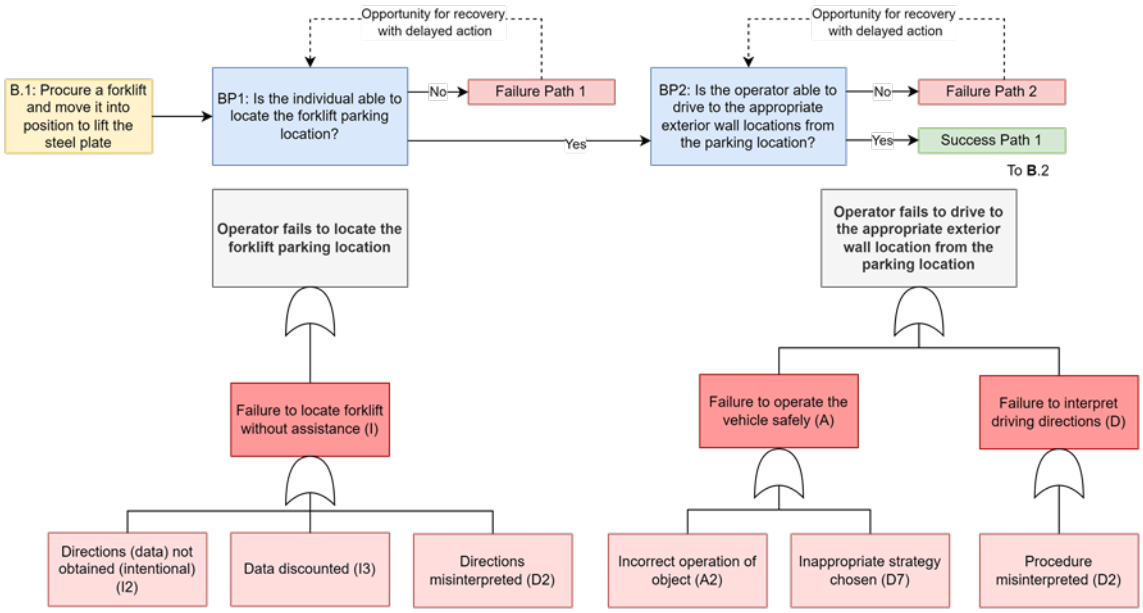

Figure 6.2: CRT and FT diagrams for Subtask 1 - Procure a forklift and move it into position to lift the steel plate

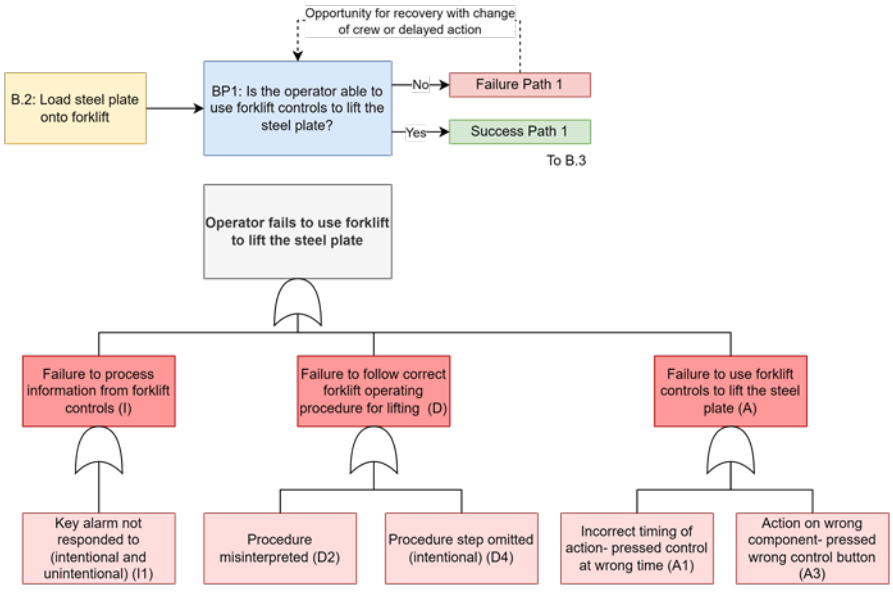

Figure 6.3: CRT and FT diagrams for Subtask 2 - Load steel plate onto forklift

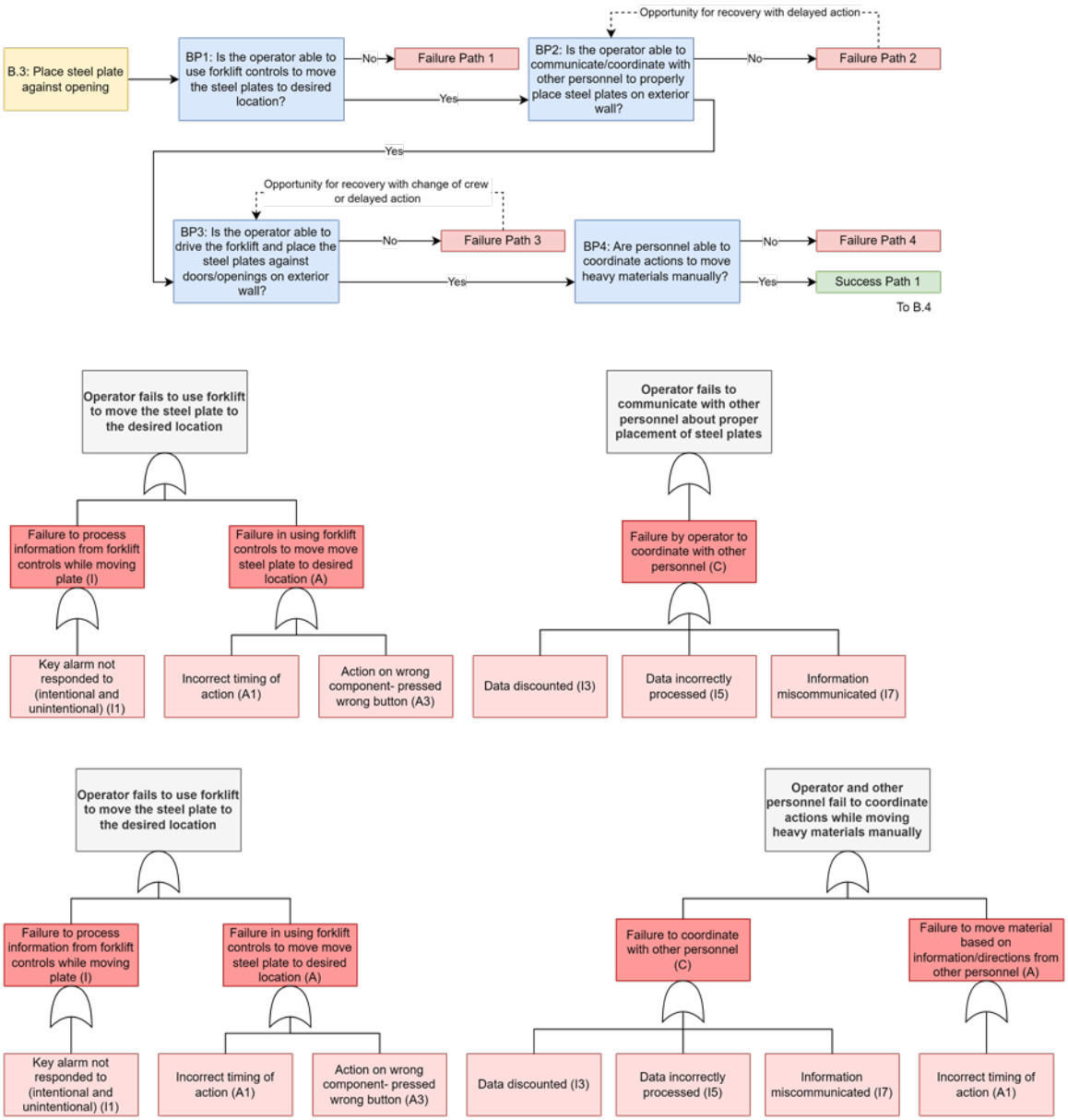

Figure 6.4: CRT and FT diagrams for Subtask 3 - Place steel plates against opening

6.5 Quantification of Error Probabilities Using BN Models

To quantify the basic events of the resultant FTs, we directly applied the BN models to find the probability of each CFM. This process was automated using GeNIe and the PySMILE Python wrapper [92] in a similar manner to the other validation exercises conducted in Chapter 4

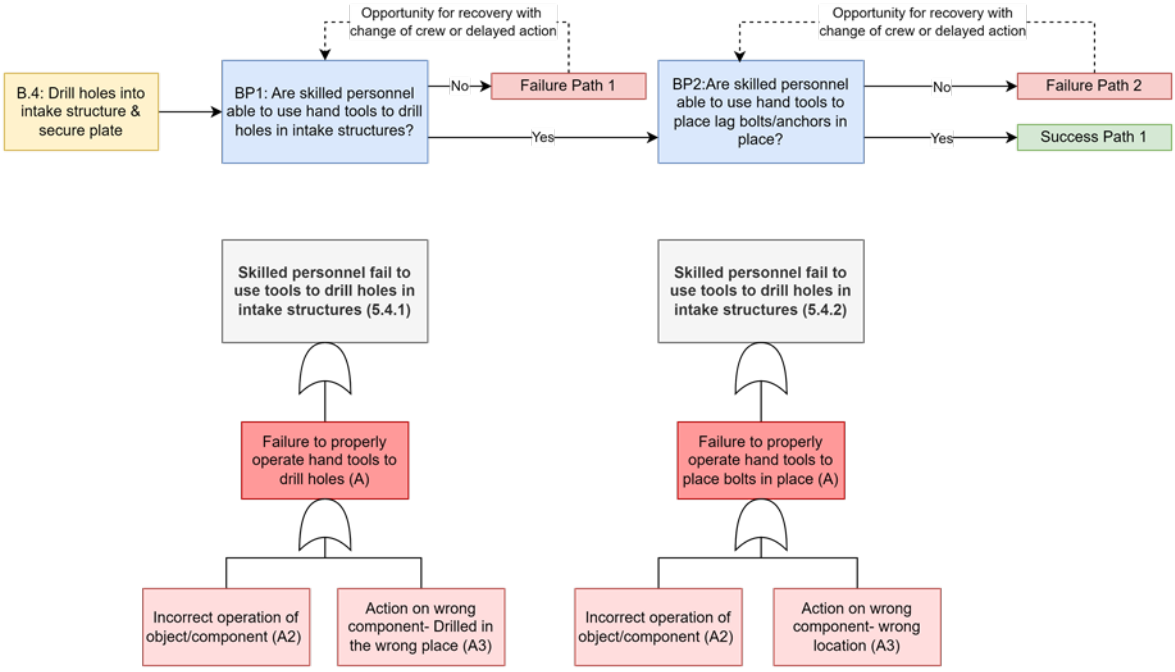

Figure 6.5: CRT and FT diagrams for Subtask 4 - Drill holes into intake structure & secure plate

and Chapter 5.

First, it was assumed that due to the unsheltered nature of the actions, the external environment would be in a degraded state. It was also assumed that the operator would be experiencing high levels of stress due to the frightening nature of the flooding situation. In addition to the environment and stress affecting all of the subtasks, it was assumed that Subtask 3 (Place steel plate against wall opening) would present high task loads, due to the need for communication, precise manipulation with machinery, and manual manipulation of heavy objects. These high loads would manifest for both information-gathering and action execution aspects of the subtask. These factors were encoded in the BN analysis by setting the following evidence on the networks: ExternEnv = Degraded; Stress = High; and for Subtask 3, TaskLoad = High for the Information phase, and TaskLoad = ModHigh for the Action phase.

In addition to setting evidence of the presence of degraded PIFs, we also set evidence that

certain CFMs irrelevant to the scenario could not occur. For example, CFM I1 (Key Alarm Not Responded To) was the only relevant failure mode to Subtask 2 (Load Steel Plate onto Forklift). Evidence was set on all other information-gathering CFMs for this task that they did not arise. After setting all evidence, the networks' beliefs were updated and the CFM probabilities extracted. Then, all evidence was cleared before repeating this procedure for the next subtask.

After obtaining the probability of each basic event, the FTs were analyzed using Boolean logic to calculate the probability of the top event. This probability was then applied to the requisite branch point in the CRT event tree. Then, the probability of each end state was calculated by multiplying the probabilities along each branch. Then, the probabilities of failure end states were summed in order to find the overall failure probability for each subtask.

The results of this analysis are summarized in Table 6.2. The HFE Probability column reports the HEP at each branch point obtained by calculating the probability of each fault tree's top event. Significant information-gathering activities render several subtasks more susceptible to failure. In particular, the potential for miscommunications in Branch Points 2 and 4 of Subtask 3 are the main contributors to relatively high failure probabilities for each branch point of 0.41. Conversely, the HEPs calculated for the Subtask 4 branch points are quite low, with both at 0.00091, as no information-gathering or decision-making is required for this step. In the Subtask Failure Probability column, the overall probability of failing the subtask is provided as calculated from the event trees. The failure probability of Subtask 1 is 0.14, Subtask 2 is 0.19, Subtask 3 is the highest at 0.83, and Subtask 4 is the lowest at 0.0018. These probabilities show that the third subtask, as the greatest opportunity for HFEs, should be made more fault-tolerant, and mitigation measures should be put in place.

Mitigation measures can be guided by the BN results. By addressing the PIFs with the

Subtask	Branch Point	HFE Probability	Subtask Failure Probability
1: Procure Forklift	1: Locate Forklift	0.13	0.14
	2: Drive Forklift	0.015	
2: Load Plate	1: Load Plate	0.19	0.19
3: Place Plate	1: Move Plate (general)	0.31	0.83
	2: Communicate on Placement	0.41	
	3: Move Plate (specific)	0.31	
	4: Manually Move Plate	0.41	
4: Drill & Secure	1: Drill Holes	0.00091	0.0018
	2: Place Bolts	0.00091	

Table 6.2: HFE probabilities for the top event of each fault tree depicted in Figures 6.2 to 6.5, and the overall failure probability for each subtask considering the CRT event tree.

largest quantitative contributions to the error probabilities, it is possible to develop a more targeted and effective plan to reduce the overall error likelihood.

6.6 Discussion and Conclusion

This novel application demonstrates the compatibility of our BN models with the Phoenix method, and further, the Phoenix method's applicability to external actions. The Phoenix method is exceedingly flexible and system-agnostic, as demonstrated by its adaptation to petroleum industry operations [140] and our application to external flooding actions [22]. We were able to apply the Phoenix framework of CRTs and FTs to this scenario. Furthermore, we found that the majority of the Phoenix method's CFMs are relevant and sufficient for the analysis of external flooding mitigation actions, with the addition of two CFMs.

Phoenix also integrates well with our BN models, which allow an analyst to easily quantify the likelihood of each CFM occurring given a set of PIFs describing the scenario's circumstances. Currently, the Phoenix method is quantified via a fully connected BN, which assumes that every

high-level PIF affects every CFM. Our BN models employ a more layered, causal understanding of the PIFs in their structure, with fewer but more substantiated arcs. This enables identification of causal chains leading to HFEs and mitigation strategies targeted to break these chains, and also significantly simplifies the computational burden of quantifying the model.

Lastly, this exercise also demonstrates the utility of incorporating our BN models into the Phoenix framework. Phoenix's CRT and FT decomposition for identifying potential errors augments the explanatory power of the BN models, and the BN models improve upon the current Phoenix quantification capabilities with their traceable structure. We anticipate completing this incorporation in the near future so that the models can be used in regulatory applications. We also hope to augment the BNs with more data as it becomes available, potentially using the framework to refine and guide future data collection efforts. This continual refinement will improve the accuracy of the networks' quantitative HEP estimates, which are more traceable and justifiable than those of any HRA method in current regulatory use.

Chapter 7: Summary of Contributions and Future Research Directions

This chapter details the engineering contributions and the publications, presentations, and models associated with the research. The chapter concludes with some directions for future research to expand upon the current work.

7.1 Engineering Contributions

7.1.1 Expansion of HRA's Causal and Theoretical Basis

The primary engineering contribution of this dissertation is the creation and validation of causal human reliability analysis models with a comprehensive theoretical and data basis. These models have unified the practice of HEP quantification and the available HRA literature basis. This work incorporates interconnected scientific theories, models, and literature to develop the cognitive causal model structures that advance the traceability and accuracy of HRA. This addresses Objective 1, as described in Chapter 1, by explicitly incorporating causality into the models' theoretical basis.

A strong basis in literature enables the robust causal characterization of information gathering, decision-making, and action execution errors during an operator team's interaction with a complex engineering system. Using these models, an analyst can explore the myriad causal path-

ways leading to HFEs to pinpoint the root causes of error. In turn, this allows for the development of more targeted preventative measures or corrective action plans to mitigate HFEs.

7.1.2 Development of a Data-Driven Quantification Process

This work synthesizes many different quantitative data sources to parameterize the models. Many of the data sources used here are being applied to comprehensive HRA BNs for the first time. This represents a novel use of the NUREG-6949 priors, multipliers from the IDHEAS, CREAM, and SPAR-H methods, and psychological literature. We also expand the use of the SACADA database in parameterizing BNs, building beyond the analysis of [85].

Furthermore, we have rigorously documented the parameterization of the networks, rendering any quantitative result directly traceable back to the data source. For all values except the CREAM and SPAR-H multipliers, it is possible to directly access the original study's data yielding that model parameter. This research documents and more thoroughly justifies the quantitative estimate from the BN than any other HRA method to date. Coupled with the models' structural basis in reliability engineering literature, this robust quantitative basis in human reliability data makes all model outputs traceable, attributable to high-quality resources, and rooted in data and science.

Not only is this quantification structure more traceable and data-driven than other available methods, but the model quantification process is straightforward and intuitive. Setting known evidence and automatically updating beliefs via a BN avoids complex calculations and can easily be used to conduct sensitivity analysis. BNs also afford versatility in their support of both prospective and retrospective reasoning. The quantification process developed in this work addresses

Objective 2 as it is easily accessible to the analyst, transparent, and data-driven.

7.1.3 Construction of Versatile, Validated Models

We have validated multiple aspects of the models on a wide variety of NPP scenarios. First, we held an expert validation session to ensure that the models were complete and accurate in their structure. The structures were also compared to narratives of abnormal event progressions, and then the models' quantitative output was validated against normal operational data. The models were then holistically validated against the U.S. HRA Empirical Study's simulated abnormal events and the analytical results of other HRA methods. A final case study incorporating the Phoenix method framework was conducted on external flooding mitigation actions. The models performed well in each scenario.

This manifold validation demonstrates the broad applicability of the models to both normal and emergency operations, inside and outside the control room. Furthermore, the validation against other HRA methods and real crews' simulator data showed significantly improved accuracy over the other methods for events with a high likelihood of error.

The models' application alongside the Phoenix method also showcases their relevance to external hazard scenarios, an area of human actions often overlooked by HRA as being skill-of-the-craft. The BN models also fit neatly into the Phoenix framework due to the similar set of CFMs used. Because Phoenix already uses BN models to quantify the method's FTs and ETs, it is very straightforward to quantify using the BN models developed in this work instead. The case study's numerical output shows that the BNs expand the breadth of scenarios covered by traditional HRA while still maintaining compatibility with cognitively based HRA methods in

current use, addressing Objective 3.

7.2 Publications and Work Products

7.2.1 Journal Publications

The work done over the course of this project was disseminated through five peer-reviewed journal papers, three of which are currently in press. The three papers which are currently in press make up the bulk of Chapters 3, 4, and 5. The paper discussing gaps and opportunities has been adapted for Chapter 2, and the paper discussing the application of Phoenix to external hazards risk assessment has been adapted for Chapter 6.

Published:

- [22] Al-Douri, A., **Camille S. Levine**, and Katrina M. Groth. (2023). "Identifying Human Failure Events for External Hazard Probabilistic Risk Assessment." *Reliability Engineering and System Safety*, 235. https://doi.org/10.1016/j.ress.2023.109236

- [25] **Levine, C. S.**, Ahmad Al-Douri, Vincent P. Paglioni, Michelle Bensi, and Katrina M. Groth. (2024). "Identifying human failure events for human reliability analysis: a review of gaps and research opportunities." *Reliability Engineering and System Safety*, 245. https://doi.org/10.1016/j.ress.2024.109967

Submitted:

- [71] **Levine, C. S.** and Katrina M. Groth. (Submitted). "Three novel causal logic-informed models for identifying and quantifying operator errors, Part I: Information-Gathering." *Reliability Engineering and System Safety*.

- [101] **Levine, C. S.** and Katrina M. Groth. (Submitted). "Three novel causal logic-informed models for identifying and quantifying operator errors, Part II: Decision-Making." *Reliability Engineering and System Safety*.

- [117] **Levine, C. S.** and Katrina M. Groth. (Submitted). "Three novel causal logic-informed models for identifying and quantifying operator errors, Part III: Action Execution." *Reliability Engineering and System Safety*.

7.2.2 Conference Presentations

In addition to the journal papers documenting this work, several conference papers were written and presented at various stages of the project. Presenting to a diverse and knowledgeable group of experts yielded invaluable feedback and insights that were used to refine the research. We presented this work at PSAM16 (Honolulu, HI), NPIC&HMIT 2023 (Knoxville, TN), PSAM17 (Sendai, Japan), and the ANS 2024 Winter Meeting (Orlando, FL). A final conference paper summarizing the work has been submitted to PSA 2025, which will be presented this summer.

Published:

- [142] Al-Douri, A., **Camille S. Levine**, and Katrina M. Groth. "Identifying human Failure Events (HFEs) for External Hazard Probabilistic Risk Assessment." *Probabilistic Safety Assessment and Management 16*, Honolulu, HI, June 26-July 1, 2022. http://hdl.handle.net/1903/29082.

- [33] Paglioni, V. P., **Camille S. Levine**, Ahmad Al-Douri, and Katrina M. Groth. "Why do Human-Machine Teams Fail: Investigating Failure Mechanisms in HRA." *NPIC&HMIT 2023*, Knoxville, TN, July 15-20, 2023.

- [44] **Levine, C. S.**, Ahmad Al-Douri, and Katrina M. Groth. "Causal Pathways Leading to Human Failure Events in Information-Gathering System Response Activities." *NPIC&HMIT 2023*, Knoxville, TN, July 15-20, 2023.

- [100] **Levine, C. S.**, Ahmad Al-Douri, and Katrina M. Groth. "Causal Pathways Leading to Decision and Action Human Failure Events: Structures and Validation." *Probabilistic Safety Assessment and Management (PSAM) 17 & Asian Symposium on Risk Assessment and Management (ASRAM)*, Sendai, Japan, October 7-11, 2024.

- [143] **Levine, C. S.** and Katrina M. Groth. "Expanding the Causal Logic Foundations of Human Reliability Analysis: Current Project Status." *ANS 2024 Winter Meeting*, Orlando, FL, November 17-21, 2024.

  Submitted:

- **Levine, C. S.** and Katrina M. Groth. "Novel Quantitative Causal Logic Models for Human Reliability Analysis." Submitted to PSA 2025, to be presented June 15-18, 2025.

7.2.3 Models and Code

A main result of this work is the information, decision, and action BN model suite, supplemented by several codes and auxiliary models which were used during parameterization and validation. The models were created using the GeNIe visual interface and parameterized in Python using PySMILE [92]. The Trilith software was also used to conduct the external hazards case study by quantifying the scenario event trees and fault trees. Additional information about the codes and their availability can be obtained by emailing Camille Levine at clevine1@umd.edu.

- Bayesian network model suite:
  - ✻ I_Phase_2024_NUMBERS.xdsl: GeNIe BN model for information-gathering human failure events.
  - ✻ D_Phase_2024_NUMBERS.xdsl: GeNIe BN model for decision-making human failure events.
  - ✻ A_Phase_2024_NUMBERS.xdsl: GeNIe BN model for action execution human failure events.
- Parameterization codes:
  - ✻ info_parameterization.py: Python code for parameterizing information-gathering network. Contains all sources used and number of subjects in studies, if applicable.
  - ✻ decision_parameterization.py: Python code for parameterizing decision-making network. Contains all sources used and number of subjects in studies, if applicable.
  - ✻ action_parameterization.py: Python code for parameterizing action execution network. Contains all sources used and number of subjects in studies, if applicable.
- Validation codes and models:
  - ✻ GRS_validation.py: Python code for validating models against GRS operational data and examining the log-differences between model outputs and data points.
  - ✻ EmpiricalStudy_validation.py: Python code for validating models against U.S. HRA Empirical Study. The code also plots other HRA methods' output from the empirical study's analysts alongside the BN model outputs for each scenario.

- Empirical_HFE1A_Quant.xdsl: Auxiliary GeNIe model used to quantify overall HEP for HFE 1A in the U.S. HRA Empirical Study.
- Empirical_HFE1B_Quant.xdsl: Auxiliary GeNIe model used to quantify overall HEP for HFE 1B in the U.S. HRA Empirical Study.
- Empirical_HFE1C_Quant.xdsl: Auxiliary GeNIe model used to quantify overall HEP for HFE 1C in the U.S. HRA Empirical Study.
- Empirical_HFE2A_Quant.xdsl: Auxiliary GeNIe model used to quantify overall HEP for HFE 2A in the U.S. HRA Empirical Study.
- Empirical_HFE3A_Quant.xdsl: Auxiliary GeNIe model used to quantify overall HEP for HFE 3A in the U.S. HRA Empirical Study.
- XHPRA_Validation.py: Python code for applying BN models to external hazards flooding case study.
- XHPRA Application- Barriers.isi: Trilith model with linked fault trees and event trees for quantifying overall HEPs for external hazards flooding case study.

7.3 Opportunities for Future Research

7.3.1 Improved Usability for Analysts

In the near-term, it would be beneficial to develop a stand-alone graphical user interface for the BN models. This would allow analysts to conduct inference with the models without searching through the models' many nodes, node states, and node probabilities. The interface could take the form of a checklist with a collapsible hierarchy of CFMs, human failure mech-

anisms, and PIFs upon which the analyst can set evidence. Evidence that poses direct conflict could be automatically flagged before running the model. The analyst could also use this checklist to select which results they are most interested in: the error probabilities, or the likelihoods of degraded PIF states. This would eliminate the need for interacting with the models through GeNIe or Python and improve the analysts' user experience.

We also intend to incorporate the BNs directly into the Phoenix HRA framework to support PRA integration. Because Phoenix currently uses BNs in its quantification process, as well as a similar palette of CFMs and PIFs to the models presented here, this should be a relatively straightforward task. UCLA has developed an online tool for conducting analyses with Phoenix, which greatly increases the ease of applying the method. This task is planned to be completed in the near future.

7.3.2 Continuous Data Augmentation

One main strength of using BNs as a model structure is that they can flexibly accept additional data as it becomes available without discarding any previous knowledge. As future human reliability data collection efforts are undertaken, the models' parameters can be updated when new information becomes available. Furthermore, some nodes (e.g. training/knowledge/experience/familiarity) were condensed or removed from the network as it was not possible to discern their individual effects from the available data. Once it becomes possible to separate and quantify the effects of these particular variables, the models' quantitative capabilities will be more complete.

7.3.3 Application to Further Domains

In addition to post-initiating event interactions with the system, the BN models hold significant promise for modeling pre-initiating event activities. Maintenance tasks are an important category of pre-initiator actions that may give rise to human failure events or physical system failures if conducted improperly, and are typically omitted from HRA analyses. Maintenance actions must be completed on most complex engineering systems, and are not limited to strictly nuclear power generation.

The models could also be applied to other system domains, such as aviation, data centers, petrochemical processing, defense, or ground transportation. To tailor the models to a particular field, Bayesian updating can be conducted with domain-specific data to account for unique aspects of operations. As human reliability data studies become available in a wider range of fields, updating the models with this data will be a straightforward process due to the flexibility of the Bayesian network model structures.

7.3.4 Preparation for Regulatory Applicability

Finally, future efforts concerning these models should focus on preparing them for use in nuclear PRA regulatory applications. Discussions are currently underway with Japanese nuclear regulators regarding the implementation of the Phoenix method in upcoming PRA analyses. It is also our hope that these models will be suitable for use in applications beyond the typical nuclear uses of HRA, especially as we have shown the BNs to be applicable to external hazards mitigation actions. To assess their utility in other fields, the models should be benchmarked against current HRA methods that are used in these additional fields as well as any existing data. We envision

that the models can be refined to support applicability and usability for a wide range of systems and scenarios.

Appendix A: Substantiation of Information-Gathering CFM & Mechanism Causes and Effects

A.1 Substantiation of Information Crew Failure Mode Causes

Each directed arc in the Bayesian network is based on a real causal relationship between the nodes which is documented in the human factors and cognitive psychology literature. The literature sources used are enumerated in Table 3.3. The first section of this appendix provides a source and quote for each mechanism's influence, and the subsequent section in turn details the influences of PIFs on the mechanisms.

The I-phase CFM I1, "Key alarm not responded to," is caused by four main human failure mechanisms: **Attention** [73], **Bias** [73, 83], **Coordination** [73, 74], and **Prioritization** [81]. The CFM may also be caused directly by issues with *human-system interface (HSI)* and in the *external environment* to operations, which are both PIFs.

Table A.1: Mechanisms and PIFs identified as causing CFM I1: "Key alarm not responded to".

Mechanism/PIF	**Substantiation**
Attention	"A major problem for SA [situational awareness] occurs when attention is directed to a subset of information and other important elements are not attended to, either intentionally or unintentionally." [73]
Continued on next page	

Mechanism/PIF	Substantiation
Bias	Unintentionally towards particular source of info: "Cue salience, therefore, will have a large impact on which portions of the environment are initially attended to." [73] Intentionally towards expectations: "The second influence on visual search that leads to departures from the serial model has to do with the top-down implications of searcher expectancies of where the target might be likely to lie." [83]
Coordination	"[Macrocognitive processes]... Individual mental model construction: Defined as individual team members using available information and knowledge to develop their mental picture of the problem situation... Develop, rationalize and visualize solution alternatives: Defined as a team member or the whole team using knowledge to describe a potential solution." [74] "Overall team SA can be conceived as the degree to which every team member possesses the SA required for his or her responsibilities... If each of two team members needs to know a piece of information, it is not sufficient that one knows perfectly but the other not at all." [73]
Prioritization	"Meeting organizational and social goals often appears to outweigh safety goals, especially in ambiguous conditions." [81]
HSI	"A major problem for SA occurs when attention is directed to a subset of information and other important elements are not attended to, either intentionally or unintentionally." [73] In addition to attentional degradation directly resulting in a key alarm being ignored, poor HSI or controls may passively direct information-gathering away from the correct information.
External Environment	"For example, vibration will reduce the quality of visual input and motor output, and noise will do the same for auditory input (Poulton, 1976)." [83]

The CFM I2, "Data not obtained" is only caused by the mechanism of **Bias** [81], and is the only information-gathering phase CFM not to be affected by attentional degradation. However, I2 is also affected directly by many individual PIFs, such as *procedure availability*, the *external environment*, the *HSI*, *perceived urgency and severity*, and *overloading*.

Table A.2: Mechanism and PIFs identified as causing CFM I2: "Data not obtained".

Mechanism/PIF	**Substantiation**
Bias	"Cues that signal a problem are not always clear-cut. Conditions can deteriorate gradually, and the decision maker's situation assessment may not keep pace. If events occur infrequently, the decision maker may not have amassed the experience to recognize the signals associated with a different CoA [course of action]." [81] A lack of experience may result in biases against certain sources of information, because the operator may automatically rule out certain infrequent scenarios, and seek out information to confirm that a more familiar scenario is present.
Procedure Availability	"Is the situation covered by normal work know-how or planned procedures? ... [leads to] Information not seen or sought" [Rasmussen's flowchart] [80] If procedures are unavailable for the scenario, it is likely that the operator will not know what cues to actively search for, or the correct cues will not catch their eye.
External Environment	"For example, vibration will reduce the quality of visual input and motor output, and noise will do the same for auditory input (Poulton, 1976)." [83]
HSI	"A major problem for SA occurs when attention is directed to a subset of information and other important elements are not attended to, either intentionally or unintentionally." [73]
Perceived Urgency & Severity	"A major problem for SA occurs when attention is directed to a subset of information and other important elements are not attended to, either intentionally or unintentionally." [73] When certain pieces of information are presented in a more salient manner, they are perceived as more urgent. Important but less salient information will not be sampled. "There are several possible reasons for differences in cognitive appraisal... One may fail to understand the risk. Here the climber may see the clouds approaching but not appreciate their implications for electrical activity and icy rock. One may be relatively more confident or even overconfident in one's ability to deal with the hazard." [83]
Continued on next page	

Mechanism/PIF	Substantiation
All Loads	"...increased tendency to sample dominant or probable sources of information under stress, cognitive tunnel vision." [73] Under high loads, the operator will fall back on more obvious information sources and block out new or conflicting information, which may result in data not being obtained.

The CFM I3, "Data Discounted," is caused by the mechanisms of **Attention** [73], **Bias** [81], **Procedure Error** [80], and **Prioritization** [80]. The CFM is not caused by any additional PIFs. All of the substantiation quotes for this CFM are the same as for the I2 and I4 CFM substantiations, with the exception of Prioritization. Discounting data sources is a similar failure mode to the incorrect decision to stop collecting data, with similar causal origins.

Table A.3: Mechanism identified as causing CFM I3: "Data Discounted".

Mechanism/PIF	Substantiation
Prioritization	"Connolly suggests distinguishing between tree-felling and hedge-clipping. Tree-felling... the decision is consequential, its goals are well defined, and there is a clear way of achieving them. In contrast, hedge trimming exemplifies decisions that are made incrementally... makes more sense to find your ways as you go... plans are of limited value when the future is uncertain and goals are ambiguous." [80] "Because high-stakes decisions entail severe losses in cases of failure, they cannot be studied... [They] increase vacillation prior to making a decision and decrease the likelihood of abandoning it once made, particularly if the decision maker can be held responsible in the case of failure." [80] As the operator progresses through a scenario that they are prioritizing due to high perceived stakes, they may fall into a tree-felling pattern of decision-making. If a data source's information conflicts with their schema of the scenario, it will be discounted as it is not useful to the planned course of action.

The CFM I4, "Decide to stop data collection," is caused by the mechanisms of **Attention** [73] and **Bias** [81]. The mechanism can additionally be caused by issues with *overloading* as well as *morale, motivation, and attitude (MMA)*, when the operator decides it is unnecessary to continue obtaining data due to time constraints or lack of motivation.

Table A.4: Mechanisms and PIFs identified as causing CFM I4: "Decide to stop data collection".

Mechanism/PIF	Substantiation
Attention	"premature closure has also been found to be more likely under stress. This includes considering less information and attending more to negative information... scanning of stimuli under stress is scattered and poorly organized." [73] High levels of stress are a main cause of attentional degradation. This frequently results in abandoning data collection due to rushing a course of action.
Bias	"Schuch (1992) investigated midair collision accidents (MAC) which are interesting because they tend to involve experienced pilots. He concluded that because experienced pilots have made repeated flights without an incident they become desensitized and stop scanning the sky." [81]
All Loads	"A higher amount of stress can have extremely negative consequences, however, as accompanying increases in autonomic functioning and aspects of the stressors can act to demand a portion of a person's limited attentional capacity (Hockey, 1986)." [73] "increased tendency to sample dominant or probable sources of information under stress, cognitive tunnel vision" [73] High loads tend to cause various kinds of stress that prevent continuous sampling of information.
MMA	"action is determined by the decision maker's values... role of exploratory action and the consequences of action" [Connolly] [80] If the operator is unmotivated or has poor work attitudes, they will be less likely to continue exploring the information that is available.

The CFM I5, "Data Misinterpreted," is caused by the mechanisms of **Attention** [73] and **Procedure Error** [82]. I5 is not caused by any additional PIFs. Misinterpreting data is a particularly interesting CFM because the operator correctly perceives that there is important data, but does not correctly interpret the meaning. It is documented in the literature that misperceptions may happen due to **Bias** as well. These expectations likely shape the failure process through the mechanism of **attention**.

Table A.5: Mechanisms identified as causing CFM I5: "Data Misinterpreted".

Mechanism/PIF	Substantiation
Attention	"operator can attend only to a subset of information or may be actively working to achieve SA, yet has erroneous or incomplete perception and integration of information." [73]
Procedure Error	"Operators also needed knowledge of the procedures, which includes not only knowledge of how to follow the individual procedure steps, but also knowledge of assumptions and logic that underlie the procedures... goal prioritization, response plans and their rationale, and procedure transition network" [82] Without a mental model of the procedures, and why the procedures are put in place, an operator may be prone to misinterpreting the context of the information they are noticing or seeking.
Bias	"information is attended to, but is misperceived, frequently due to the influence of prior expectations" [77]

The CFM I6, "Reading Error," is only caused by the mechanism of **Attention** [73]. Reading errors can also be caused by issues with a poor *HSI*. With a poor interface structure, it will be difficult for the operator to accurately characterize visual information, and there will be low information availability.

Table A.6: Mechanism and PIF identified as causing CFM I6: "Reading Error".

Mechanism/PIF	Substantiation
Attention	Substantiated by [73] in a similar way to CFM I5. Rather than being misinterpreted, the information is incorrectly read.
HSI	"environmental cues with highly salient features will tend to capture attention away from current goal-directed processing." [73]

The CFM I7, "Data Miscommunicated," is caused by issues with **Attention** [74] and **Coordination** [74]. This is an especially complex CFM because it is closely tied to the macrocognitive phase of Collaboration, requiring multiple independent actors in the scenario. Additionally, there is a multitude of ways that miscommunication may manifest. There may be issues with the content or format of the information sent (potentially related to the *HSI*), the receiver's understanding of the information, the timing of the message, or a lack of confirmation of receipt.

Table A.7: Mechanisms and PIF identified as causing CFM I7: "Data Miscommunicated".

Mechanism/PIF	Substantiation
Attention	"In naturalistic environments, team members are often required to perform multiple simultaneous tasks. In multi-tasking environments, tools that support attention management will be especially important." [74] For scenarios that involve splitting multiple tasks between multiple team members, poor attention to a task can result in a degraded awareness the system and of others' knowledge of the system. This can lead to erroneous communication content or timing.
Coordination	"We expect that group members who share these labels will create shared cognition, have knowledge interoperability and increase their performance." [74] "Team Knowledge Development- Defined as all team members participating to clarify information for building common team knowledge. ... The externalized processes are the cognitive activities of exchanging information with the other team members through verbal and non-verbal communication." [74]
HSI	"electronically mediated communication is degraded in terms of social and contextual clues." [74] Depending on the communication interface of the system, information transfer may be degraded if context cues are lost.

The CFM I8, "Wrong data attended to," is caused by the mechanisms of **Attention** [73] and **Procedure Error** [80]. Procedure errors may arise in two different ways: the operator may not realize that the task is not covered by the procedure, or the operator may not respond to the correct task-defining information, in both cases defaulting to the most familiar cues. I8 may also be caused by issues with the *HSI*.

Table A.8: Mechanisms and PIFs identified as causing CFM I8: "Wrong data attended to".

Mechanism/PIF	Substantiation
Attention	Substantiated by [73] in a similar way to I1 and I2. In this case, not only is the correct data not obtained, but attention is spent on the wrong data source.
Continued on next page	

Mechanism/PIF	Substantiation
Procedure Error	"Is the situation covered by normal work know-how or planned procedures?" potentially leads to "Operator responds to familiar cue which is incomplete part of available information" [Rasmussen's flowchart] [80] With a poor grasp of the procedures, the operator may get distracted by familiar information and fail to notice additional relevant cues.
HSI and external environment	"Major problem for SA when attention is directed to a subset of information and other important elements are not attended to... designs that restrict access to SA elements (via information filtering, for instance) will contribute to this problem." "problem of information overload in many systems still must be considered. [and addressed by] filtering of extraneous information and reduction of data by processing and integrating ..." "environmental cues with highly salient features will tend to capture attention away from current goal-directed processing." [73]

Finally, the CFM I9, "Data not sampled with appropriate frequency," is caused by the mechanisms of **Attention** [73], **Bias** [81], and **Coordination** [73]. I9 may additionally be caused by *overloading* or *stress*. I9 brings in aspects of scenario dynamics, such as the consideration that different operational states have different temporal information-gathering requirements. "In dynamic environments, many decisions are required across a fairly narrow space of time, and tasks are dependent on an ongoing, up-to-date analysis of the environment." [73]

Table A.9: Mechanisms and PIFs identified as causing CFM I9: "Data not sampled with appropriate frequency".

Mechanism/PIF	Substantiation
Attention	"Complex tasks with multiple input sources appear to be particularly sensitive to the effects of stressors... performance decrements... took place during the information input stage." [73]
Bias	"Cues that signal a problem are not always clear-cut. Conditions can deteriorate gradually, and situation assessment may not keep pace. If events occur infrequently, the decision maker may not have amassed the experience to recognize the signals associated with a different CoA [course of action]." [81]
Continued on next page	

Mechanism/PIF	Substantiation
Coordination	"If each of two team members needs to know a piece of information, it is not sufficient that one knows perfectly but the other not at all... Information transmission... It may constitute a verbal exchange or separate, direct viewing of displays, with each individual independently acquiring information on the status of the aircraft." [73] Poor coordination can result in knowledge gaps between team members. One member may assume that the other has a complete grasp of the scenario, even if this is not the case. This results in degraded information-gathering frequency.
All Loads	"A different amount of relevance may be indicated for different goals... At least some SA on all elements has been found to be needed, even if this conveys merely that the element is not very important at the moment." [73] When all data elements are presented as being important, the resulting high workload will overwhelm the operator, and they will only be able to sample the information at a limited frequency.
HSI	"Because the state of the environment is constantly changing, often in complex ways, a major portion of the operator's job becomes that of obtaining and maintaining good SA." [73] Poor HSI makes it difficult for the operator to keep apprised of dynamic situations. Checking information to update situational awareness may be trivial, or may require a significant amount of the operator's effort.

A.2 Substantiation of Information Mechanism Causes

The mechanism of **attentional** failure is broadly caused by several person- and situation/stressor-based PIFs, as well as by the mechanism of **bias**, which is described in more detail below. Selective attention is used to detect and select important cues in the environment from which to gain information. Attentional failures can be caused by exceedance of *sensory limits* due to task complexity, *overloading*, *stress* which may be caused by overloading, and *fatigue* which may be caused by stress.

Table A.10: PIFs identified as causing the mechanism of Attention for Information-Gathering.

PIF	Substantiation
Sensory Limits	"An increase in the amount of information presented causes a relative decline in efficiency... the greater the amount of overlap between two messages the lower the relative efficiency. ... The increase was not as large as the increase in information presented, but it seems to imply that messages conveying little information may be dealt with simultaneously, while those containing much information may not. This point is very relevant to the idea of a limited capacity for the nervous system." [75] "The situation assessment included plausible goals, critical cues, expectancies, and an obvious course of action" [78] "increasing the salience of cues that are important within the context of the situation" [78] "[For skill-based behavior,] Information (sensory input) at this level is processed as signals: It triggers action directly without explicit consideration of what the information means or what the decision maker's goals are." [Rasmussen] [80] "Level of performance is significantly determined by sensory factors... there is an optimal level of attentional resources for a given task and further attentional resources devoted to the task will not result in an improvement in performance." [76] "as the size of the attentional field increases, the density of processing resources within the field decreases" [76]
Continued on next page	

PIF	Substantiation
All Loads	"Low SA with Low Workload: The operator may have little idea of what is going on and is not actively working to find out because of inattentiveness, vigilance problems, or low motivation... Low SA with High Workload: ... the operator can attend to only a subset of information... erroneous or incomplete perception and integration of information" [73] "A simplistic decision analysis that separated different goals might have been misleading, whereas a more sophisticated decision analysis would be difficult to carry out under these time pressures" [78] "If there is not adequate time, the decision maker is prepared to implement the course of action that experience has generated as the most likely to be successful" [78] "In addition to temporal pressures, high cognitive load (e.g., due to uncertain, dynamic and excessive information) may reduce sharing unique information and result in poor task performance (Stasser and Titus 1987)." [74] "under various forms of stress, people tend to narrow their field of attention to include only a limited number of central aspects... complex tasks with multiple input sources appear to be particularly sensitive to the effects of stressors." [73] "Some 22.9% of the SA errors occurred when a pilot or controller either became momentarily distracted by other relevant tasks or experienced an overall high level of workload... also experienced a failure to monitor or observe data as a result of distraction by situations not related to flying the aircraft." [77]
Continued on next page	

PIF	Substantiation
Stress	"Under perceived danger, a decrease in attention has been observed for peripheral information." [73] "A higher amount of stress can have extremely negative consequences, however, as accompanying increases in autonomic functioning and aspects of the stressors can act to demand a portion of a person's limited attentional capacity." [73] "...analytical decision making [rather than recognition-primed decision making] seems to prevail when the available data are abstract and alphanumeric rather than perceptual, when the problems are very combinatorial, and when there is a strong requirement to justify the course of action chosen." [78] "...operator can attend only to a subset of information or may be actively working to achieve SA, yet has erroneous or incomplete perception and integration of information." [73] "increased tendency to sample dominant or probable sources of information under stress, cognitive tunnel vision" [73] "premature closure has also been found to be more likely under stress. This includes considering less information and attending more to negative information... scanning of stimuli under stress is scattered and poorly organized." [73]
Fatigue	"In fact, maintaining sustained, vigilant attention in low-workload situations is both fatiguing and stressful" [83] "In addition, researchers have reported that tasks particularly sensitive to sleep disruption are those involving judgment (Krueger et al., 1985), learning or storing new material (Williams et al., 1959), as well as those tasks involving self-initiated cognitive activity, like maintaining situation awareness and planning (Banderet et al., 1981)" [83]

The mechanism of **bias** is similar to attention, but it is primarily an individual cognitive process instead of a task-driven process. It is caused by person-based PIFs related to knowledge base such as *knowledge*, *training*, and *familiarity*, as well as perceived characteristics of the situation. Biases may be caused by an operator's *perception of responsibility* for their individual actions as well as their *perception of the situation's urgency and severity*. Crucially, although biases may be grounded in the reality of the situation, they are first filtered through the individual's perception.

Table A.11: PIFs identified as causing the mechanism of Bias for Information-Gathering.

PIF	**Substantiation**
Perceived situation: urgency, severity, decision impact	"When multiple goals are compatible with each other, several may be active at once. When goals are incompatible, their associated priority level for the identified situation class determines which shall be invoked." [73] "... loss of skills an operator may experience by virtue of not having been an active perceiver, decision maker, or controller during the time that automation assumed responsibility for the task... may make the operator less trustful of his or her own performance and hence more likely to continue to use automation ... [and the] ability to intervene appropriately should the system fail." [83] "Connolly suggests distinguishing between tree-felling and hedge-clipping. Tree-felling... the decision is consequential, its goals are well defined, and there is a clear way of achieving them. In contrast, hedge trimming exemplifies decisions that are made incrementally... makes more sense to find your ways as you go... plans are of limited value when the future is uncertain and goals are ambiguous." [80] "Because high-stakes decisions entail severe losses in cases of failure, they cannot be studied... [They] probably... increase vacillation prior to making a decision and decrease the likelihood of abandoning it once made, particularly if the decision maker can be held responsible in the case of failure." [80]
Continued on next page	

PIF	Substantiation
All Loads	"Low SA with Low Workload: The operator may have little idea of what is going on and is not actively working to find out because of inattentiveness, vigilance problems, or low motivation... Low SA with High Workload: ... the operator can attend to only a subset of information... erroneous or incomplete perception and integration of information" [73] "increased tendency to sample dominant or probable sources of information under stress..." [73] "premature closure has also been found to be more likely under stress. This includes considering less information and attending more to negative information." [73]
Operator's knowledge base: knowledge, training, familiarity	"Schemata and mental models are developed as a function of training and experience. A novice may have only a vague idea of important system components and sketchy rules or heuristics..." [73] "Because abnormal events tend to be quite infrequent, the correct responses may not be familiar." [81] "Operators also used their knowledge of the physical interconnections of plant systems, and... plant parameter behavior, to generate and narrow potential explanations of plant symptoms" [82] "Schuch (1992) ... concluded that because experienced pilots have made repeated flights without an incident they become desensitized and stop scanning the sky [causing midair collisions]." [81] "A major component of this skill is learning the strategy of how to allocate resources differentially in the most optimal fashion, and how to schedule sequential activities so that different tasks are performed at the best time." [83]

Team **coordination** is the process of sharing information and responsibilities between operators working towards the same set of goals. Errors in coordination are prone to the influence of organizational and team-based factors, such as *safety culture*, *staffing*, *supervision*, *role awareness*, team *cohesion*, and *communication availability and quality*. Issues with hiring and management processes cause individual workers to not be aware of their responsibilities, such as alarm response, data source monitoring, and communicating information to others. Coordination errors may also be caused by gaps in the operator's declarative knowledge, which may be due to lack of *training* or low levels of general *knowledge and experience*.

Table A.12: PIFs identified as causing the mechanism of Coordination for Information-Gathering.

PIF	Substantiation
Safety Culture	"An organization's culture is conceptually tied to the predominant national culture of the organization and it manifests itself in multiple ways; for example, different role structures" [74] "Need to foster and support accurate mental models through training and control room displays and decision aids" [82] "We saw repeated cases where operators stopped to discuss as a group whether the procedure path... would eventually lead them to take the actions they recognized to be important for safe recovery" [82] "Organizational factors emphasize productivity which may conflict with safety... peer pressure may encourage risky behavior... Meeting organizational and social goals often appears to outweigh safety goals..." [81]
Staffing	"any examples of performance problems in coalitions have been documented that are related to unanticipated differences in work styles or organizational procedures (such as Siegel 1998)" [74]
Supervision	Forms a Definitional logic idiom with Cohesion and Role awareness [35].
Role Awareness	"When people become interrelated as part of a group they acquire and maintain social relationships... [which] include group composition, communication structure, division of labor (roles), interpersonal relationship structure and power structure." [74] "*performance depends on members' sharing of role-specific information* in a timely and cooperative fashion." [74] "team members must become aware of which team members hold unique information by being able to communicate effectively" [74]
Cohesion	"Team members develop shared mental models by proceeding through the three phases of mental model convergence: orientation, differentiation, and integration." [74]
Communication	"*...performance depends on members' sharing of role-specific information* in a timely and cooperative fashion." [74] "electronically mediated communication is degraded in terms of social and contextual clues." [74] "Technology mediated communication... Low information richness... may decrease inhibition, decrease self-regulation, increase self-absorption, and increase counternormative behavior... ultimately deteriorating communication among team members." [74]
Continued on next page	

PIF	Substantiation
Declarative knowledge base: Knowledge/Training	"Organizations can benefit by analyzing their group tasks; when groups face scenarios for which there are patterns, the patterns should be labeled and tools created to allow groups to communicate these labels with minimal task interruption." [74] "If a team member is trying to communicate to their team members that they are seeing a pattern ... then only one message need be sent to signal recognition of the pattern. However, if that team member is communicating all the discrete elements ... then several more messages will need to be sent." [74] "In addition to temporal pressures, high cognitive load (e.g., due to uncertain, dynamic and excessive information) may reduce sharing unique information and result in poor task performance." [74]

The mechanism of **prioritization** is a process associated with the operator's *morale, motivation, and attitudes (MMA)*. This refers to the tasks that an operator chooses to address first when faced with conflicting or simultaneous goals [24]. Like attentional failures, prioritization issues can also be brought about by biases in which the operator is biased towards addressing a particular information source. Prioritization issues can also be brought about directly through issues with the operator's MMA, as well as poor *HSI*, *perceived situation urgency and severity*, and *perceived decision impact*.

Table A.13: PIFs identified as causing the mechanism of Prioritization for Information-Gathering.

PIF	Substantiation
MMA	"action is determined by the decision maker's values... role of exploratory action and the consequences of action" [Connolly] [80]
HSI	"A major problem for SA occurs when attention is directed to a subset of information and other important elements are not attended to, either intentionally or unintentionally... designs that restrict access to SA elements (via information filtering, for instance) will contribute to this problem." "problem of information overload in many systems still must be considered. The filtering of extraneous information (not related to SA needs) and reduction of data (by processing and integrating low level data) ..." "environmental cues with highly salient features will tend to capture attention away from current goal-directed processing." [73] "Because the state of the environment is constantly changing, often in complex ways, a major portion of the operator's job becomes that of obtaining and maintaining good SA." [73] "electronically mediated communication is degraded in terms of social and contextual clues." [74]
Perceived situation characteristics: urgency, severity, decision impact	"There are several possible reasons for differences in cognitive appraisal... One may fail to understand the risk... One may be relatively more confident or even overconfident in one's ability to deal with the hazard." [83] "Is the situation covered by normal work know-how or planned procedures? ... [leads to] Information not seen or sought" [Rasmussen's flowchart] [80] "When multiple goals are compatible with each other, several may be active at once. When goals are incompatible, their associated priority level for the identified situation class determines which shall be invoked." [73]

The mechanism of **procedure error** is brought about when the procedures are incorrectly applied to the situation, so information is gathered incorrectly. This may happen because the operator simply does not have the depth of *experience* or *training* needed to deal with the situation properly, or due to complicated situational characteristics that cause high *workload*. Organizational factors like *safety culture* and *procedure quality* may lead the operator to have low confidence in the procedures or be less inclined to follow instructions as well.

Table A.14: PIFs identified as causing the mechanism of Procedure Error for Information-Gathering.

PIF	Substantiation
Operator's knowledge base: knowledge, training, familiarity	"If there is not adequate time, the decision maker is prepared to implement the course of action that experience has generated as the most likely to be successful" [78] "Analytical strategies are more frequently used by decision makers with less experience" [78] "The danger of misapplying recognitional decision strategies is that personnel will lack the experience needed to identify effective courses of action... mentally simulate the pitfalls... optimize when necessary" [78] "At the same time there was a second strand of activity that was self-initiated and cognitively based... crews exhibited evidence of SA and response planning that enabled them to handle aspects of the situation that were not fully covered by the procedures" [82] "Cues that signal a problem are not always clear-cut. Conditions can deteriorate gradually, and the decision maker's situation assessment may not keep pace. If events occur infrequently, the decision maker may not have amassed the experience to recognize the signals associated with a different CoA [course of action]." [81]
All Loads	"A simplistic decision analysis that separated different goals might have been misleading, whereas a more sophisticated decision analysis would be difficult to carry out under these time pressures" [78] "If there is not adequate time, the decision maker is prepared to implement the course of action that experience has generated as the most likely to be successful" [78]
Safety Culture	"Need to foster and support accurate mental models through training and control room displays and decision aids" [82] "We saw repeated cases where operators stopped to discuss as a group whether the procedure path... would eventually lead them to take the actions they recognized to be important for safe recovery" [82]
Procedure Quality	"Is the situation covered by normal work know-how or planned procedures? ... [leads to] Information not seen or sought" [Rasmussen's flowchart] [80]

Appendix B: Substantiation of Decision-Making CFM & Mechanism Causes and Effects

B.1 Substantiation of Decision Crew Failure Mode Causes

Each directed arc in the Bayesian network is based on a real causal relationship between the nodes which is documented in the human factors and cognitive psychology literature. The literature sources used are enumerated in Table 4.1. The first section of this appendix provides a source and quote for each mechanism's influence, and the subsequent section in turn details the influences of PIFs on the mechanisms.

The Decision-phase CFM D1, "State Misdiagnosed", is caused by two mechanisms: **Perceived Urgency/Severity** [83] and **Mental Model** [73]. D1 can also be caused directly with *human-system interface (HSI)* issues. Furthermore, this CFM may be the way in which an information-gathering failure manifests in error: "SA [situational awareness], therefore, is based on far more than simply perceiving information about the environment. It includes comprehending the meaning of that information in an integrated form, comparing it with operator goals, and providing projected future states of the environment that are valuable for decision making." [73] When the operator believes that they have accurate information, but actually have incorrect or incomplete situational awareness due to the constraints of information-gathering processes, mis-

Table B.1: Mechanisms and PIFs identified as causing CFM D1: "State Misdiagnosed".

Mechanism/PIF	Substantiation
Perceived Urgency/Severity	"There are several possible reasons for differences in cognitive appraisal. One may fail to perceive the circumstances of risk... One may fail to understand the risk... One may be relatively more confident or even overconfident in one's ability to deal with the hazard." [83] A low perceived severity may lead to a poor appraisal of the situation at hand.
Mental Model	"The key to using these models to achieve SA rests on the ability of the individual to recognize key features in the environment–critical cues that will map to key features in the model. The model can then provide for much of the higher levels of SA (comprehension and projection) without loading working memory." [73] If the operator has a faulty mental model, environmental cues will not be accurately mapped to mental constructs, which will make it difficult to correctly characterize the state.
HSI	"...situations where the crews were most negatively affected by the presence of degraded indications were the situations where they kept working on a problem without stopping to reflect on the information they received from the HSI and without considering alternative explanations for what they saw." [105, 107] Without a satisfactory HSI to communicate the system state, the operator may ignore the information that they do receive.

diagnoses are possible: "Finally, of the information displayed by the system and that directly acquirable from the environment, there may be incomplete or inaccurate transmission to the human operator because of perceptual, attention, and working memory constraints." [73]

The CFM D2, "Procedure Misinterpreted", is only caused by one mechanism, **Skills** [113]. Following a procedure is a largely cognitive task dependent on the operator's knowledge base and cognitive abilities, which comprises their skill level. D2 can also be caused by issues with *coordination* and *procedure quality*. If procedures are poor or unavailable, it is almost certain that the operator will misinterpret the required tasks or underlying logic, so the CFM is defined as being in a failed state.

Table B.2: Mechanism and PIFs identified as causing CFM D2: "Procedure Misinterpreted".

Mechanism/PIF	Substantiation
Skills	"People who are highly experienced with a task tend to process information at the skill-based level, reacting to the raw perceptual elements at an automatic, subconscious level; without the need to interpret and integrate cues or consider possible alternate actions but instead responding to cues and patterns that are already associated with actions." [113]
Coordination	"In the complex case, the diagnosis time went considerably up when the quality of the team cognition was bad... The extended time for diagnosis in the complex case could be seen in that the crews had difficulties interpreting and following the procedures, since there was a mismatch between procedures and the plant situation." [105, 110] A high task complexity may lead to degraded communications, that in turn impacts the team's ability to coordinate procedural actions.
Procedure Quality	"They showed that due to the complexity made by the lack of certain indications that the procedures asked for, many different paths through the procedures were observed, and these led to the extended diagnosis time." [102, 105]

The CFM D3, "Failure to Adapt Procedures to Situation", is also caused by only one mechanism, **Mental Model** [81]. If the operator's mental model of the situation is incorrect, they will be more likely to incorrectly apply procedures. D3 may also be caused by issues with the *procedure quality*, *task complexity*, and *overloading*. Like D2, the CFM is defined as being in a failed state if the procedure quality is poor.

Table B.3: Mechanism and PIFs identified as causing CFM D3: "Failure to Adapt Procedures to Situation".

Mechanism/PIF	Substantiation
Mental Model	"A common pattern was the crew's decision to continue with their original plan when conditions suggested that other courses of action might be more prudent... consequences were not anticipated or evaluated." [81] Operators are reluctant to abandon their initial mental model, so a faulty mental model may particularly exacerbate the situation during a dynamic scenario.
Continued on next page	

Mechanism/PIF	Substantiation
Procedure Quality	"But procedures are not without problems, as indicated by some of the examples listed above. They are essentially feed-forward control devices – prepared atone time and place to be applied at some future time and place – and they suffer, along with all such control systems, the problem of dealing with local variations." [103] Procedures are unable to properly cover the range of dynamic scenarios that operators may encounter.
Task Complexity	"While many times the operator may 'know' through training or experience that certain information is not being shown, people are so visually dominant that in times of stress or when task load is high, they can easily forget or not attend to this non-present information." [104]
All Loads	"As situations degrade, risk and time pressure may rise. These conditions may limit the decision maker's ability to project the situation into the future and mentally simulate the consequences of a course of action." [81] The previous quote from [104] for Task Complexity also applies.

The CFM D4, "Procedure Step Omitted", is caused by the mechanisms of **Bias** [104, 108] and **Prioritization** [104]. The crew may adhere to internal biases about the necessity or the importance of the procedure steps, and omit necessary tasks. This CFM is not directly caused by any additional PIFs.

Table B.4: Mechanisms identified as causing CFM D4: "Procedure Step Omitted".

Mechanism/PIF	Substantiation
Bias	"Biases in the evaluation of compound events are particularly significant in the context of planning... the general tendency to overestimate the probability of conjunctive events leads to unwarranted optimism in the evaluation of the likelihood that a plan will succeed..." [108] An operator may omit steps due to overconfidence in the success likelihood of their plan. "If people focus on the wrong goals, they may not be receptive to the correct information, or may not seek out needed information at all. In many cases, they may actually be working toward achieving a goal that is not so important, while all the time thinking they are achieving what the task at hand calls for." [104] Issues with biases can lead to incorrect prioritization.
Continued on next page	

Mechanism/PIF	Substantiation
Prioritization	"Prioritization of tasks should, in the end, be up to the person who knows the relative importance of what is happening in the environment... Requiring a physical action to acknowledge that alarm can again take time away from other tasks, or can lead to distraction if left unresponded to." [104]

The CFM D5, "Inappropriate Transfer to Another Procedure", is caused by the mechanisms of **Bias** [104, 112], **Prioritization** [104], and **Mental Model** [83]. D5 may additionally be caused by high *stress* levels. This CFM mainly stems from internal biases and issues with the operators' mental models that cause them to think that the current procedure is less applicable to the situation than another procedure.

Table B.5: Mechanisms and PIFs identified as causing CFM D5: "Inappropriate Transfer to Another Procedure".

Mechanism/PIF	Substantiation
Bias	"...confirmation bias, which is the tendency to overly prioritize and seek for information that benefits existing views... essentially did not change the hypothesis that it was only one steam generator that was faulty- they simply changed their mind about which one it was... Finally, the medium scoring crews tended to search for information that backed up (rather than falsified) their views" [112] In this situation, confirmation bias led the crew to transfer their actions from one inappropriate procedure to another.
Prioritization	Substantiated by [104] in the same manner as CFM D4, "Procedure Step Omitted". Rather than omitting a step, the crew may be led towards an incorrect course of action due to overconfidence or incorrectly prioritized goals.
Mental Model	"If the working hypothesis, action, or plan fails to meet minimum criteria (which are heavily affected by time constraints), the decision maker may generate a new hypothesis, action, or plan. When a plan is finally selected, it is executed and the person monitors the environment to update their situation assessment and to determine whether changes in procedures must be made." [83]
Continued on next page	

Mechanism/PIF	Substantiation
Stress	"Functional magnetic resonance imaging (fMRI) scans have demonstrated that as stress increases, the efficiency of the prefrontal cortex breaks down and we become more instinctive or primal in action." [109] Under stress, the operator may revert to their default actions rather than considering whether it is an appropriate response to the scenario.

The CFM D6, "Delayed Action", is caused by the mechanism of **Mental Model** [104]. *Stress* or poor *information use* may also cause poor situational awareness leading to the delay of actions. Similar to the CFM D1, this CFM may also be the way in which an Information-phase failure manifests [73]. If the operators have incomplete or incorrect information, they may decide to delay their actions to wait for additional information to become available.

Table B.6: Mechanism and PIFs identified as causing CFM D6: "Delayed Action".

Mechanism/PIF	Substantiation
Mental Model	"If, with equally good SA, the person has a low level of confidence in that SA, however, they most likely will not act on it, choosing to gather more information or behave protectively, and thus be ineffective." [104]
Stress	"Indecision, hesitation, failure to sense (hear, see, feel), failure to act, over-reaction or under-reaction should be expected when the SNS (sympathetic nervous system) is triggered." [109]
Information Use	"...may delay action... lack of confidence can contribute to indecisiveness and delays that produce poor outcomes. This illustrates the importance of communicating information reliability so that people can properly calibrate their confidence level." [104]

The CFM D7, "Inappropriate Strategy Chosen", is caused by the mechanisms of **Skills** [113] and **Mental Model** [73]. It can also be caused directly by *overloading*, *stress*, and *coordination* issues. This CFM is the result of an incorrect choice made, which may be due to a lack of knowledge or the operator's confidence in their knowledge base to address the task at hand.

Table B.7: Mechanisms and PIFs identified as causing CFM D7: "Inappropriate Strategy Chosen".

Mechanism/PIF	Substantiation
Skills	"People who are highly experienced with a task tend to process information at the skill-based level, reacting to the raw perceptual elements at an automatic, subconscious level; without the need to interpret and integrate cues or consider possible alternate actions but instead responding to cues and patterns that are already associated with actions." [113] Highly skilled individuals are more likely to have action scripts associated with the correct strategy.
Mental Model	"When an individual has a well-developed mental model... expectations regarding future states... a direct, single-step link between recognized situation classifications and typical actions." [73] Similar to the Skills mechanism, individuals with suitable mental models are also more likely to be able to predict future states of the system and are thus better able to choose a correct strategy.
All Loads	"As situations degrade, risk and time pressure may rise. These conditions may limit the decision maker's ability to project the situation into the future and mentally simulate the consequences of a course of action." [81]
Stress	"Stress can interfere with the retrieval of multiple hypotheses, and constrains working memory capacity, thus limiting evaluation of options... Stress may interfere with evaluation and recognition of the inappropriateness of wrong responses, e.g. shutting down the wrong engine." [81]
Coordination	"In the complex case, the diagnosis time went considerably up when the quality of the team cognition was bad... The extended time for diagnosis in the complex case could be seen in that the crews had difficulties interpreting and following the procedures, since there was a mismatch between procedures and the plant situation." [105, 110] "The conclusion was that the more complex the tasks get, the more does (bad) teamwork impact performance." [105, 110] "In addition to temporal pressures, high cognitive load (e.g., due to uncertain, dynamic and excessive information) may reduce sharing unique information and result in poor task performance (Stasser and Titus 1987)." [74] With degraded communication, and thus poor coordination, task performance is degraded due to issues with choosing and implementing a strategy. This is worsened for complex scenarios.

The CFM D8, “Premature Termination of Procedure”, similar to the previously discussed CFM A4, was also added to the Phoenix CFM framework to address a gap during its application to ex-control room actions. Like CFM D6 (“Delay Action”), this CFM occurs when a correct assessment of system needs is made but the action is not carried out to the proper extent. The crew may believe that the system state has stabilized due to an incorrect **perceived urgency/severity** level [114], and so they terminate the procedure prematurely. An example of this might be the construction of a sandbag wall in preparation for a flooding event, and not building the wall high enough. Additionally, the crew may be aware that the actions need to be executed, but they have a faulty **mental model** [83] of predicted system dynamics.

Table B.8: Mechanisms identified as causing CFM D8: "Premature Termination of Procedure".

Mechanism/PIF	Substantiation
Perceived Urgency/Severity	"Thus, for example, in predicting the future temperature of a process on the basis of historical trend data of the exponential growth, people would be likely to underestimate its future values... This is the inference that most exponentially increasing quantities do eventually encounter self-correcting mechanisms that slow the rate of growth. For example, exponential population increases will encounter natural (through disease) or artificial means (i.e., of birth control) to lower the rate of growth. Exponentially increasing temperatures will often trigger fire extinguishing efforts, or opening pressure relief valves that will reduce the rate of growth. So, the long-term memory of experience will lead the decision maker—accurately—to infer that the rapidly growing quantity will eventually slow its rate of growth... At the same time, other research indicates that people (e.g., stock analysts) may sometimes be overly risky or extreme in their projection of quantities that are not exponentially growing as above (De Bondt & Thaler 2002), leading to an overreaction in their trading (e.g., choice) behavior." [114]
Mental Model	"The evaluation process relies heavily on mental simulation to help assess the hypothesis, action, or plan under consideration (Orasanu 1993). In this process, the decision maker searches the environment for further data (as in medical diagnosis). The use of cognitive simulations to generate ideas about additional cues to be obtained would explain why people tend to look for confirming evidence. This is because only the generation and running of 'false' mental models would yield disconfirming cues to search for. This seems unlikely to occur in many situations, especially those with a time urgency." [83] With a poor mental model, the operator will be unable to correctly simulate future conditions. They may deem it unnecessary to continue actions, when it is in fact necessary.

B.2 Substantiation of Decision Mechanism Causes

The mechanism of **perceived urgency/severity** is the operator's appraisal of the potential negative consequences associated with poorly handling the scenario. With a high *perceived decision weight* or perceived individual responsibility, or extreme *overloading*, the perceived severity is defined as being high. Furthermore, when the task complexity is high, leading to high perceived severity, it becomes likely that the operator makes a decision without properly deliberating.

Table B.9: PIFs identified as causing the mechanism of Perceived Urgency/Severity for Decision-making.

PIF	**Substantiation**
Perceived Decision Weight	Forms a definitional dependency idiom with the mechanism [35].
All Loads	"As situations degrade, risk and time pressure may rise. These conditions may limit the decision maker's ability to project the situation into the future and mentally simulate the consequences of a course of action." [81] Under a high workload, the operator will feel significant pressure to make a quick, accurate decision, leading to the perception of a severe situation. When the operator is under too much pressure, their accuracy will become limited.
Task Complexity	"Pierce (1996) found that people over-rely on rapid, intuitive decisions rather than perform the more difficult deliberate analyses... decision aids might support human decision making by counteracting this 'shortcut' or satisficing tendency—at least when it is important and there is ample time for analytical processing (e.g., life-threatening decisions)." [83] Under highly complex, highly severe situations, there is a tendency to make quick decisions. This can be dangerous–without deliberation in situations where it is necessary to gain a full appraisal of the situation before acting, poor decisions are likely.

The mechanism of **skills** refers to the operator's skill at assessing a situation and determining the appropriate course of action. Decision-making skills can be trained or acquired, and improve with high amounts of *training*, *familiarity*, *knowledge and experience*, and suitable *information use*. Skills are also limited by PIFs outside of the operator's control, such as their

cognitive *personal and physical abilities*, and their *stress* levels.

Table B.10: PIFs identified as causing the mechanism of Skills for Decision-making.

PIF	**Substantiation**
Personal and physical abilities	"Experts also tend to develop superior skills for acquiring and processing information in their environment that must be learned or trained... Endsley and Bolstad (1994) found that military pilots with better SA were better at attention sharing, pattern matching, spatial abilities, and perceptual speed... Gugerty and Tirre (1997) found evidence that people with better SA performed better on measures of working memory, visual processing, temporal processing, and time-sharing ability." [104]
Operator's knowledge base: familiarity, training, knowledge	"experts spend more time trying to understand the problem and construct a mental model of the task environment than novices..." [74] "these behavioral patterns are consistent with the tendency not to seek additional information due to [misapplied] expertise... their expertise may have guided them not to continue, as conserving resources would lead to greater success" [112] "a number of areas have been identified which can be addressed through training to improve SA (Endsley and Robertson 2000b) ... task management, development of comprehension, projection and contingency planning, information seeking and self-checking activities, basic and higher order cognitive skills, training of team SA skills" [104] "Training is a very important factor that can compensate for suboptimal design of human-system interfaces and human–computer interaction." [105, 111] Skills crucial to decision-making can be learned through training.
Continued on next page	

PIF	Substantiation
Information use	"People are able to use information in their judgments on the reliability of sensors in order to weight the data presented, although they may not do so optimally (Berg 1990; Bolstad & Endsley 1998; Montgomery & Sorkin 1996)." [104] "A common factor for the lower scoring crews was that they did not show appropriate patience in watching the effects of the procedure actions... misleading information about whether or not the preceding procedure had been effective" [112]
Stress	"Each form of stress can be equally debilitating to precision-based skills. A combination of any two stressors can magnify the deterioration of precision skills. Combinations of three or more stressors can trigger an immediate cascade of all precision skills. Further, survival stress trumps all combinations of stressors." [109] "Several types of stress factors exist that may act to influence SA, including... importance or consequences of events... prestige, job advancement or loss, mental load, and time pressure." [73]

The mechanism of **bias** is similar to the bias mechanism for information-gathering tasks. Internal biases may stem from operators' *training*, *familiarity*, and *knowledge*. With poor *procedure quality*, individuals are predisposed to follow familiar actions instead of new procedures that are difficult to understand. *Perceived decision weight* can also lead to expectations of the scenario progression. In the case of low perceived decision responsibility, the operator will be biased towards taking as few actions as possible.

Table B.11: PIFs identified as causing the mechanism of Bias for Decision-making.

PIF	Substantiation
Procedure quality	"Another common problem is a failure to generate the correct solution plan, even when the problem representation is clear. Sometimes this is due to fixation on previous plans that worked in the past. It is often difficult to search long-term memory for alternative solutions because memory activation tends to not spread to "distant" or weakly associated concepts. People are also prone to showing 'functional fixedness', looking at the functionality of objects only in terms of their normal use." [83] With poor procedure quality, the operator will be biased against using the procedures at all, instead falling back on typical patterns.
Continued on next page	

PIF	Substantiation
Perceived decision weight	"Reliance refers to the degree to which an operator will trust in the automation... Compliance, on the other hand, refers to whether the operator will immediately respond to an automated alert and follow its recommendations... People who are more prone to be complacent have been shown to have lower SA... other factors influencing the degree to which people will trust in or rely on automation include time constraints and subjective workload, how much people trust in their own abilities to do the task, and effort needed to engage with the automation (Lee and See 2004)." [104] In situations where automation is present, or the operator is not fully responsible for their decision, they are more likely to become reliant, compliant, and complacent with the decisions made for them. Thus, they become less aware of the scenario, and are biased against their own decisions.
Operator's knowledge base: familiarity, knowledge	"...people act wisely when they use their experience or the experience of others." [80] "assess the frequency of a class or the probability of an event by the ease with which instances or occurrences can be brought to mind... the reliance on availability leads to predictable biases... biases due to the retrievability of instances, due to the effectiveness of a search set, due to imaginability..." [108] "behavioral patterns are consistent with the tendency not to seek additional information due to [misapplied] expertise..." [112] High levels of knowledge and familiarity can cause biases, but poor knowledge base can lead to other biases such as the information availability heuristic.

The mechanism of **prioritization**, also caused by **bias**, is similar to the prioritization mechanism for information-gathering failures. If the operator has an exceedingly high *perceived decision weight* and *perceived urgency* for the scenario, they will be unable to properly prioritize tasks. With poor *morale, motivation, and attitude*, the operator will not be motivated to pursue goals in the correct order. Finally, organizational *safety culture* may cause the operator to value profit, timeliness, or other mission objectives over more important safety objectives.

Table B.12: PIFs identified as causing the mechanism of Prioritization for Decision-making.

PIF	**Substantiation**
Perceived situation characteristics: decision weight, urgency	"Under conditions of severe overload, people may shift their operational strategy. Instead of trying to perform all operations manually and using the automation as a backup, they may rely on the automation to prioritize their tasks, attending only to the problems alerted by the automation." [104]
Safety Culture	"Pilots may be willing to take a risk with safety (a possible loss) to arrive on time (a sure benefit) ... Meeting organizational and social goals often appears to outweigh safety goals, especially in ambiguous conditions." [81] "Organizational managers have a variety of means at their disposal: administrative controls (prescriptive rules and procedures), individual controls (selection, training and motivators), group controls (supervision, norms and targets) and technical controls (automation, engineered safety features, physical barriers) ... the balance between them is very much a reflection of the organizational culture." [103] With a poor safety culture, the operator may incorrectly prioritize other goals, leading to degraded conditins that give rise to an initiating event or exacerbate the situation at hand.
Morale, motivation, and attitude	Forms a definitional dependency idiom with the mechanism [35].

The mechanism of **mental model** refers to the operator's internal appraisal of the scenario's causes and effects. With sufficient *training*, *familiarity*, and *knowledge*, the operator should be able to correctly diagnose off-normal conditions. Poor *information use* or low-quality *communication* from team members may lead the operator to draw incorrect conclusions. *Overloading* may also lead them to consider a more limited set of options for diagnosis. This mechanism is not a PIF from the hierarchy used, but resurfaces in cognitive literature to describe the process and product of an individual's situational assessment.

Table B.13: PIFs identified as causing the mechanism of Mental Model for Decision-making.

PIF	Substantiation
Operator's knowledge base: training, familiarity, knowledge	"This provision of mental models [default values] allows experts to have access to reasonable defaults that provide more effective decisions than those of novices who simply have missing information (or poorer defaults)." [73] "experts spend more time trying to understand the problem and construct a mental model of the task environment than novices..." [106] "The RPD [recognition primed decision-making] model underscores the crucial role of domain-specific knowledge or experience in proficient decision making: No step in the model can be executed effectively without such knowledge... Rather, these features include making finer distinctions and setting plausible goals within situations, drawing better analogies among situations, imagining richer potential developments, producing appropriate action, and recognizing inappropriate action more quickly." [80] "This provision of mental models [default values] allows experts to have access to reasonable defaults that provide more effective decisions than those of novices who simply have missing information (or poorer defaults)." [73]
All Loads	"As situations degrade, risk and time pressure may rise. These conditions may limit the decision maker's ability to project the situation into the future and mentally simulate the consequences of a course of action." [81]
Information use	"... The situations were not recognized as ones that should trigger a change of course of action, due to the ambiguity of the cues" [81] "... default values for certain features of a system can be used if exact current values are not known... when more details become available, their SA becomes more accurate, possibly leading to better decisions, but they are still able to make reasonable decisions without perfect information." [73] "A common factor for the lower scoring crews was that they did not show appropriate patience in watching the effects of the procedure actions... misleading information about whether or not the preceding procedure had been effective" [112]
Continued on next page	

PIF	Substantiation
Communication quality	"[The highest scoring] crew decided early on that they needed information from external sourced information. Therefore, they communicated frequently... [the lowest scoring crew] decided to not spend resources on communicating with a field operator regarding alternative information" [112] Communication within a team and with external sources is integral to the development of a rich and useful mental model.

Appendix C: Substantiation of Action Execution CFM & Mechanism Causes and Effects

C.1 Substantiation of Action Crew Failure Mode Causes

Each directed arc in the Bayesian network is based on a real causal relationship between the nodes which is documented in the human factors and cognitive psychology literature. The literature sources used are enumerated in Table 5.1. The first section of this appendix provides a source and quote for each mechanism's influence, and the subsequent section in turn details the influences of PIFs on the mechanisms.

The A-phase CFM A1, "Incorrect Timing", is caused by four main human failure mechanisms: **Coordination** [119], **Procedure Error** [120], **Skills** [120], and **Attention** [114]. It may additionally be caused by a poor *human-system interface (HSI)* or poor *tool availability*. This is one of the more complicated Action-phase crew failure modes, as it can be caused by all of the identified Action-phase failure mechanisms. A deficient HSI can be the cause of delayed actions due to the extra effort needed to interact with the system. For situations where the required tools are not available to the operator, this CFM is defined as being in a failure state using the definitional dependency idiom.

Table C.1: Mechanisms and PIFs identified as causing CFM A1: "Incorrect Timing".

Mechanism/PIF	Substantiation
Coordination	"Increasingly researchers have found support for another property, that of TMM [team mental model] quality. TMM sharedness is defined as the common understanding of 'expected collective behavior patterns' (Ayoko and Chua 2014) as well as shared knowledge and expectations of team members during the group work... It is also important to account for time in understanding the dynamics of interactions (Raghavan, Steeg, Galstyan, and Tartakovsky 2014) and, therefore, the development of both TMM sharedness and quality properties." [119] Without a quality shared understanding of the situation, it will require more time to select a course of action, which may lead to its poorly timed execution.
Procedure Error	"Since memory is fallible, particularly when interruptions occur, written procedures and checklists are used in many operational environments to reduce operator workload and decrease the probability of errors. However, since retrieving the appropriate page from a manual may take an unacceptable amount of time in an emergency situation, critical items (e.g., the first few steps in an emergency checklist) are committed to memory." [120] Misapplication of procedures in emergency scenarios may lead to a poor understanding of the timing of situation dynamics, leading to incorrectly timed actions.
Skills	"Different components of a complex task may have different functional priority with respect to the overall goals of the task or temporal priority (if they are not completed by a deadline, they can no longer be performed at all). Operators may adopt different resource allocation and scheduling strategies to satisfy external instructions or personal goals... Shifting task priorities and resource allocation policies in response to transient changes in task demands is particularly difficult (Tsang and Wickens 1988) and may inhibit an effective response to workload transition." [120] The operator may be unskilled at managing the task priorities, leading to delayed actions or premature actions.
Continued on next page	

Mechanism/PIF	Substantiation
Attention	"First, poorer performance by those who are stressed by job-related factors (e.g., poor working conditions, inequitable wages) may be related to the lack of attention, resources, or effort put into the job (i.e., low motivation). In contrast, the greater safety hazards of some who suffer life event stress may be related to distraction or diversion of inattention; that is, attention diverted from the job-related task at hand to thinking about the source of stress (Wine 1971)." [83] "Multitasking can also be described as dividing attention between tasks rather than between information channels" [114] Poorer performance may take the form of delayed actions, perhaps due to starting a task, getting distracted, and realizing too late that it was left unfinished.
HSI	"Performance may be degraded under normal operating conditions when display–control relations are incompatible or inconsistent with population stereotypes." [121] "Hence, these procedures must be clear and simply phrased and should be as consistent as possible with operations. Ideally procedural instructions of what to do should be redundantly coded with speech as well as with print or pictures, should avoid arbitrary symbolic coding, and ...avoid negatives." [114] "Responses are faster and more accurate when there is a spatial correspondence between the position and direction of movement of stimuli and control responses, [population stereotypes, and modalities] ... If information must be transformed from one format or orientation to another to complete a task, workload and response time are increased. For example, King et al. (1989) found that response time and subjective workload increased when mental rotation was required for comparison." [120]
Continued on next page	

Mechanism/PIF	Substantiation
Tool Availability	Forms a definitional dependency idiom with the CFM [35].

The CFM A2, "Incorrect Operation", is caused by three different human failure mechanisms: **Procedure Error** [120], **Skills** [120], and **Attention** [83, 114]. It may be caused by other factors relating to the operator's interaction with the scenario, such as the *human-system interface (HSI)*, the *quality of the tools* that are available, and their *stress levels*. If only poor-quality or broken tools are available to the operator, then this CFM is defined as being in a failed state using the definitional dependency idiom. Operator skill is particularly important to this CFM as well, which can be caused by a lack of training or simply a lack of experience in implementing the action.

Table C.2: Mechanisms and PIFs identified as causing CFM A2: "Incorrect Operation".

Mechanism/PIF	Substantiation
Procedure Error	"Since memory is fallible, particularly when interruptions occur, written procedures and checklists are used in many operational environments to reduce operator workload and decrease the probability of errors. However, since retrieving the appropriate page from a manual may take an unacceptable amount of time in an emergency situation, critical items (e.g., the first few steps in an emergency checklist) are committed to memory." [120]
Skills	"Error will increase if control inputs are too little, too late, or in the wrong direction, if operators overcontrol (i.e., make control inputs that are faster or larger than necessary), they create additional workload for themselves (i.e., they must compensate for the errors that they generate) and run the risk of destabilizing the system. Thus, although minimizing error is the goal of most control activities, smoothness and stability may be equally important." [120]
Continued on next page	

Mechanism/PIF	Substantiation
Attention	"First, poorer performance by those who are stressed by job-related factors (e.g., poor working conditions, inequitable wages) may be related to the lack of attention, resources, or effort put into the job (i.e., low motivation). In contrast, the greater safety hazards of some who suffer life event stress may be related to distraction or diversion of inattention; that is, attention diverted from the job-related task at hand to thinking about the source of stress (Wine 1971)." [83] "Multitasking can also be described as dividing attention between tasks rather than between information channels" [114] In addition to poor timing, a lack of executive attention may also lead to performance issues in the form of poorly executed operation.
HSI	"More interesting is the fact that population stereotypes are found when there is no direct relation between the display and control... clockwise to right or up principle, Warrick's principle, clockwise to increase principle, scale-slide principle... Performance may be degraded under normal operating conditions when display–control relations are incompatible or inconsistent with population stereotypes." [121] Performance degradation may happen when the operator must interact with a poorly designed HSI: i.e., an HSI that does not conform to their inherent expectations of how it should behave.
Stress	"Hockey (1986) concludes that there is a general effect of noise and/or anxiety stress on the speed-accuracy tradeoff, shifting performance to a less accurate but not slower level. In their study of pilot judgment, Wickens et al. (1988) found that judgments were less accurate but not necessarily slower under the combined stress effects of noise, time pressure, and threat of loss of income. The tendency of those under the stress of an emergency to shift performance from accurate to fast (but error prone) responding has been cited as a concern in operator response to complex failures in nuclear power control rooms." [120] High stress can cause accuracy errors due to rushing.
Continued on next page	

Mechanism/PIF	Substantiation
Tool Quality	Forms a definitional dependency idiom with the CFM [35].

The CFM A3, "Wrong Component", is caused by just one mechanism, **Attention** [83, 114]. It may also be caused by high amounts of *stress* to the operator or a poor *HSI*. For this CFM to occur, they would be taking the correct action at the right time, but on the wrong component. The resulting error is a simple cognitive slip, which may manifest due to divided attention or the presence of extreme stressors.

Table C.3: Mechanisms and PIFs identified as causing CFM A3: "Wrong Component".

Mechanism/PIF	Substantiation
Attention	This error is substantiated by [83, 114] in a similar way to the previous two CFMs (A1 and A2): performance degradation due to a lack of executive attention may take the form of absently acting on the wrong component.
Stress	This CFM is substantiated by [120] in a similar way to CFM A2, "Incorrect Operation": rushing may also lead the operator to perform the correct actions, but on the wrong component.
HSI	This CFM is substantiated by [121] in a similar way to CFM A2, "Incorrect Operation".

The CFM A4, "No Action Taken", was introduced to fill a gap identified after applying the Phoenix HRA method to ex-control room actions. Although the CFM A1 "Incorrect Timing of Action" contains failure scenarios where crews forget altogether to take the required action (i.e., infinite delay time), it is useful to make a distinction between a delay and an action that will not be taken. This could be due to lack of prepared *tool availability*, or the mechanisms of **Attention** [114] and **Coordination** [119]. "No Action Taken" refers to a scenario in which the correct information is gathered and the correct decision is made, but the actions are simply not carried out. For example, a crew might go to pick up equipment from an external location, but the vehicle is out of fuel (lack of resource availability), or the task is not properly assigned to an individual so it is assumed to be done, but never ends up being completed (poor coordination).

Additionally, the individual operator may simply forget that the task was assigned due to extreme load and stressful circumstances surrounding the situation (attentional error).

Table C.4: Mechanisms and PIFs identified as causing CFM A4: "No action Taken".

Mechanism/PIF	**Substantiation**
Attention	"Working memory loss describes just that. Under stress, people appear to be less capable of using working memory to store or rehearse new material or to perform computations and other attention-demanding mental activities (Wickens et al. 1991; Stokes & Kite 1994; Hockey 1986)." [83]
Coordination	"In other words, it is 'being on the same page' in regard to the task. The TMM quality refers to the degree to which the TMM reflects the right working model of the task... TMM sharedness may be more important to team effectiveness as shared expectations lead toward more cohesive work in accomplishing tasks." [119]
Tool Availability	Forms a definitional dependency idiom with the CFM [35].

C.2 Substantiation of Action Mechanism Causes

The mechanism of **team coordination** involves delegating and completing responsibilities as part of a cohesive operations unit. Errors in coordination are mostly brought about by issues with organizational and team-based factors, including *cohesion*, *role awareness*, *communication availability and quality*, *safety culture*, and *staffing*. Issues with safety culture may lead to poor staff coverage or poor personnel availability. A lack of role awareness may lead operators to believe that an issue is being addressed when, in fact, it is not. Without team cohesion and good communication, coordination of team members will also be poor.

Table C.5: PIFs identified as causing the mechanism of Coordination for Action-taking.

PIF	**Substantiation**
Cohesion	"One reason that teams often perform below the initial expectations of management is that there has been an inadequate amount of training and team-building in advance." [83]
Continued on next page	

PIF	Substantiation
Role Awareness, Communication Availability	"Individual task knowledge development: Defined as individual team members collecting and processing data or information, or responding to clarification requested by other team members about their assigned task... The externalized processes are the cognitive activities of exchanging task knowledge with the other team members through verbal and non-verbal communication." [74] If the crew members are not aware of their roles (or the others' roles) on the team, they will not be able to effectively communicate vital task information to each other.
Safety Culture	"Organizational managers have a variety of means at their disposal: administrative controls (prescriptive rules and procedures), individual controls (selection, training and motivators), group controls (supervision, norms and targets) and technical controls (automation, engineered safety features, physical barriers) ... the balance between them is very much a reflection of the organizational culture." [103] Depending on the organization's safety culture, the environment may be conducive to collaboration or it may be actively hostile.
Staffing	"...any examples of performance problems in coalitions have been documented that are related to unanticipated differences in work styles or organizational procedures (such as Siegel 1998)" [74] "One reason that teams often perform below the initial expectations of management is that there has been an inadequate amount of training and team-building in advance" [83] If the members of a team are hastily chosen due to management issues, or if they are poorly suited to working with other, there will likely be issues with their coordination.

The mechanism of **procedure error** refers to incorrect following of the correct procedure. Because operators may be recalling procedure steps from memory due to a lack of *procedure availability*, *training* levels are important in whether or not they succeed. A poor *HSI* or poor *safety culture* can also lead to the operator being unable or unwilling to adhere to the procedures with the correct levels of precision.

Table C.6: PIFs identified as causing the mechanism of Procedure Error for Action-taking.

PIF	Substantiation
Safety Culture	"Organizational managers have a variety of means at their disposal: administrative controls (prescriptive rules and procedures), individual controls (selection, training and motivators), group controls (supervision, norms and targets) and technical controls (automation, engineered safety features, physical barriers) ... the balance between them is very much a reflection of the organizational culture." [103]
Procedure Availability	"New procedures and regulations, however, should be introduced with caution and careful analysis of their potentially less obvious side effects. Sometimes the costs (and burden on the operator) to implement may far outweigh the benefit." [120] New procedures may make the information less available to the operator, due to differences between expectation and reality. This leaves them prone to incorrect interpretations. However, a lack of procedures, or difficult to find procedures, guarantee that there will be issues in following the procedure.
Training	"Schemata and mental models are developed as a function of training and experience in a given environment. A novice in an area may have only a vague idea of important system components and sketchy rules or heuristics for determining the behavior he or she should employ with the system." [73] "Because abnormal events tend to be quite infrequent, the correct responses may not be familiar." [81] Insufficient training may leave the operator unable to properly employ the correct procedure.
HSI	"Performance may be degraded under normal operating conditions when display–control relations are incompatible or inconsistent with population stereotypes." [121] With incompatible HSI, the operator may believe they are implementing the response correctly, but in fact are operating the controls incorrectly.

The mechanism of **skills** refers to the operator not having the necessary skills to correctly carry out the action. This can be due to a operator's lacking knowledge base, including their *training*, *familiarity*, and general *knowledge and expertise*. A poor training program may be caused by a poor *safety culture*. The operator may also have insufficient *personal and physical abilities* to complete the task.

Table C.7: PIFs identified as causing the mechanism of Skills for Action-taking.

PIF	Substantiation
Safety Culture	"Organizational managers have a variety of means at their disposal: administrative controls (prescriptive rules and procedures), individual controls (selection, training and motivators), group controls (supervision, norms and targets) and technical controls (automation, engineered safety features, physical barriers) ... the balance between them is very much a reflection of the organizational culture." [103]
Operator's knowledge base: knowledge, training, familiarity	"... it does no good for a cue to be present at retrieval time if no link between the cue and the target has been established... In general, retrieval cues need not be environmental—they can be drawn from long-term knowledge about the task and thus be part of the problem solver's internal mental context" [123] "Because abnormal events tend to be quite infrequent, the correct responses may not be familiar." [81]
Personal and physical abilities	"Another common screening device is the standardized test. These tests measure cognitive and physical abilities and personality. Other tests provide a prospective employer with a work sample. For example, if you apply for a cashier's position in a store, you will probably be tested on your arithmetic skills. Of all these tests, it turns out that, for people with no experience, 'the most valid predictor of future performance is general cognitive ability'" [121]

The mechanism of **attentional failure** for action-taking is caused mainly by situation and stressor-based PIFs. Unlike the attentional failure mechanism for information-gathering tasks, this type of attention is executive attention, or the ability to focus resources on the execution of a task. Executive attention is used to focus on carrying out steps of a procedure and heavily relies on working memory. Similar resources to selective attention are required, and *fatigue*, *stress*, and *overloading* can cause attentional failures. To execute actions properly, a moderate *perceived urgency* and reasonably good *external environmental conditions* must be present.

Table C.8: PIFs identified as causing the mechanism of Attention for Action-taking.

PIF	**Substantiation**
Fatigue	"this study showed a reduction in both the global performance, i.e., with slower response speed and less accurate but more omitted responses, and the overall correction rate after sleep deprivation" [122] Sleep deprivation, unsurprisingly, leads to degradation in executive attention to a task.
Stress	"...the greater safety hazards of some who suffer life event stress may be related to distraction or diversion of inattention; that is, attention diverted from the job-related task at hand to thinking about the source of stress (Wine 1971)." [83]
All Loads	"Another implication of the motor program concept is that it should take longer to program more complex movements. It is true that response times to initiate a movement are longer for more complex movements (Henry & Rogers 1960; Klapp 1977). Thus, the speed with which a person can react to a stimulus is directly related to the complexity of the movements that must follow." [121] Executive attention is impaired by higher task loads. "As the demands imposed on short-term memory are increased, tasks are perceived as being more difficult. Thus, tasks that require short-term memory often impose greater workload than do tasks that require retrieval of information from long-term memory. For example, Berg and Sheridan (1984) found that pilot workload was significantly higher and flying performance was significantly worse for flights that imposed short-term memory demands than for those that required long-term memory." [120]
Continued on next page	

PIF	Substantiation
Perceived Urgency	"One may be relatively more confident or even overconfident in one's ability to deal with the hazard. Finally, if people appraise that they are more in control of the situation, they are less likely to experience stress than if they feel that other agents are in control (Bowers, Weaver, & Morgan 1995)." [83] With a low perceived urgency, individuals are less likely to pay attention due to overconfidence that they are in control of the situation.
External Environment	"Inadequate environmental conditions are major contributors to stress. Stress is also produced by a variety of other factors, including the social environment, task demands, and long-term confinement. High levels of stress can result in illness and poor performance." [121] "Many stressors simply impose a distraction and thus divert selective attention away from task-relevant processing. Loud or intermittent noises or even the conversation at a nearby table at the library will serve as a source of such distraction (Baldwin 2012)." [114]

Appendix D: Conditional Probability Tables and Multipliers Used for Parameterization

To ensure that the BN models were capable of quantifying the probabilities of CFMs, as well as degraded PIFs and the presence of human failure mechanisms, it was necessary to parameterize the networks with a variety of data and quantitative models. The combinatorial explosion of some nodes' CPTs made this a nontrivial task, with many nodes having dozens of combinations of parent node states, and some CFMs having hundreds. The SACADA database, with nearly 24,000 entries, covered many of these combinations of PIFs and PIF states. However, for PIFs or states for which no data was collected, we parameterized the nodes through additional means.

By applying a multiplier to the probability of a "degraded" or otherwise adversely affected child node in the presence of a degraded parent node, it is possible to double, triple, or quadruple the number of values populated in the CPT. For example, the SACADA database considers both the CFM D4 "Procedure Step Skipped" and the mechanism of bias. There is not sufficient data to quantify the probability of CFM D4 occurring conditioned on the presence of bias, so this value must be derived from the literature. A multiplier of 1.607 is obtained from Blissett et al [144] to uniformly increase the probability of CFM D4 in the presence of bias. This multiplier can be applied to all probabilities of D4 = Yes conditioned on Bias = Yes, holding all other conditions

equal. Then, all probabilities of D4 = No conditioned on Bias = Yes can be obtained by taking the complement of this value. As Bias is a binary node, the procedure effectively doubles the amount of values populated in the CPT for CFM D4.

For the tables in this appendix containing the probabilities and conditional probability multipliers of each BN node, the full list of parent node states is omitted for brevity. It can be assumed that all parent PIFs are in a nominal state aside from the PIF(s) specified in the "Condition" column of the table. When no PIF states are specified, the given probability is the CPT value for that node with all parent node states nominal.

D.1 Information-Gathering Probabilities and Multipliers

For the information-gathering network, it was possible to obtain most of the marginal probabilities from the SACADA database and NUREG-6949, combining values with Bayesian updating where necessary. For the conditional probabilities, the Definitional dependency idiom [35] was applied where the parent PIF state was deemed to have a certain effect on the child node. Many multipliers, especially for the CFMs, were derived from the IDHEAS-ECA method. As some CFMs (I3, I6, I7, and I8) were not included in the SACADA database, it was necessary to obtain these base probabilities from other HRA methods, including Phoenix, IDHEAS-ECA, and CREAM.

Table D.1: Marginal probabilities of Bayesian network nodes for the information-gathering network.

Node	State	Probability	Citation
Task Complexity	Very High	0.159	[84]
	Moderately High	0.341	
	Nominal	0.500	
Non-task Load	High	0.209	[87]
	Nominal	0.791	
Time Load	Inadequate	0.0435	[84, 87]
	Nominal	0.7840	
	Extra	0.1725	
Safety Culture	Poor	0.023	[84]
	Nominal	0.819	
	Good	0.158	
Supervision	Poor	0.00018	[87]
	Nominal	0.99982	
HSI	Missing, Misleading	0.05822	[84, 87]
	Poor	0.14046	
	Nominal	0.79801	
	Good	0.00331	
Training, Knowledge, Experience, and Familiarity	Low	0.09424	[87]
	Nominal	0.90576	
MMA	Poor	0.00035	[87]
	Nominal	0.99965	
Procedure Avail/Qual	Not Available	0.00052	[84, 87]
	Incomplete	0.06914	
	Poor	0.00312	
	Nominal	0.92722	
External Environment	Degraded	0.0016	[87]
	Nominal	0.9984	
Communication Quality	Poor	0.00225	[87, 145]
	Nominal	0.99775	

Table D.2: Probabilities and associated conditional multipliers of Bayesian network nodes for the decision-making network.

Node (State)	**Multiplier or Probability Value**	**Condition**	**Citation**
Task Load (High)	0.0816	Moderately High Task Complexity	[87]
Task Load (High)	2.5x multiplier	Very High Task Complexity	[1]
Task Load (High)	0.5x multiplier	Nominal Task Complexity	[1]
Staffing (Poor)	1 (deterministic)	Poor Safety Culture	[35]
Staffing (Nominal)	1 (deterministic)	Nominal or Good Safety Culture	[35]
Sensory Limits, PPA (Overloaded)	0.132	Nominal Task Complexity	[87]
Sensory Limits, PPA (Overloaded)	1 (deterministic)	Very High Task Complexity	[35]
Sensory Limits, PPA (Overloaded)	2x multiplier	Moderately High Task Complexity	[1]
All Loads (Very High)	0.0805	Nominal Time Load, Non-task Load, Staffing, HSI; Low Task Load	[87]
All Loads (Moderately High)	0.2861	Nominal Time Load, Non-task Load, Staffing, HSI; Low Task Load	[87]
All Loads (Very High)	1 (deterministic)	Inadequate Time Load, Missing or Misleading HSI	[35]
All Loads (Moderately High)	1 (deterministic)	High Task Load, High Non-task Load, Poor HSI	[35]
All Loads (Moderately High)	1.09x multiplier	Poor Staffing	[146]
All Loads (Nominal)	0.5x multiplier	Good HSI, Extra Time	[1]
Cohesion (Poor)	7.95x multiplier	Low Training, Knowledge, Experience, Familiarity	[147]

Node (State)	Multiplier or Probability Value	Condition	Citation
Cohesion (Poor)	5.585x multiplier	Poor Supervision	[148]
Cohesion (Poor)	1.044x multiplier	Poor Staffing	[146]
Perceived Urgency & Severity (Very High)	0.0964	Moderately High Task Complexity	[87]
Perceived Urgency & Severity (Moderately High)	0.7622	Moderately High Task Complexity	[87]
Perceived Urgency & Severity (Very High)	1 (deterministic)	Very High All Loads	[35]
Perceived Urgency & Severity (Very High)	0.5x multiplier	Nominal All Loads	[16]
Perceived Urgency & Severity (Nominal)	2x multiplier	Nominal All Loads	[16]
Stress (Extreme)	0.00048	–	[84, 87]
Stress (High)	0.01785	–	[84, 87]
Stress (Nominal)	0.98167	–	[84, 87]
Stress (Extreme, High)	4x multiplier	Very High All Loads	[16]
Stress (Extreme, High)	2x multiplier	Moderately High All Loads	[16]
Prioritization (Degraded)	0.198	Moderately High Perceived Urgency/Severity	[149]
Prioritization (Degraded)	0.407x multiplier	Yes Bias	[150]
Prioritization (Degraded)	1.07x multiplier	Poor Safety Culture	[151]
Prioritization (Degraded)	4.83x multiplier	Degraded MMA	[152]
Prioritization (Degraded)	1.43x multiplier	Very High Perceived Urgency/Severity	[153]
Bias (Yes)	0.00202	Very High Perceived Urgency/Severity	[87]
Bias (Yes)	1.30x multiplier	High Training, Knowledge, Experience, and Familiarity	[154]
Bias (Yes)	1.29x multiplier	Moderately High Perceived Urgency/Severity	[155]

Node (State)	Multiplier or Probability Value	Condition	Citation
Bias (Yes)	3.65x multiplier	Nominal Perceived Urgency/Severity	[156]
Attention (Degraded)	0.00029	–	[87]
Attention (Degraded)	1.28x multiplier	Degraded Alertness/Fatigue	[157]
Attention (Degraded)	76.6x multiplier	Overloaded Sensory Limits	[87]
Attention (Degraded)	3.24x multiplier	Moderately High All Loads	[87]
Attention (Degraded)	11.28x multiplier	Very High All Loads	[87]
Attention (Degraded)	7.521x multiplier	High Stress	[87]
Attention (Degraded)	56.57x multiplier	Extreme Stress	[87]
Attention (Degraded)	3.65x multiplier	Yes Bias	[95]
Procedure Error (Yes)	0.0270	–	[158]
Procedure Error (Yes)	1.5x multiplier	Poor Procedures	[17]
Procedure Error (Yes)	2.71x multiplier	Incomplete Procedures	[159]
Procedure Error (Yes)	4.16x multiplier	Missing Procedures	[159]
Procedure Error (Yes)	2.6x multiplier	Low Training, Knowledge, Experience, and Familiarity	[160]
Procedure Error (Yes)	2x multiplier	Very High All Loads	[16]
Procedure Error (Yes)	1.414x multiplier	Moderately High All Loads	[16]
Procedure Error (Yes)	2x multiplier	Poor Safety Culture	[16]
Procedure Error (Yes)	0.8x multiplier	Good Safety Culture	[16]
Coordination (Degraded)	0.00053	–	[87]
Coordination (Nominal)	1.18x multiplier	Good Staffing	[161]
Coordination (Degraded)	1.33x multiplier	Poor Cohesion	[145]
Coordination (Degraded)	1.24x multiplier	Degraded Communication	[145]
Coordination (Degraded)	1647x multiplier	Low Training, Knowledge, Experience, and Familiarity	[147]
I1 Key Alarm Not Responded To (Yes)	0.00107	–	[87]
I1 Key Alarm Not Responded To (Yes)	0.6x multiplier	Yes Bias	[93]

Node (State)	Multiplier or Probability Value	Condition	Citation
I1 Key Alarm Not Responded To (Yes)	467.3x multiplier	Degraded Attention	[87]
I1 Key Alarm Not Responded To (Yes)	2x multiplier	Poor Coordination	[17]
I1 Key Alarm Not Responded To (Yes)	1.1x multiplier	Poor Prioritization	[17]
I1 Key Alarm Not Responded To (Yes)	1.5x multiplier	Poor HSI	[17]
I1 Key Alarm Not Responded To (Yes)	5x multiplier	Missing/Misleading HSI	[17]
I1 Key Alarm Not Responded To (Yes)	0.5x multiplier	Good HSI	[16]
I1 Key Alarm Not Responded To (Yes)	1.7x multiplier	Poor External Environment	[17]
I2 Data Not Obtained (Yes)	0.00164	Moderately High Perceived Urgency/Severity	[87]
I2 Data Not Obtained (Yes)	0.00806	Very High All Loads	[87]
I2 Data Not Obtained (Yes)	2.5x multiplier	Moderately High All Loads	[17]
I2 Data Not Obtained (Yes)	1.5x multiplier	Poor HSI	[17]
I2 Data Not Obtained (Yes)	11.1x multiplier	Missing/Misleading HSI	[87]
I2 Data Not Obtained (Yes)	0.841x multiplier	Good HSI	[17]
I2 Data Not Obtained (Yes)	1.7x multiplier	Poor External Environment	[17]
I2 Data Not Obtained (Yes)	1.5x multiplier	Poor Procedure Quality	[17]
I2 Data Not Obtained (Yes)	2.2x multiplier	Incomplete Procedure Quality	[17]
I2 Data Not Obtained (Yes)	1.1x multiplier	Missing Procedure Quality	[17]
I2 Data Not Obtained (Yes)	2.76x multiplier	Moderately High Perceived Urgency/Severity	[87]

Node (State)	Multiplier or Probability Value	Condition	Citation
I2 Data Not Obtained (Yes)	6.53x multiplier	Very High Perceived Urgency/Severity	[87]
I2 Data Not Obtained (Yes)	1136x multiplier	Yes Bias	[162]
I3 Data Discounted (Yes)	0.0016	No Bias	[17]
I3 Data Discounted (Yes)	1.1x multiplier	Poor Prioritization	[17]
I3 Data Discounted (Yes)	156.25x multiplier	Yes Bias	[17]
I3 Data Discounted (Yes)	43.1x multiplier	Degraded Attention	[163]
I3 Data Discounted (Yes)	40.4x multiplier	Yes Procedure Error	[164]
I4 Decide to Stop Collecting Data (Yes)	0.00169	–	[87]
I4 Decide to Stop Collecting Data (Yes)	0.00686	Moderately High All Loads	[87]
I4 Decide to Stop Collecting Data (Yes)	0.00444	Very High All Loads	[87]
I4 Decide to Stop Collecting Data (Yes)	48.59x multiplier	Degraded Attention	[87]
I4 Decide to Stop Collecting Data (Yes)	23.4x multiplier	Yes Bias	[17]
I4 Decide to Stop Collecting Data (Yes)	380.4x multiplier	Poor MMA	[165]
I5 Data Misinterpreted (Yes)	0.00089	–	[87]
I5 Data Misinterpreted (Yes)	0.3695	Yes Procedure Error	[164]
I5 Data Misinterpreted (Yes)	1035x multiplier	Degraded Attention	[166]
I6 Reading Error (Yes)	0.000846	–	[23]
I6 Reading Error (Yes)	1.66x multiplier	Degraded Attention	[167]
I6 Reading Error (Yes)	1.5x multiplier	Poor HSI	[17]
I6 Reading Error (Yes)	5x multiplier	Missing/Misleading HSI	[17]
I6 Reading Error (Yes)	0.5x multiplier	Good HSI	[16]
I7 Data Miscommunicated (Yes)	0.006	–	[23]

Node (State)	Multiplier or Probability Value	Condition	Citation
I7 Data Miscommunicated (Yes)	2x multiplier	Poor Coordination	[17]
I7 Data Miscommunicated (Yes)	97.92x multiplier	Poor Communication	[168]
I7 Data Miscommunicated (Yes)	1.22x multiplier	Degraded Attention	[169]
I8 Wrong Data Attended To (Yes)	0.001	–	[16]
I8 Wrong Data Attended To (Yes)	1.5x multiplier	Poor HSI	[17]
I8 Wrong Data Attended To (Yes)	5x multiplier	Missing/Misleading HSI	[17]
I8 Wrong Data Attended To (Yes)	0.5x multiplier	Good HSI	[16]
I8 Wrong Data Attended To (Yes)	13x multiplier	Yes Procedure Error	[163]
I8 Wrong Data Attended To (Yes)	2.5x multiplier	Degraded Attention	[170]
I9 Data Checked with Inappropriate Frequency (Yes)	0.00093	–	[87]
I9 Data Checked with Inappropriate Frequency (Yes)	2x multiplier	Poor Coordination	[17]
I9 Data Checked with Inappropriate Frequency (Yes)	1.5x multiplier	Poor HSI	[17]
I9 Data Checked with Inappropriate Frequency (Yes)	5x multiplier	Missing/Misleading HSI	[17]
I9 Data Checked with Inappropriate Frequency (Yes)	0.5x multiplier	Good HSI	[16]
I9 Data Checked with Inappropriate Frequency (Yes)	2.5x multiplier	Moderately High All Loads	[17]
I9 Data Checked with Inappropriate Frequency (Yes)	3.365x multiplier	Very High All Loads	[87]

Node (State)	Multiplier or Probability Value	Condition	Citation
I9 Data Checked with Inappropriate Frequency (Yes)	5.65x multiplier	Yes Bias	[17]
I9 Data Checked with Inappropriate Frequency (Yes)	979x multiplier	Degraded Attention	[171]

D.2 Decision-making Probabilities and Multipliers

As the SACADA database takes a very granular view of the cognitive errors that can occur during operations, many of the probabilities for decision-making errors and PIFs could be found in SACADA. Psychological literature sources were particularly necessary here to disentangle the causes of human failure mechanisms such as prioritization and bias. All CFMs' nominal probabilities were quantified using SACADA except for the new CFM D8 (Premature Termination of Actions), which was found in Jang et al.'s study of advanced control rooms [172].

Table D.3: Marginal probabilities of Bayesian network nodes for the information-gathering network.

Node	State	Probability	Citation
Safety Culture	Poor	0.023	[84]
	Nominal	0.819	
	Good	0.158	
Time Load	Inadequate	0.0595	[84, 87]
	Nominal	0.799	
	Extra	0.1415	
Non-task Load	High	0.263	[87]
	Nominal	0.737	
Task Complexity	Very High	0.0185	[84, 87]
	Moderately High	0.0579	
	Nominal	0.9236	
Continued on next page			

Node	State	Probability	Citation
MMA	Poor	0.00035	[87]
	Nominal	0.99965	
Alertness/Fatigue	Degraded	0.159	[84]
	Nominal	0.841	
HSI	Missing/Misleading	0.00059	[84, 87]
	Poor	0.06500	
	Nominal	0.93277	
	Good	0.00164	
PPA	Degraded	0.0014	[87]
	Nominal	0.9986	

Table D.4: Probabilities and associated conditional multipliers of Bayesian network nodes for the decision-making network.

Node (State)	Multiplier or Probability Value	Condition	Citation
Task Load (High)	0.335	Very High Task Complexity	[87]
Task Load (Nominal)	0.665	Very High Task Complexity	[87]
Task Load (High)	0.155	Moderately High Task Complexity	[87]
Task Load (Nominal)	0.845	Moderately High Task Complexity	[87]
Task Load (High)	0.178	Nominal Task Complexity	[87]
Task Load (Nominal)	0.812	Nominal Task Complexity	[87]
Procedure Quality (Not Available)	0.00052	–	[84, 87]
Procedure Quality (Incomplete)	0.06914	–	[84, 87]
Procedure Quality (Poor)	0.00312	–	[84, 87]
Procedure Quality (Not Available, Incomplete, Poor)	2x multiplier	Poor Safety Culture	[1]
Procedure Quality (Not Available, Incomplete, Poor)	0.8x multiplier	Good Safety Culture	[1]

Node (State)	Multiplier or Probability Value	Condition	Citation
Training, Knowledge, Experience, Familiarity (Low)	0.00079	–	[87]
Training, Knowledge, Experience, Familiarity (Low)	2x multiplier	Poor Safety Culture	[1]
Training, Knowledge, Experience, Familiarity (Low)	0.8x multiplier	Good Safety Culture	[1]
All Loads (Very High)	0.0935	Nominal Time Load, Task Load, Non-task Load	[87]
All Loads (Moderately High)	0.359	Nominal Time Load, Task Load, Non-task Load	[87]
All Loads (Very High)	1 (deterministic)	Inadequate Time Load	[35]
All Loads (Moderately High)	1 (deterministic)	High Task Load or High Non-task Load	[35]
All Loads (Nominal)	1 (deterministic)	Extra Time Load, Nominal/Low Task Load, Nominal Non-task Load	[35]
All Loads (Very High, Moderately High)	0.5x multiplier	Extra Time Load, Nominal Task Load and Non-task Load	[16]
Communication Quality (Poor)	0.5	Very High Task Complexity	[17]
Communication Quality (Poor)	0.05	Moderately High Task Complexity	[17]
Communication Quality (Poor)	0.00293	Nominal Task Complexity	[87]
Information Use (Degraded)	0.00175	–	[87]
Information Use (Degraded)	1 (deterministic)	Poor MMA	[35]
Stress (Extreme)	0.00024	–	[87]
Stress (High)	0.00263	–	[87]
Stress (High)	1 (deterministic)	Degraded Alertness/Fatigue	[35]
Stress (Extreme, High)	2x multiplier	Moderately High All Loads	[16]

Node (State)	Multiplier or Probability Value	Condition	Citation
Stress (Extreme, High)	4x multiplier	Very High All Loads	[16]
Coordination (Degraded)	0.00185	–	[87]
Coordination (Degraded)	5.125x multiplier	Degraded Communication	[173]
Skills (Degraded)	0.00185	–	[87]
Skills (Degraded)	2x multiplier	High Stress	[1]
Skills (Degraded)	27x multiplier	Poor Information Use	[87]
Skills (Degraded)	41.6x multiplier	Poor PPA	[87]
Skills (Degraded)	57.3x multiplier	Low Training, Knowledge, Experience, and Familiarity	[87]
Bias (Yes)	0.00494	–	[87]
Bias (Yes)	0.846x multiplier	Moderately High Perceived Urgency/Severity	[174]
Bias (Yes)	2x multiplier	Very High Perceived Urgency/Severity	[174]
Bias (Yes)	1x multiplier	Low Training, Knowledge, Experience, and Familiarity	[175]
Bias (Yes)	1.094x multiplier	Poor Procedure Quality	[161]
Bias (Yes)	1.197x multiplier	Incomplete Procedure Quality	[161]
Bias (Yes)	1.432x multiplier	Missing Procedure Quality	[161]
Bias (Yes)	1.29x multiplier	High Perceived Decision Weight	[176]
Prioritization (Degraded)	0.00088	High Perceived Decision Weight	[87]
Prioritization (Degraded)	1.18x multiplier	Moderately High Perceived Urgency/Severity	[87]
Prioritization (Degraded)	0.795x multiplier	Very High Perceived Urgency/Severity	[87]
Prioritization (Degraded)	2x multiplier	High Stress	[1]
Prioritization (Degraded)	5x multiplier	Extreme Stress	[1]

Node (State)	Multiplier or Probability Value	Condition	Citation
Prioritization (Degraded)	240.8x multiplier	Poor Safety Culture	[151]
Prioritization (Degraded)	0.8x multiplier	Good Safety Culture	[1, 16]
Prioritization (Degraded)	710x multiplier	Degraded MMA	[152]
Prioritization (Degraded)	1.23x multiplier	Yes Bias	[177]
Prioritization (Degraded)	0.7x multiplier	Low Perceived Decision Weight	[153]
Mental Model (Degraded)	0.00273	–	[87]
Mental Model (Degraded)	0.0037	Moderately High All Loads	[87]
Mental Model (Degraded)	0.0047	Very High All Loads	[87]
Mental Model (Degraded)	61.1x multiplier	Poor Information Use	[87]
Mental Model (Degraded)	73.3x multiplier	Low Training, Knowledge, Experience, and Familiarity	[87]
Mental Model (Degraded)	330x multiplier	Degraded Communication	[178]
D1 Misdiagnosis of Plant State (Yes)	0.00099	–	[87]
D1 Misdiagnosis of Plant State (Yes)	0.00019	Moderately High Perceived Urgency/Severity	[87]
D1 Misdiagnosis of Plant State (Yes)	0.00113	Very High Perceived Urgency/Severity	[87]
D1 Misdiagnosis of Plant State (Yes)	0.5x multiplier	Good HSI	[16]
D1 Misdiagnosis of Plant State (Yes)	9.78x multiplier	Poor HSI	[87]
D1 Misdiagnosis of Plant State (Yes)	50.5x multiplier	Missing/Misleading HSI	[17]
D1 Misdiagnosis of Plant State (Yes)	202x multiplier	Degraded Mental Model	[87]
D2 Procedure Misinterpreted (Yes)	0.00312	–	[87]

Node (State)	Multiplier or Probability Value	Condition	Citation
D2 Procedure Misinterpreted (Yes)	0.05263	Poor Coordination	[87]
D2 Procedure Misinterpreted (Yes)	26.7x multiplier	Degraded Skills	[87]
D2 Procedure Misinterpreted (Yes)	1.25x multiplier	Poor Procedures	[87]
D2 Procedure Misinterpreted (Yes)	5x multiplier	Incomplete Procedures	[17]
D2 Procedure Misinterpreted (Yes)	17x multiplier	Missing Procedures	[17]
D3 Fail to Adapt Procedures (Yes)	0.00068	–	[87]
D3 Fail to Adapt Procedures (Yes)	0.00033	Moderately High All Loads	[87]
D3 Fail to Adapt Procedures (Yes)	0.00111	Very High All Loads	[87]
D3 Fail to Adapt Procedures (Yes)	86.5x multiplier	Poor Mental Model	[87]
D3 Fail to Adapt Procedures (Yes)	3.9x multiplier	Very High Task Complexity	[179]
D3 Fail to Adapt Procedures (Yes)	1.97x multiplier	Moderately High Task Complexity	[179]
D3 Fail to Adapt Procedures (Yes)	1.47x multiplier	Poor Procedures	[17]
D3 Fail to Adapt Procedures (Yes)	5x multiplier	Incomplete Procedures	[17]
D3 Fail to Adapt Procedures (Yes)	17x multiplier	Missing Procedures	[17]
D4 Procedure Step Skipped (Yes)	0.0581	–	[87, 158]
D4 Procedure Step Skipped (Yes)	1.607x multiplier	Yes Bias	[144]
D4 Procedure Step Skipped (Yes)	1.1x multiplier	Degraded Prioritization	[17]
D5 Inappropriate Transfer to Procedure (Yes)	0.00212	–	[87]
D5 Inappropriate Transfer to Procedure (Yes)	1.1x multiplier	Degraded Prioritization	[17]

Node (State)	Multiplier or Probability Value	Condition	Citation
D5 Inappropriate Transfer to Procedure (Yes)	70.5x multiplier	Degraded Mental Model	[87]
D5 Inappropriate Transfer to Procedure (Yes)	1.2x multiplier	High Stress	[17]
D5 Inappropriate Transfer to Procedure (Yes)	134.76x multiplier	Extreme Stress	[87]
D6 Decide to Delay Action (Yes)	0.00141	–	[87]
D6 Decide to Delay Action (Yes)	0.12195	Poor Mental Model	[87]
D6 Decide to Delay Action (Yes)	1.2x multiplier	High Stress	[17]
D6 Decide to Delay Action (Yes)	5x multiplier	Extreme Stress	[1]
D7 Inappropriate Strategy (Yes)	0.00032	–	[87]
D7 Inappropriate Strategy (Yes)	0.00074	Moderately High All Loads	[87]
D7 Inappropriate Strategy (Yes)	0.00095	Very High All Loads	[87]
D7 Inappropriate Strategy (Yes)	312.5x multiplier	Degraded Skills	[87]
D7 Inappropriate Strategy (Yes)	195.3x multiplier	Poor Mental Model	[87]
D7 Inappropriate Strategy (Yes)	150.1x multiplier	Degraded Coordination	[87]
D7 Inappropriate Strategy (Yes)	1.2x multiplier	High Stress	[17]
D7 Inappropriate Strategy (Yes)	5x multiplier	Extreme Stress	[1]
D8 Premature Termination of Action (Yes)	0.00857	–	[172]
D8 Premature Termination of Action (Yes)	1.1x multiplier	Moderately High Perceived Urgency/Severity	[17]
D8 Premature Termination of Action (Yes)	5x multiplier	Very High Perceived Urgency/Severity	[16]

Node (State)	Multiplier or Probability Value	Condition	Citation
D8 Premature Termination of Action (Yes)	5x multiplier	Poor Mental Model	[16]

Parent Node States			Conditional Probability for Perceived Urgency/Severity		
All Loads	Task Complexity	Perceived Decision Weight	Very High	Mod High	Nominal
Very High	Very High	Low	0.463	0.537	0
		High	1	0	0
	Mod High	Low	0.213	0.787	0
		High	1	0	0
	Nominal	Low	0.614	0.322	0.064
		High	0	1	0
Mod High	Very High	Low	0.372	0.628	0
		High	1	0	0
	Mod High	Low	0.168	0.668	0.164
		High	0	1	0
	Nominal	Low	0.513	0.437	0.050
		High	0	1	0
Nominal	Very High	Low	0.265	0.735	0
		High	0	1	0
	Mod High	Low	0.168	0.832	0
		High	0	1	0
	Nominal	Low	0.302	0.571	0.127
		High	0	1	0

Table D.5: Conditional probability table for Perceived Urgency/Severity for the decision-making network. All deterministic (1 or 0) probabilities derived from Paglioni's definitional idiom [35] and all other values calculated from SACADA [87].

D.3 Action Execution Probabilities and Multipliers

The action execution BN, like the previous two networks, also required supplementation of the SACADA database with literature resources and multipliers from the IDHEAS, CREAM, and SPAR-H methods. The majority of these multipliers were derived from the IDHEAS method.

The addition of A-phase CFM A4 (No Action Taken) in the BN framework posed no particular challenge as this is a an error mode quantified in the SACADA database. Although there are fewer nodes in the action execution network, the A-phase CFMs are highly connected to the mechanism and PIF layers, with many conditional relationships to be parameterized.

Table D.6: Marginal probabilities of Bayesian network nodes for the action execution network.

Node	**State**	**Probability**	**Citation**
PPA	Degraded	0.00148	[87]
	Nominal	0.99852	
Task Complexity	Very High	0.00277	[84, 87]
	Moderately High	0.00709	
	Nominal	0.99014	
Safety Culture	Poor	0.023	[84]
	Nominal	0.819	
	Good	0.159	
External Environment	Degraded	0.028	[180]
	Nominal	0.972	
Time Load	Inadequate	0.0859	[84, 87]
	Nominal	0.7981	
	Extra	0.1160	
Non-task Load	High	0.254	[87]
	Nominal	0.746	
Cohesion	Poor	0.112	[145]
	Nominal	0.888	
Role Awareness	Poor	0.345	[181]
	Nominal	0.645	
Communication Quality	Poor	0.00192	[87]
	Nominal	0.99808	
HSI	Missing or Misleading	0.00040	[84, 87]
	Poor	0.00266	
	Nominal	0.49049	
	Good	0.50645	
Alertness/Fatigue	Degraded	0.159	[84]
	Nominal	0.841	
Perceived Urgency/Severity	Very High	0.293	[87]
	Moderately High	0.586	

Continued on next page

Node	State	Probability	Citation
Perceived Urgency/Severity	Nominal	0.121	[87]

Table D.7: Probabilities and associated conditional multipliers of Bayesian network nodes for the action execution network.

Node (State)	Multiplier or Probability Value	Condition	Citation
Training, Knowledge, Experience, and Familiarity (Low)	0.013	Nominal/Good Safety Culture	[182]
Training, Knowledge, Experience, and Familiarity (Low)	0.9978	Poor Safety Culture	[151]
Procedure Quality (Nominal)	0.97970	Nominal/Good Safety Culture	[84, 87]
Procedure Quality (Poor)	0.00522	Nominal/Good Safety Culture	[84, 87]
Procedure Quality (Incomplete)	0.01421	Nominal/Good Safety Culture	[84, 87]
Procedure Quality (Missing)	0.00087	Nominal/Good Safety Culture	[84, 87]
Procedure Quality (Poor, Incomplete, Missing)	49.04x multiplier	Poor Safety Culture	[183]
Tool Availability (Nominal)	1 (deterministic)	Good/Nominal Safety Culture	[35]
Tool Availability (Poor)	0.257	Poor Safety Culture	[184]
Tool Quality (Nominal)	1 (deterministic)	Good/Nominal Safety Culture	[35]
Tool Quality (Poor)	0.257	Poor Safety Culture	[184]
Stress (Nominal)	0.99709	–	[84, 87]
Stress (High)	0.00251	–	[84, 87]
Stress (Extreme)	0.00040	–	[84, 87]
Stress (High, Extreme)	12.8x multiplier	Very High All Loads	[87]
Stress (High, Extreme)	3.58x multiplier	Moderately High All Loads	[87]
Stress (High, Extreme)	199.5x multiplier	Degraded External Environment	[185]

Node (State)	Multiplier or Probability Value	Condition	Citation
All Loads (Very High)	0.07764	Nominal Time Load, Non-task Load, Task Load	[87]
All Loads (Moderately High)	0.39011	Nominal Time Load, Non-task Load, Task Load	[87]
All Loads (Very High)	1 (deterministic)	Inadequate Time Load	[35]
All Loads (Moderately High)	1 (deterministic)	High Task Load or High Non-task Load	[35]
All Loads (Nominal)	1 (deterministic)	Low Task Load, Nominal Non-Task Load, and Extra/Nominal Time Load	[35]
All Loads (Moderately High, Very High)	0.5x multiplier	Extra Time	[16]
Coordination (Degraded)	0.00192	–	[87]
Coordination (Degraded)	1.6x multiplier	Poor Safety Culture	[16]
Coordination (Degraded)	2x multiplier	Poor Role Awareness	[186]
Coordination (Degraded)	1.33x multiplier	Poor Cohesion	[145]
Coordination (Degraded)	1.24x multiplier	Degraded Communication	[145]
Procedure Error (Yes)	0.00030	–	[87]
Procedure Error (Yes)	0.5x multiplier	Good HSI	[16]
Procedure Error (Yes)	5x multiplier	Missing/Misleading HSI	[16]
Procedure Error (Yes)	3.38x multiplier	Poor HSI	[17]
Procedure Error (Yes)	47.88x multiplier	Poor Procedure Quality	[159]
Procedure Error (Yes)	38.1x multiplier	Incomplete Procedure Quality	[159]
Procedure Error (Yes)	239x multiplier	Missing Procedure Quality	[159]
Procedure Error (Yes)	0.84x multiplier	Low Training, Knowledge, Experience, and Familiarity	[187]
Skills (Degraded)	0.00118	–	[87]
Skills (Degraded)	0.1	Degraded PPA	[87]

Node (State)	Multiplier or Probability Value	Condition	Citation
Skills (Degraded)	1.9x multiplier	Low Training, Knowledge, Experience, and Familiarity	[188]
Attention (Degraded)	0.00122	–	[87]
Attention (Degraded)	0.00126	Moderately High Perceived Urgency/Severity	[87]
Attention (Degraded)	0.00101	Very High Perceived Urgency/Severity	[87]
Attention (Degraded)	1.28x multiplier	Degraded Alertness/Fatigue	[157]
Attention (Degraded)	60.1x multiplier	Extreme Stress	[189]
Attention (Degraded)	2x multiplier	High Stress	[1]
Attention (Degraded)	0.398x multiplier	Moderately High All Loads	[87]
Attention (Degraded)	2.80x multiplier	Very High All Loads	[87]
Attention (Degraded)	1.1x multiplier	Degraded External Environment	[190]
A1 Incorrect Timing (Yes)	0.00067	–	[115]
A1 Incorrect Timing (Yes)	6.5x multiplier	Missing/Misleading HSI	[17]
A1 Incorrect Timing (Yes)	3.38x multiplier	Poor HSI	[17]
A1 Incorrect Timing (Yes)	0.5x multiplier	Good HSI	[1]
A1 Incorrect Timing (Yes)	2x multiplier	Degraded Coordination	[17]
A1 Incorrect Timing (Yes)	2x multiplier	Poor Tool Availability	[17]
A1 Incorrect Timing (Yes)	2.3x multiplier	Degraded Attention	[17]
A1 Incorrect Timing (Yes)	2x multiplier	Yes Procedure Error	[17]
A1 Incorrect Timing (Yes)	6.1x multiplier	Degraded Skills	[17]
A2 Incorrect Operation (Yes)	0.00029	–	[87]
A2 Incorrect Operation (Yes)	1.1x multiplier	Poor Tool Quality	[17]

Node (State)	Multiplier or Probability Value	Condition	Citation
A2 Incorrect Operation (Yes)	6.5x multiplier	Missing/Misleading HSI	[17]
A2 Incorrect Operation (Yes)	3.38x multiplier	Poor HSI	[17]
A2 Incorrect Operation (Yes)	0.5x multiplier	Good HSI	[1]
A2 Incorrect Operation (Yes)	1.2x multiplier	High Stress	[17]
A2 Incorrect Operation (Yes)	5x multiplier	Extreme Stress	[1]
A2 Incorrect Operation (Yes)	123.1x multiplier	Degraded Skills	[115, 116]
A2 Incorrect Operation (Yes)	15.2x multiplier	Yes Procedure Error	[17]
A2 Incorrect Operation (Yes)	2.3x multiplier	Degraded Attention	[17]
A3 Wrong Component (Yes)	0.00030	–	[87]
A3 Wrong Component (Yes)	6.5x multiplier	Missing/Misleading HSI	[17]
A3 Wrong Component (Yes)	3.38x multiplier	Poor HSI	[17]
A3 Wrong Component (Yes)	0.5x multiplier	Good HSI	[1]
A3 Wrong Component (Yes)	1.2x multiplier	High Stress	[17]
A3 Wrong Component (Yes)	5x multiplier	Extreme Stress	[1]
A3 Wrong Component (Yes)	2.3x multiplier	Degraded Attention	[17]
A4 No Action Taken (Yes)	0.00044	–	[87]
A4 No Action Taken (Yes)	0.125	Degraded Attention	[87]
A4 No Action Taken (Yes)	2x multiplier	Degraded Coordination	[17]
A4 No Action Taken (Yes)	2x multiplier	Poor Tool Availability	[17]

In addition to the parameterization of nodes through the sources detailed in Tables D.6 and

Task Complexity Node State	Pr(Very High Task Load)	Pr(Mod High Task Load)	Pr(Nominal Task Load)
Very High	0.316	0.577	0.107
Mod High	0.25	0.50	0.25
Nominal	0.158	0.628	0.214

Table D.8: Conditional probability table for Task Load in the action execution network. Very High and Moderately High Task Load probabilities are derived from SACADA [87], while the Nominal probabilities make use of a multiplier from SPAR-H [1].

D.7, it was found during validation against the GRS operational data ([115, 116]) that certain effect-modulating dependencies were being ignored. Effect-modulating dependency accounts for the nonlinear effects of multiple PIFs, causing a downstream effect larger than their linear combination would suggest [35, 85]. This concept has also been considered in the ATHEANA method's "error forcing contexts", where PIFs and plant states combine to create a worse situation than any PIF on its own.

For example, the combination of degraded procedures and an operator's lack of knowledge base (whether due to insufficient training or the individual's lack of prior experience) causes an especially difficult-to-succeed situation. Not only can the operator not reference supportive, well-written procedures, but they also cannot fall back on their situational knowledge. Therefore, when procedures are poor, incomplete, or missing, and training/knowledge/experience/familiarity is low, a 100x multiplier is applied to the likelihood of the Procedure Error mechanism occurring.

Node 1	Node 2	Child Node	Effect Multiplier
Procedure Availability/ Quality	Training, Knowledge, Experience, and Familiarity	Procedure Error	100x
Stress	All Loads	Attention	100x
HSI	Attention	A1 Incorrect Timing	10x
HSI	Attention	A2 Incorrect Action	10x
HSI	Attention	A3 Wrong Component	10x

Table D.9: Instances of effect-modulating idiom in the action execution network and the error multipliers assumed.

Bibliography

[1] David I. Gertman, Harold Blackman, Julie Marble, Curtis L. Smith, and Ronald L. Boring. *The SPAR-H Human Reliability Analysis Method*. Tech. rep. NUREG/CR-6883. Washington, D.C.: U.S. Nuclear Regulatory Commission, 2005.

[2] Jing Xing, Gareth Parry, Mary Presley, John Forester, Stacey Hendrickson, and Vinh Dang. *An Integrated Human Event Analysis System (IDHEAS) for Nuclear Power Plant Internal Events At-Power Application*. Tech. rep. NUREG/CR-2199. Washington, D.C.: U.S. Nuclear Regulatory Commission, May 2016.

[3] Barry Kirwan, W. Huw Gibson, and Brian Hickling. "Human error data collection as a precursor to the development of a human reliability assessment capability in air traffic management." In: *Reliability Engineering & System Safety* 93.2 (Feb. 2008), pp. 217–233. DOI: 10.1016/j.ress.2006.12.005.

[4] Mark A. Sujan, David Embrey, and Huayi Huang. "On the application of Human Reliability Analysis in healthcare: Opportunities and challenges." In: *Reliability Engineering & System Safety* 194 (Feb. 2020), p. 106189. DOI: 10.1016/j.ress.2018.06.017.

[5] Simon French et al. "Human Reliability Analysis: A Review and Critique." In: *Manchester Business School Working Paper* 589 (2009).

[6] Katrina M. Groth, Reuel Smith, and Ramin Moradi. "A hybrid algorithm for developing third generation HRA methods using simulator data, causal models, and cognitive science." In: *Reliability Engineering and System Safety* 191 (Nov. 2019), p. 106507. DOI: 10.1016/j.ress.2019.106507.

[7] Ali Mosleh and Y. H. James Chang. "Model-based human reliability analysis: prospects and requirements." In: *Reliability Engineering & System Safety* 83.2 (Feb. 2004), pp. 241–253. DOI: 10.1016/j.ress.2003.09.014.

[8] Alan D. Swain. "Human reliability analysis: Need, status, trends and limitations." In: *Reliability Engineering & System Safety* 29.3 (Jan. 1990), pp. 301–313. DOI: 10.1016/0951-8320(90)90013-D.

[9] Dennis Bley, Erasmia Lois, Alan Kolaczkowski, John Forester, John Wreathall, and Susan Cooper. *An overview of the evolution of human reliability analysis in the context of probabilistic risk assessment.* Tech. rep. SAND2008-5085. Albuquerque, NM: Sandia National Laboratories, Jan. 2009. DOI: 10.2172/950689.

[10] Nsimah J. Ekanem, Ali Mosleh, and Song-Hua Shen. "Phoenix – A model-based Human Reliability Analysis methodology: Qualitative Analysis Procedure." In: *Reliability Engineering & System Safety* 145 (Jan. 2016), pp. 301–315. DOI: 10.1016/j.ress.2015.07.009.

[11] Huafei Liao, John Forester, Vinh N. Dang, Andreas Bye, Y. H. James Chang, and Erasmia Lois. "Assessment of HRA method predictions against operating crew performance: Part II: Overall simulator data, HRA method predictions, and intra-method comparisons." In: *Reliability Engineering & System Safety* 191 (2019), p. 106510.

[12] Huafei Liao, John Forester, Vinh N. Dang, Andreas Bye, Y. H. James Chang, and Erasmia Lois. "Assessment of HRA method predictions against operating crew performance: Part III: Conclusions and achievements." In: *Reliability Engineering & System Safety* 191 (Nov. 2019), p. 106511. DOI: 10.1016/j.ress.2019.106511.

[13] Stacey Langfitt Hendrickson, Gareth Parry, John A. Forester, Vinh N. Dang, April M. Whaley, Stuart Lewis, Erasmia Lois, and Jing Xing. *Towards an Improved HRA Method.* Tech. rep. SAND2012-1319C. Albuquerque, NM: Sandia National Laboratories, Feb. 2012.

[14] Kilian Zwirglmaier, Daniel Straub, and Katrina M. Groth. "Capturing cognitive causal paths in human reliability analysis with Bayesian network models." In: *Reliability Engineering and System Safety* 158.October 2016 (2017), pp. 117–129. DOI: 10.1016/j.ress.2016.10.010.

[15] Alan D. Swain and H. E. Guttmann. *NUREG/CR-1278: Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications*. Tech. rep. Albuquerque, NM: Sandia National Laboratories, U.S. Nuclear Regulatory Commission, 1983.

[16] Erik Hollnagel. *Cognitive reliability and error analysis method: CREAM*. 1st ed. Oxford ; New York: Elsevier, 1998. ISBN: 978-0-08-042848-2.

[17] Jing Xing, James Chang, and Jonathan DeJesus Segarra. *Integrated Human Event Analysis System for Event and Condition Assessment (IDHEAS-ECA)*. Tech. rep. NUREG-2256. Washington, D.C.: U.S. Nuclear Regulatory Commission, Oct. 2022.

[18] Vipul Garg, Gopika Vinod, and Vivek Kant. "Auto-CREAM: Software application for evaluation of HEP with basic and extended CREAM for PSA studies." In: *Reliability Engineering & System Safety* 236 (Aug. 2023), p. 109318. DOI: 10.1016/j.ress.2023.109318.

[19] Y. H. James Chang and Ali Mosleh. "Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1: Overview of the IDAC Model." In: *Reliability Engineering & System Safety* 92.8 (Aug. 2007), pp. 997–1013. DOI: 10.1016/j.ress.2006.05.014.

[20] Rossella Bisio, Alexandra Fernandes, and Claire Blackett. "A Framework to Analyze Human Performance Outside the Control Room." In: *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference* (June 21–26, 2020). Venice, Italy.

[21] Y. H. James Chang and Ali Mosleh. "Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 4: IDAC causal model of operator problem-solving response." In: *Reliability Engineering & System Safety* 92.8 (Aug. 2007), pp. 1061–1075. DOI: 10.1016/j.ress.2006.05.011.

[22] Ahmad Al-Douri, Camille S. Levine, and Katrina M. Groth. "Identifying Human Failure Events (HFEs) for External Hazard Probabilistic Risk Assessment." In: *Reliability Engineering & System Safety* 235 (2023), p. 109236. DOI: https://doi.org/10.1016/j.ress.2023.109236.

[23] Nsimah Joseph Ekanem. "A model-based human reliability analysis methodology (Phoenix method)." PhD dissertation. College Park, MD: University of Maryland, 2013. DOI: http://hdl.handle.net/1903/14831.

[24] Katrina M. Groth and Ali Mosleh. "A data-informed PIF hierarchy for model-based Human Reliability Analysis." In: *Reliability Engineering and System Safety* 108 (2012), pp. 154–174.

[25] Camille S. Levine, Ahmad Al-Douri, Vincent Philip Paglioni, Michelle Bensi, and Katrina M. Groth. "Identifying human failure events for human reliability analysis: A review of gaps and research opportunities." In: *Reliability Engineering & System Safety* 245 (May 2024), p. 109967. ISSN: 0951-8320. DOI: 10.1016/j.ress.2024.109967.

[26] Jan-Erik Holmberg and Marja Liinasuo. *Non-PSA applications of HRA*. Tech. rep. 14124_R003. Espoo, Finland: VTT Technical Research Centre of Finland, 2017.

[27] Vincent Philip Paglioni, Torrey Mortenson, and Katrina M. Groth. "The human failure event: what is it and what should it be?" In: *Proceedings of the 16th Probabilistic Safety Assessment and Management Conference (PSAM16)* (June 26–July 1, 2022). Honolulu, HI.

[28] M. T. Barriere, J. Wreathall, S. E. Cooper, D. C. Bley, W. J. Luckas, and A. M. Ramey-Smith. *Multidisciplinary Framework for Human Reliability Analysis with an Application to Errors of Commission and Dependency*. Tech. rep. NUREG/CR-6265. Washington, D.C.: US Nuclear Regulatory Commission, Aug. 1995.

[29] S. E. Cooper, A. M. Ramey-Smith, J. Wreathall, and G. W. Parry. *A technique for human error analysis (ATHEANA)*. Tech. rep. NUREG/CR–6350, BNL-NUREG–52467. Washington, D.C.: U.S. Nuclear Regulatory Commission, May 1996. DOI: 10.2172/249298.

[30] American Society of Mechanical Engineers (ASME) and American Nuclear Society (ANS). *Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications*. New York, 2022.

[31] Vincent P. Paglioni and Katrina M. Groth. "Dependency definitions for quantitative human reliability analysis." In: *Reliability Engineering & System Safety* 220 (Apr. 2022), p. 108274. DOI: 10.1016/j.ress.2021.108274.

[32] April M. Whaley, Jing Xing, Ronald Laurids Boring, Stacey M. L. Henrickson, Jeffrey C. Joe, Katya L. Le Blanc, and Stephanie L. Morrow. *Cognitive Basis for Human Reliability Analysis (NUREG - 2114)*. Tech. rep. NUREG/CR-2114. Washington, D.C.: U.S. Nuclear Regulatory Commission, Jan. 2016.

[33] Vincent P. Paglioni, Camille S. Levine, Ahmad Al-Douri, and Katrina M. Groth. "Why Do Human-Machine Teams Fail: Investigating Failure Mechanisms in Human Reliability Analysis." In: *13th Nuclear Plant Instrumentation, Control & Human-Machine Interface Technologies (NPIC&HMIT 2023)* (July 15–20, 2023). Knoxville, TN.

[34] Y. H. James Chang and Ali Mosleh. "Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 2: IDAC performance influencing factors model." In: *Reliability Engineering & System Safety* 92.8 (Aug. 2007), pp. 1014–1040. DOI: 10.1016/j.ress.2006.05.010.

[35] Vincent Philip Paglioni and Katrina M Groth. "Dependency idioms for quantitative human reliability analysis." In: *Nuclear Science and Engineering* 198.8 (Sept. 2023), pp. 1645–1667. DOI: 10.1080/00295639.2023.2250159.

[36] Michelle Bensi, Katrina M. Groth, Zeyun Wu, Ray Schneider, Somayeh\ Mohammadi, Tao Liu, Ahmad Al-Douri, and Camille Levine. "Identifying and Prioritizing Sources of Uncertainty in External Hazard Probabilistic Risk Assessment: Project Activities and Progress." In: *Probabilistic Safety Assessment and Management (PSAM) 16* (June 26–July 1, 2022). Honolulu, HI.

[37] Michelle Bensi, Zeyun Wu, Katrina Groth, Zhegang Ma, Ray Schneider, Tao Liu, Kaveh Faraji Najarkolaie, Ahmad Al-Douri, and Camille Levine. "Identification and Prioritization of Sources of Uncertainty in External Hazard Probabilistic Risk Assessment: Project Update." In: *18th International Probabilistic Safety Assessment and Analysis (PSA 2023)* (July 15–20, 2023). Knoxville, TN.

[38] J. Williams. "HEART–A proposed method for assessing and reducing human error." In: *Ninth Advances in Reliability Technology Symposium* (Apr. 2–4, 1986). Bradford, UK.

[39] Julie Marble, Huafei Liao, Mary Presley, John Forester, Andreas Bye, Vinh Dang, and Erasmia Lois. "Results and Insights Derived from the Intra-Method Comparisons of the US HRA Empirical Study." In: *Proceedngs of the 11th Probabilistic Safety Assessment and Management/European Safety and Reliability 2012* (2012).

[40] Claire Blackett, Sondre Øie, and Kristian Gould. "Lessons Learned from Applying a New HRA Method for the Petroleum Industry." In: *Reliability Engineering & System Safety* (Oct. 2018). DOI: 10.1016/j.ress.2018.10.001.

[41] Jaewhan Kim, Wondea Jung, and Jinkyun Park. "AGAPE-ET: Advanced guidelines for human reliability analysis of emergency tasks." In: *2007 IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting* (Aug. 26, 2007–Aug. 31, 2008), pp. 314–321. DOI: 10.1109/HFPP.2007.4413226.

[42] Ronald L. Boring, Martin Rasmussen, Thomas A. Ulrich, and Nancy J. Lybeck. "Aggregation of Autocalculated Human Error Probabilities from Tasks to Human Failure Events in a Dynamic Human Reliability Analysis Implementation." In: *Probabilistic Safety Assessment and Management (PSAM) 14* (Sept. 16–21, 2018). Los Angeles, CA.

[43] Yochan Kim, Sun Yeong Choi, Jinkyun Park, and Jaewhan Kim. "Empirical study on human error probability of procedure-extraneous behaviors." In: *Reliability Engineering & System Safety* 227 (Nov. 2022), p. 108727. DOI: 10.1016/j.ress.2022.108727.

[44] Camille S. Levine, Ahmad Al-Douri, and Katrina M. Groth. "Causal Pathways Leading to Human Failure Events in Information-Gathering System Response Activities." In: *13th Nuclear Plant Instrumentation, Control & Human-Machine Interface Technologies (NPIC&HMIT 2023)* (July 15–20, 2023). Knoxville, TN.

[45] Katrina M. Groth and Ali Mosleh. "Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model." In: *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 226.4 (2012), pp. 361–379. DOI: 10.1177/1748006X11428107.

[46] Burak Polat. "A Model-Based Human Reliability Analysis Methodology for the Physical Protection of Nuclear Facilities." PhD thesis. Raleigh, NC: North Carolina State University, 2022.

[47] Y. H. James Chang and Ali Mosleh. "Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 5: Dynamic probabilistic simulation of the IDAC model." In: *Reliability Engineering & System Safety* 92.8 (Aug. 2007), pp. 1076–1101. DOI: 10.1016/j.ress.2006.05.012.

[48] Yunfei Zhao and Carol Smidts. "CMS-BN: A cognitive modeling and simulation environment for human performance assessment, part 1 — methodology." In: *Reliability Engineering & System Safety* 213 (Sept. 2021), p. 107776. DOI: 10.1016/j.ress.2021.107776.

[49] Ali Mosleh. "PRA: a perspective on strengths, current limitations, and possible improvements." In: *Nuclear Engineering and technology* 46.1 (2014), pp. 1–10.

[50] Taotao Zhou, Mohammad Modarres, and Enrique López Droguett. "Multi-unit nuclear power plant probabilistic risk assessment: A comprehensive survey." In: *Reliability Engineering & System Safety* 213 (Sept. 2021), p. 107782. DOI: 10.1016/j.ress.2021.107782.

[51] Yolande Akl and Karl Fleming. *Summary report of the international workshop on multi-unit probabilistic safety assessment*. Tech. rep. Ottawa, Canada: Canadian Nuclear Safety Commission, 2015.

[52] Shawn St Germain, Ron Boring, Georgeta Banaseanu, Yolande Akl, and Hayat Chatri. "Multi-unit considerations for human reliability analysis." In: *PSAM Topical Conference on Human Reliability, Quantitative Human Factors, and Risk Management* (June 7–9, 2017). Munich, Germany.

[53] Jooyoung Park, Awwal Mohammed Arigi, and Jonghyun Kim. "Treatment of human and organizational factors for multi-unit HRA: Application of SPAR-H method." In: *Annals of Nuclear Energy* 132 (2019), pp. 656–678. DOI: 10.1016/j.anucene.2019.06.053.

[54] Awwal Mohammed Arigi, Gayoung Park, and Jonghyun Kim. "Dependency analysis method for human failure events in multi-unit probabilistic safety assessments." In: *Reliability Engineering and System Safety* 203 (2020), p. 107112. DOI: 10.1016/j.ress.2020.107112.

[55] Bernhard Reer, Vinh N. Dang, and Stefan Hirschberg. "The CESA method and its application in a plant-specific pilot study on errors of commission." In: *Reliability Engineering & System Safety* 83.2 (Feb. 2004), pp. 187–205. DOI: 10.1016/j.ress.2003.09.010.

[56] Awwal Mohammed Arigi, Jooyoung Park, Hak-Kyu Lim, and Jonghyun Kim. "Analysis of human-induced initiating events in a multi-unit loss of offsite power scenario." In: *Journal of Nuclear Science and Technology* 57.1 (Jan. 2020), pp. 121–132. DOI: 10.1080/00223131.2019.1661305.

[57] Xuhong He, Cilla Andersson, Anders Olsson, Julia Ljungbjörk, Anders Karlsson, Lovisa Nordlof, Karl Gustafsson, Lasse Tunturivuori, and Per Hellström. "Errors of Commission in HRA – NPSAG Phase 1 project." In: *Probabilistic Safety Assessment and Management (PSAM) 14* (Sept. 16–21, 2018). Los Angeles.

[58] Yongchan Kim and Jonghyun Kim. "Identification of human-induced initiating events in the low power and shutdown operation using the Commission Error Search and Assessment method." In: *Nuclear Engineering and Technology* 47.2 (Mar. 2015), pp. 187–195. DOI: 10.1016/j.net.2014.12.006.

[59] J. Julius, E. Jorgenson, G. W. Parry, and A. M. Mosleh. "A procedure for the analysis of errors of commission in a Probabilistic Safety Assessment of a nuclear power plant at full power." In: *Reliability Engineering & System Safety* 50.2 (Jan. 1995), pp. 189–201. DOI: 10.1016/0951-8320(95)00075-D.

[60] Rajiv Prasad, Garill Coles, A. Dalton, N. Kohn, K. Branch, A. Bittner, and R. Taylor. *Effects of Environmental Conditions on Manual Actions for Flood Protection and Mitigation*. Tech. rep. NUREG/CR-7256. Richland, WA: Pacific Northwest National Laboratory, U.S. Nuclear Regulatory Commission, May 2020.

[61] Ronald Laurids Boring. "Top-Down and Bottom-Up Definitions of Human Failure Events in Human Reliability Analysis." In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58.1 (Sept. 2014), pp. 563–567.

[62] Dhruv Pandya, Luca Podofillini, Frank Emert, Antony J. Lomax, and Vinh N. Dang. "Developing the foundations of a cognition-based human reliability analysis model via mapping task types and performance-influencing factors: Application to radiotherapy." In: *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 232.1 (Feb. 2018), pp. 3–37. DOI: 10.1177/1748006X17731903.

[63] D. E. Embrey. "SHERPA: A systematic human error reduction and prediction approach." In: *Proceedings of the international topical meeting on advances in human factors in nuclear power systems* (Apr. 21–24, 1986). Knoxville, TN, pp. 184–193.

[64] S. T. Shorrock and B. Kirwan. "TRACEr: a technique for the retrospective of cognitive errors in ATM." In: *Engineering Psychology and Cognitive Ergonomics: Volume Three-Transportation Systems, Medical Ergonomics and Training* (1999).

[65] Claire Blackett, Jan Erik Farbrot, Sondre Øie, and Marius Fernander. *The Petro-HRA Guideline, Rev. 1, Vol. 2*. Tech. rep. IFE/E-2022/002. Kjeller, Norway: Institute for Energy Technology, Jan. 2022.

[66] Judea Pearl. *Causality: models, reasoning, and inference*. 2nd. Cambridge, U.K. ; New York: Cambridge University Press, 2009. ISBN: 978-0-521-89560-6.

[67] Katrina M. Groth and Laura P. Swiler. "Bridging the gap between HRA research and HRA practice: A Bayesian network version of SPAR-H." In: *Reliability Engineering & System Safety* 115 (July 2013), pp. 33–42. DOI: 10.1016/j.ress.2013.02.015.

[68] Lusine Mkrtchyan, Luca Podofillini, and Vinh N. Dang. "Bayesian belief networks for human reliability analysis: A review of applications and gaps." In: *Reliability Engineering & System Safety* 139 (July 2015), pp. 1–16. DOI: 10.1016/j.ress.2015.02.006.

[69] Martin Neil, Norman Fenton, and Lars Nielson. "Building large-scale Bayesian networks." In: *The Knowledge Engineering Review* 15.3 (Sept. 2000), pp. 257–284. DOI: 10.1017/S0269888900003039.

[70] Norman Fenton, Martin Neil, and David A. Lagnado. "A General Structure for Legal Arguments About Evidence Using Bayesian Networks." In: *Cognitive Science* 37.1 (2013), pp. 61–102. DOI: 10.1111/cogs.12004.

[71] Camille S. Levine and Katrina M. Groth. "Three Novel Causal Logic-Informed Models for Identifying and Quantifying Operator Errors, Part I: Information-Gathering." In: *Reliability Engineering & System Safety* (2025).

[72] Katrina M. Groth and Stacey Langfitt Hendrickson. *Draft Bayesian Networks for IDHEAS*. 2012.

[73] Mica R. Endsley. "Toward a Theory of Situation Awareness in Dynamic Systems." In: *Human Factors* 37.1 (Mar. 1995). Publisher: SAGE Publications Inc, pp. 32–64. DOI: 10.1518/001872095779049543.

[74] Michael P. Letsky, Norman W. Warner, Stephen Fiore, and C.A.P. Smith. *Macrocognition in teams: Theories and methodologies*. CRC Press, Jan. 2008.

[75] Donald E. Broadbent. *Perception and Communication*. Elsevier, 1958. ISBN: 978-1-4832-2582-1.

[76] Charles W. Eriksen and James D. St. James. "Visual attention within and around the field of focal attention: A zoom lens model." In: *Perception & Psychophysics* 40.4 (July 1986), pp. 225–240. DOI: 10.3758/BF03211502.

[77] Debra Jones and Mica Endsley. "Sources of situation awareness errors in aviation." In: *Aviation, space, and environmental medicine* 67 (July 1996), pp. 507–12.

[78] Gary A. Klein. "A recognition-primed decision (RPD) model of rapid decision making." In: *Decision making in action: models and methods*. Norwood, NJ: Ablex Pub, 1993, pp. 138–147. ISBN: 978-0-89391-794-4.

[79] Gary Klein, Brian Moon, and Robert Hoffman. "Making Sense of Sensemaking 2: A Macrocognitive Model." In: *Intelligent Systems, IEEE* 21 (Oct. 2006), pp. 88–92. DOI: 10.1109/MIS.2006.100.

[80] Raanan Lipshitz. "Converging themes in the study of decision making in realistic settings." In: *Decision making in action: Models and methods*. Westport, CT, US: Ablex Publishing, 1993, pp. 103–137. ISBN: 978-0-89391-794-4.

[81] Judith Orasanu and Lynne Martin. "Errors in Aviation Decision Making: A Factor in Accidents and Incidents." In: *Proceedings of the Workshop on Human error, Safety, and Systems Development* (May 29–31, 1998). Warrenton, VA, pp. 100–107.

[82] Emilie M. Roth. "Analysis of decision making in nuclear power plant emergencies: An investigation of aided decision making." In: *Naturalistic Decision Making*. Ed. by Caroline E. Zsambok and Gary Klein. Psychology Press, 1997, pp. 175–182. ISBN: 978-1-317-77960-5.

[83] Christopher D. Wickens, John Lee, Sallie E. Gordon-Becker, and Yili Liu. *An Introduction to Human Factors Engineering*. Pearson Prentice Hall, 2004. ISBN: 978-0-13-183736-2.

[84] Erasmia Lois, Ali Mosleh, Sankaran Mahadevan, and Dennis Bley. *The Employment of Empirical Data and Bayesian Methods in Human Reliability Analysis: A Feasibility Study*. Tech. rep. NUREG/CR-6949, INL/EXT-06-11670. Washington, D.C.: U.S. Nuclear Regulatory Commission, 2007.

[85] Vincent P. Paglioni. "Improving the Foundational Knowledge of Dependency in Human Reliability Analysis." PhD thesis. College Park, MD: University of Maryland, Apr. 2023.

[86] H. Gibson, G. Basra, and B. Kirwan. "Development of the CORE-DATA database." In: *Safety & Reliability Journal, Safety and Reliability Society, Manchester* 19.1 (1999), pp. 6–20.

[87] Y. H. James Chang et al. "The SACADA database for human reliability and human performance." In: *Reliability Engineering & System Safety*. Special issue of selected articles from ESREL 2012 125 (May 2014), pp. 117–133. DOI: 10.1016/j.ress.2013.07.014.

[88] A. J. Spurgin, P. Moieni, C. D. Gaddy, G. Parry, D. D. Orvis, J. P. Spurgin, V. Joksimovitch, D. P. Gaver, and G. W. Hannaman. "Operator reliability experiments using power plant simulators." In: *Electric Power Research Institute, EPRI NP-6937* 1 (1990).

[89] B. Hallbert, Ronald Boring, D. Gertman, D. Dudenhoffer, April Whaley, J. Marble, Jeffrey Joe, and Erasmia Lois. *Human Event Repository and Analysis (HERA) System, Overview*. Tech. rep. NUREG/CR-6903. Idaho Falls, ID: Idaho National Laboratory,U.S. Nuclear Regulatory Commission, 2006.

[90] Jing Xing, Y. H. James Chang, and Jonathan DeJesus Segarra. *DRAFT- Integrated Human Event Analysis System for Human Reliability Data (IDHEAS-DATA)*. Tech. rep. RIL-2021-XX. Washington, D.C.: U.S. Nuclear Regulatory Commission, 2021.

[91] Vincent P. Paglioni and Katrina M. Groth. "Creating formative HRA dependency models using the HRA dependency idioms and SACADA data, Part II: Model quantification." In: *Annals of Nuclear Energy* 208 (Dec. 2024), p. 110761. ISSN: 0306-4549. DOI: 10.1016/j.anucene.2024.110761.

[92] BayesFusion. *SMILE: Structural Modeling, Inference, and Learning Engine*. Apr. 2024.

[93] Becky L. Hooey, Christopher D. Wickens, Ellen Salud, Angelia Sebok, Shaun Hutchins, and Brian F. Gore. "Predicting the unpredictable: estimating human performance parameters for off-nominal events." en. In: *Proceedings of the 15th International Symposium on Aviation Psychology*. Dayton, OH, Apr. 2009.

[94] Xiaojun Wu, Manrong She, Zhizhong Li, Fei Song, and Wei Sang. "Effects of integrated designs of alarm and process information on diagnosis performance in digital nuclear power plants." In: *Ergonomics* 60.12 (Dec. 2017), pp. 1653–1666. ISSN: 0014-0139. DOI: 10.1080/00140139.2017.1335884.

[95] Anina N. Rich, Melina A. Kunar, Michael J. Van Wert, Barbara Hidalgo-Sotelo, Todd S. Horowitz, and Jeremy M. Wolfe. "Why do we miss rare targets? Exploring the boundaries of the low prevalence effect." In: *Journal of Vision* 8.15 (Nov. 2008), p. 15. ISSN: 1534-7362. DOI: 10.1167/8.15.15.

[96] *OPIS: Operational Performance Information System for Nuclear Power Plant*.

[97] Michael Barriere et al. *Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA)*. Tech. rep. NUREG-1624. Washington, D.C.: U.S. Nuclear Regulatory Commission, May 2000.

[98] *1979 NRC Annual Report*. Tech. rep. NUREG-0690. Washington, D.C.: U.S. Nuclear Regulatory Commission, Sept. 1979.

[99] Alexandra Fernandes, Rossella Bisio, and Claire Blackett. "Operator Actions Outside the Control Room: A Field Study." In: *Engineering Psychology and Cognitive Ergonomics, Cognition and Design*. Springer, July 2020, pp. 30–41.

[100] Camille S. Levine, Ahmad Al-Douri, and Katrina M. Groth. "Causal Pathways Leading to Decision and Action Human Failure Events: Structures and Validation." In: *Proceedings of the 17th Probabilistic Safety Assessment and Management Conference (PSAM17) & Asian Symposium on Risk Assessment and Management (ASRAM2024)* (Oct. 7–11, 2024). Sendai, Japan.

[101] Camille S. Levine and Katrina M. Groth. "Three Novel Causal Logic-Informed Models for Identifying and Quantifying Operator Errors, Part II: Decision-Making." In: *Reliability Engineering & System Safety* (2025).

[102] Salvatore Massaiu and Lars Holmgren. *Diagnosis and Decision-Making with Emergency Operating Procedures in Non-Typical Conditions: A HAMMLAB Study with U.S. Operators*. Tech. rep. HWR-1121. Halden, Norway: IFE, Aug. 2014.

[103] James Reason. "Safety paradoxes and safety culture." In: *Injury Control and Safety Promotion* 7.1 (Mar. 2000), pp. 3–14. DOI: 10.1076/1566-0974(200003)7:1;1-V;FT003.

[104] Mica R. Endsley and Debra Jones. *Designing for Situation Awareness: an Approach to User-Centered Design*. 2nd. CRC Press, 2016. ISBN: 978-1-4200-6358-5.

[105] Andreas Bye. "Future needs of human reliability analysis: The interaction between new technology, crew roles and performance." In: *Safety Science* 158 (Feb. 2023), p. 105962. DOI: 10.1016/j.ssci.2022.105962.

[106] Raanan Lipshitz and Orit ben Shaul. "Schemata and Mental Models in Recognition-Primed Decision Making." In: *Naturalistic Decision Making*. Ed. by Caroline E. Zsambok and Gary Klein. Lawrence Elbaum Associates, 1997, pp. 293–303.

[107] Espen Nystad, Magnhild Kaarstad, and Robert McDonald. "Crew Decision-making in Situations with Degraded Information." In: *Proceedings of the 29th European Safety and Reliability Conference (ESREL)* (Sept. 22–26, 2019). Hannover, Germany. DOI: 10.3850/978-981-11-2724-3_0523-cd.

[108] Amos Tversky and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases." In: *Science, New Series* 185.4157 (1974), pp. 1124–1131.

[109] Bruce Siddle. *The Stress Paradox*. Oct. 2008.

[110] Per Braarud and Bertil Johansson. *Team Cognition in a Complex Accident Scenario*. Tech. rep. HWR-955. Halden, Norway: IFE, Mar. 2010.

[111] Michael Hildebrandt and Robert McDonald. *Procedure Use, Communication, and Teamwork in a Digital Control Room*. Tech. rep. HWR-1313. Halden, Norway: IFE, Dec. 2020.

[112] Emil Andersen, Igor Kozine, and Anja Maier. "Biased Decision-Making in Realistic Extra-Procedural Nuclear Control Room Scenarios." In: *Design Computing and Cognition '18*. Ed. by John S. Gero. Springer International Publishing, 2019, pp. 555–573. ISBN: 978-3-030-05362-8. DOI: 10.1007/978-3-030-05363-5_30.

[113] Frank Greitzer, Robin Podmore, Marck Robinson, and Pamela Ey. "Naturalistic Decision Making for Power System Operators." In: *International Journal of Human–Computer Interaction* 26.2-3 (Mar. 2010), pp. 278–291. DOI: 10.1080/10447310903499070.

[114] Christopher D. Wickens, Justin G. Hollands, Simon Banbury, and R. Parasuraman. *Engineering psychology and human performance*. 4th ed. Boston, MA: Pearson, 2012. ISBN: 978-0-205-02198-7.

[115] Wolfgang Preischl and Mario Hellmich. "Human error probabilities from operational experience of German nuclear power plants." In: *Reliability Engineering & System Safety* 109 (Jan. 2013), pp. 150–159. DOI: 10.1016/j.ress.2012.08.004.

[116] Wolfgang Preischl and Mario Hellmich. "Human error probabilities from operational experience of German nuclear power plants, Part II." In: *Reliability Engineering & System Safety* 148 (Apr. 2016), pp. 44–56. ISSN: 0951-8320. DOI: 10.1016/j.ress.2015.11.011.

[117] Camille S. Levine and Katrina M. Groth. "Three Novel Causal Logic-Informed Models for Identifying and Quantifying Operator Errors, Part III: Action Execution." In: *Reliability Engineering & System Safety* (2025).

[118] Gareth W. Parry, B. O. Y. Lydell, A. J. Spurgin, P. Moieni, and A. Beare. *An approach to the analysis of operator actions in probabilistic risk assessment*. Tech. rep. TR-100259. Palo Alto: Electronic Power Research Institute, 1992.

[119] Jeffy Jahfar Poozhithara, Deanna M. Kennedy, Spencer Onstot, Agnė Januškevičiūtė, and Marjanthi Cekrezi. "Predictive Algorithm for Team Mental Model Convergence." In: *IEEE Transactions on Computational Social Systems* 10.2 (Apr. 2023), pp. 640–655. DOI: 10.1109/TCSS.2022.3169726.

[120] Beverly Huey and Christopher D. Wickens. *Workload Transition: Implications for Individual and Team Performance*. Tech. rep. Washington, D.C.: National Research Council, Jan. 1993.

[121] Robert W. Proctor and Trisha Van Zandt. *Human Factors in Simple and Complex Systems*. 3rd ed. Boca Raton: CRC Press, Oct. 2017. ISBN: 978-1-315-15681-1. DOI: 10.1201/9781315156811.

[122] Shulan Hsieh, Cheng-Yin Tsai, and Ling-Ling Tsai. "Error correction maintains post-error adjustments after one night of total sleep deprivation." In: *Journal of Sleep Research* 18.2 (2009), pp. 159–166. DOI: 10.1111/j.1365-2869.2008.00730.x.

[123] Erik M. Altmann and J. Gregory Trafton. "Memory for goals: an activation-based model." In: *Cognitive Science* 26.1 (2002), pp. 39–83. DOI: 10.1207/s15516709cog2601_2.

[124] Alessandro Valenza, Harry Charlier, Antonino Bianco, and Davide Filingeri. "Independent and interactive effects of thermal stress and mental fatigue on manual dexterity." In: *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology* 319.6 (Dec. 2020), R703–R711. ISSN: 1522-1490. DOI: 10.1152/ajpregu.00226.2020.

[125] John Forester et al. *The U.S. HRA Empirical Study: Assessment of HRA Method Predictions against Operating Crew Performance on a U.S. Nuclear Power Plant Simulator*. Tech. rep. NUREG-2156. Washington, D.C.: U.S. Nuclear Regulatory Commission, June 2016.

[126] A. M. Kolaczkowski, J. A. Forester, Erasmia Lois, and S. E. Cooper. *Good Practices for Implementing Human Reliability Analysis (HRA)*. Tech. rep. NUREG-1792. Washington, D.C.: U.S. Nuclear Regulatory Commission, Apr. 2005, p. 110.

[127] International Atomic Energy Agency. *Seismic Hazards in Site Evaluation for Nuclear Installations*. Tech. rep. Specific Safety Guide No. SSG-9. Vienna, Austria: IAEA, 2010.

[128] International Atomic Energy Agency. *Meteorological and Hydrological Hazards in Site Evaluation for Nuclear Installations*. Tech. rep. DS417. Vienna, Austria: IAEA, 2011.

[129] Michelle Bensi, Somayeh Mohammadi, Nilanjana Ghosh, Tao Liu, Camille S. Levine, Ahmad Al-Douri, Ray Schneider, Zeyun Wu, Katrina Groth, and Zhegang Ma. "Uncertainty in External Hazard Probabilistic Risk Assessment (PRA): A Structured Taxonomy." In: *Probabilistic Safety Assessment and Management (PSAM) 16* (June 26–July 1, 2022). Honolulu, HI.

[130] Taotao Zhou, Mohammad Modarres, and Enrique López Droguett. "A Review of Multi-Unit Nuclear Power Plant Probabilistic Risk Assessment Research." In: *Volume 2: Plant Systems, Structures, Components, and Materials; Risk Assessments and Management* (July 22–26, 2018). London, England: American Society of Mechanical Engineers. ISBN: 978-0-7918-5144-9. DOI: 10.1115/ICONE26-81130.

[131] Charles Miller, Amy Cubbage, Daniel Dorman, Jack Grobe, Gary Holahan, and Nathan Sanfilippo. *Recommendations for Enhancing Reactor Safety in the 21st Century*. Tech. rep. SECY-11-0093. Washington, D.C.: U.S. Nuclear Regulatory Commission, July 2011.

[132] Hayat Chatri, Smain Yalaoui, and Yolande Akl. "Human Reliability Analysis (HRA) methodologies used in the Canadian Probabilistic Safety Assessment (PSA) for external events." In: *Probabilistic Safety Assessment and Management [PSAM] 14* (Sept. 16–21, 2018). Los Angeles, CA.

[133] Ali Ayoub, Andrej Stankovski, Wolfgang Kröger, and Didier Sornette. "Precursors and startling lessons: Statistical analysis of 1250 events with safety significance from the civil nuclear sector." In: *Reliability Engineering & System Safety* 214.107820 (2021).

[134] Yolande Akl and Karl Fleming. *Summary report of the international workshop on multi-unit probabilistic safety assessment*. Tech. rep. Ottawa, Canada: Canadian Nuclear Safety Commission, 2015.

[135] Mohammad Modarres, Taotao Zhou, and Mahmoud Massoud. "Advances in multi-unit nuclear power plant probabilistic risk assessment." In: *Reliability Engineering & System Safety* 157 (2017), pp. 87–100.

[136] Diego Mandelli, Carlo Paris, Andrea Alfonsi, Dan Maljovec, Ronald L. Boring, S. Ewing, Shawn St. Germain, Curtis L. Smith, Cristian Rabiti, and Martin Rasmussen. "Multi-unit dynamic PRA." In: *Reliability Engineering & System Safety* 185 (2019), pp. 303–317.

[137] Antonios Zoulis and Y. H. James Chang. "Assessing human failures events for ex-control room actions." In: *ANS PSA 2013 International Topical Meeting on Probabilistic Safety Assessment and Analysis* (Sept. 22–27, 2013). Columbia, SC: American Nuclear Society.

[138] Marilia A. Ramos, Ingrid Bouwer Utne, and Ali Mosleh. "Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events." In: *Safety Science* 116 (July 2019), pp. 33–44. DOI: 10.1016/j.ssci.2019.02.038.

[139] Marilia A. Ramos, Christoph A. Thieme, Ingrid B. Utne, and Ali Mosleh. "Human-system concurrent task analysis for maritime autonomous surface ship operation and safety." In: *Reliability Engineering & System Safety* 195 (Mar. 2020), p. 106697. DOI: 10.1016/j.ress.2019.106697.

[140] Marilia Ramos, Enrique López Droguett, Ali Mosleh, and Márcio das Chagas Moura. "A human reliability analysis methodology for oil refineries and petrochemical plants operation: Phoenix-PRO qualitative framework." In: *Reliability Engineering & System Safety* 193 (Jan. 2020), p. 106672. DOI: 10.1016/j.ress.2019.106672.

[141] Shuai Chen, Li Zhang, and Tao Qing. "A human reliability analysis methodology based on an extended Phoenix method for severe accidents in nuclear power plants: Qualitative analysis framework." In: *Reliability Engineering & System Safety* 214 (Oct. 2021), p. 107750. DOI: 10.1016/j.ress.2021.107750.

[142] Ahmad Al-Douri, Camille S. Levine, and Katrina M. Groth. "Identifying Human Failure Events (HFEs) for External Hazard Probabilistic Risk Assessment." In: *PSAM 16* (June 26–July 1, 2022). Honolulu, HI.

[143] Camille S. Levine and Katrina M. Groth. "Expanding the Causal Logic Foundations of Human Reliability Analysis: Current Project Status." In: *ANS 2024 Winter Meeting* (Nov. 17–21, 2024). Orlando, FL.

[144] Sarah Blissett, Deric Morrison, David McCarty, and Matthew Sibbald. "Should learners reason one step at a time? A randomised trial of two diagnostic scheme designs." In: *Medical Education* 51.4 (2017), pp. 432–441. ISSN: 1365-2923. DOI: 10.1111/medu.13221.

[145] Jinkyun Park, Wondea Jung, and Joon-Eon Yang. "Investigating the effect of communication characteristics on crew performance under the simulated emergency condition of nuclear power plants." In: *Reliability Engineering & System Safety* 101 (May 2012), pp. 1–13. ISSN: 0951-8320. DOI: 10.1016/j.ress.2012.01.003.

[146] B. Hallbert, Angelia Sebok, and D. Morisseau. *A Study of Control Room Staffing Levels for Advanced Reactors*. Tech. rep. NUREG/IA-0137. Halden, Norway: OECD Halden Reactor Project, U.S. Nuclear Regulatory Commission, Nov. 2000.

[147] Hui-Ling Tung and Yu-Hsuan Chang. "Effects of empowering leadership on performance in management team." In: *Journal of Chinese Human Resources Management* 2.1 (Jan. 2011), pp. 43–60. ISSN: 2040-8005. DOI: 10.1108/20408001111148720.

[148] Kelly L. Zellars, Bennett J. Tepper, and Michelle K. Duffy. "Abusive supervision and subordinates' organizational citizenship behavior." In: *Journal of Applied Psychology* 87.6 (Dec. 2002), pp. 1068–1076. ISSN: 00219010. DOI: 10.1037/0021-9010.87.6.1068.

[149] Alexandra M. Freund and Martin J. Tomasik. "Managing conflicting goals through prioritization? The role of age and relative goal importance." en. In: *PLOS ONE* 16.2 (Feb. 2021). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0247047.

[150] Nikos Gorgoraptis, Raquel F. G. Catalao, Paul M. Bays, and Masud Husain. "Dynamic Updating of Working Memory Resources for Visual Objects." In: *Journal of Neuroscience* 31.23 (June 2011), pp. 8502–8511. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0208-11.2011.

[151] Christopher B. Frazier, Timothy D. Ludwig, Brian Whitaker, and D. Steve Roberts. "A hierarchical factor analysis of a safety culture survey." In: *Journal of Safety Research* 45 (June 2013), pp. 15–28. ISSN: 0022-4375. DOI: 10.1016/j.jsr.2012.10.015.

[152] Michelle Richardson and Charles Abraham. "Conscientiousness and achievement motivation predict performance." In: *European Journal of Personality* 23.7 (Nov. 2009), pp. 589–605. ISSN: 0890-2070. DOI: 10.1002/per.732.

[153] Aaron M. Schmidt and Richard P. DeShon. "What to do? The effects of discrepancies, incentives, and time on dynamic goal prioritization." In: *Journal of Applied Psychology* 92.4 (July 2007), pp. 928–941. ISSN: 1939-1854, 0021-9010. DOI: 10.1037/0021-9010.92.4.928.

[154] Hamood M. Alenezi and Markus Bindemann. "The Effect of Feedback on Face-Matching Accuracy." In: *Applied Cognitive Psychology* 27.6 (2013), pp. 735–753. ISSN: 1099-0720. DOI: 10.1002/acp.2968.

[155] J. Kerstholt. *Time pressure effects on decision making in a dynamic task environment*. Tech. rep. IZF 1991 B-8. Soesterberg, The Netherlands: Institute For Perception RVO-TNO, 1991.

[156] Raja Parasuraman, Robert Molloy, and Indramani L. Singh. "Performance Consequences of Automation-Induced 'Complacency'." In: *The International Journal of Aviation Psychology* 3.1 (Jan. 1993), pp. 1–23. DOI: 10.1207/s15327108ijap0301_1.

[157] Christopher D. Wickens, Shaun D. Hutchins, Lila Laux, and Angelia Sebok. "The Impact of Sleep Disruption on Complex Cognitive Tasks: A Meta-Analysis." In: *Human Factors* 57.6 (Sept. 2015), pp. 930–946. ISSN: 0018-7208. DOI: 10.1177/0018720815571935.

[158] A. N. Beare and R. E. Dorris. "A Simulator-Based Study of Human Errors in Nuclear Power Plant Control Room Tasks." In: *Proceedings of the Human Factors Society Annual Meeting* 27.2 (Oct. 1983). Publisher: SAGE Publications, pp. 170–174. ISSN: 0163-5182. DOI: 10.1177/154193128302700213.

[159] Ar Ryum Kim, Jinkyun Park, Ji Tae Kim, Jaewhan Kim, and Poong Hyun Seong. "Study on the identification of main drivers affecting the performance of human operators during low power and shutdown operation." In: *Annals of Nuclear Energy* 92 (June 2016), pp. 447–455. ISSN: 0306-4549. DOI: 10.1016/j.anucene.2016.02.010.

[160] Andrew R. Dattel, Lisa Kossuth, Chelsea C. Sheehan, H. Justin Green, Crystal Giannini, Jennifer Decker, Hayley M. Mericle-Swingle, and Suzanne A. Crockett. "Effects of Conceptual Training and Procedural Training for Teaching Aviation Instrument Holding Patterns." In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57.1 (Sept. 2013), pp. 1445–1449. DOI: 10.1177/1541931213571323.

[161] Fei-Hui Huang and Sheue-Ling Hwang. "Experimental studies of computerized procedures and team size in nuclear power plant operations." In: *Nuclear Engineering and Design* 239.2 (Feb. 2009), pp. 373–380. ISSN: 0029-5493. DOI: 10.1016/j.nucengdes.2008.10.009.

[162] Brandon Phillips, Victor R. Prybutok, and Daniel A. Peak. "Decision Confidence, Information Usefulness, and Information Seeking Intention in the Presence of Disconfirming Information." In: *Informing Science: The International Journal of an Emerging Transdiscipline* 17 (2014), pp. 001–024. ISSN: 1547-9684, 1521-4672. DOI: 10.28945/1932.

[163] Shameem Hameed, Thomas Ferris, Swapnaa Jayaraman, and Nadine Sarter. "Using Informative Peripheral Visual and Tactile Cues to Support Task and Interruption Management." In: *Human Factors* 51.2 (Apr. 2009), pp. 126–135. ISSN: 0018-7208. DOI: 10.1177/0018720809336434.

[164] Alla Keselman and Catherine Arnott Smith. "A classification of errors in lay comprehension of medical documents." In: *Journal of Biomedical Informatics* 45.6 (Dec. 2012), pp. 1151–1163. ISSN: 1532-0464. DOI: 10.1016/j.jbi.2012.07.012.

[165] Marc H. Anderson. "Social networks and the cognitive motivation to realize network opportunities: a study of managers' information gathering behaviors." In: *Journal of Organizational Behavior* 29.1 (2008), pp. 51–78. ISSN: 1099-1379. DOI: 10.1002/job.459.

[166] Serkan Cak, Bilge Say, and Mine Misirlisoy. "Effects of working memory, attention, and expertise on pilots' situation awareness." In: *Cognition, Technology & Work* 22.1 (Feb. 2020), pp. 85–94. ISSN: 1435-5566. DOI: 10.1007/s10111-019-00551-w.

[167] Clara Casco, Patrizio E. Tressoldi, and Annamaria Dellantonio. "Visual Selective Attention and Reading Efficiency are Related in Children." In: *Cortex* 34.4 (Jan. 1998), pp. 531–546. ISSN: 0010-9452. DOI: 10.1016/S0010-9452(08)70512-4.

[168] L. Lingard, S. Espin, S. Whyte, G. Regehr, G. R. Baker, R. Reznick, J. Bohnen, B. Orser, D. Doran, and E. Grober. "Communication failures in the operating room: an observational classification of recurrent types and effects." In: *BMJ Quality & Safety* 13.5 (Oct. 2004), pp. 330–334. ISSN: 2044-5415, 2044-5423. DOI: 10.1136/qshc.2003.008425.

[169] Nazbanou Nozari and Gary S. Dell. "Feature migration in time: Reflection of selective attention on speech errors." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38.4 (2012), pp. 1084–1090. ISSN: 1939-1285. DOI: 10.1037/a0026933.

[170] David L. Strayer and William A. Johnston. "Driven to Distraction: Dual-Task Studies of Simulated Driving and Conversing on a Cellular Telephone." In: *Psychological Science* 12.6 (Nov. 2001), pp. 462–466. ISSN: 0956-7976. DOI: 10.1111/1467-9280.00386.

[171] Fiona M. Donald and Craig H. M. Donald. "Task disengagement and implications for vigilance performance in CCTV surveillance." In: *Cognition, Technology & Work* 17.1 (Feb. 2015), pp. 121–130. ISSN: 1435-5566. DOI: 10.1007/s10111-014-0309-8.

[172] Inseok Jang, Wondea Jung, and Poong Hyun Seong. "Human error and the associated recovery probabilities for soft control being used in the advanced MCRs of NPPs." In: *Annals of Nuclear Energy* 87 (Jan. 2016), pp. 290–298. ISSN: 0306-4549. DOI: 10.1016/j.anucene.2015.09.011.

[173] Mathias Basner et al. "Psychological and Behavioral Changes during Confinement in a 520-Day Simulated Interplanetary Mission to Mars." In: *PLOS ONE* 9.3 (Mar. 2014), e93298. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0093298.

[174] Peter Wright. "The harassed decision maker: Time pressures, distractions, and the use of evidence." In: *Journal of Applied Psychology* 59.5 (1974), pp. 555–561. ISSN: 1939-1854. DOI: 10.1037/h0037186.

[175] Willem A. Wagenaar and Gideon B. Keren. "Calibration of probability assessments by professional blackjack dealers, statistical experts, and lay people." In: *Organizational Behavior and Human Decision Processes* 36.3 (Dec. 1985), pp. 406–416. ISSN: 0749-5978. DOI: 10.1016/0749-5978(85)90008-1.

[176] George Cvetkovich. "Cognitive Accommodation, Language, and Social Responsibility." In: *Social Psychology* 41.2 (1978), pp. 149–155. ISSN: 0147-829X. DOI: 10.2307/3033574.

[177] Peter S. Kutchukian, Nadya Y. Vasilyeva, Jordan Xu, Mika K. Lindvall, Michael P. Dillon, Meir Glick, John D. Coley, and Natasja Brooijmans. "Inside the Mind of a Medicinal Chemist: The Role of Human Bias in Compound Prioritization during Drug Discovery." In: *PLOS ONE* 7.11 (Nov. 2012), e48476. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0048476.

[178] Pia Christina Sander. "Shared knowledge in complex teams: an investigation of the shared mental model construct." PhD thesis. Maastricht University, 2016. DOI: 10.26481/dis.20160331ps.

[179] Ernestina J. A. Schreuder and Tina Mioch. "The effect of time pressure and task completion on the occurrence of cognitive lockup." In: *CEUR Proceedings 4th Workshop HCP Human Centered Processes*. Feb. 2011.

[180] Mona J. Bird, David L. MacIntosh, and Phillip L. Williams. "Occupational Exposures During Routine Activities in Coal-Fueled Power Plants." In: *Journal of Occupational and Environmental Hygiene* 1.6 (June 2004), pp. 403–413. ISSN: 1545-9624. DOI: 10.1080/15459620490453346.

[181] Hosik Yoo, Jeong-Ho Lee, and Jinho Chung. "An analysis of the survey results on nuclear security culture for personnel at nuclear facilities." In: *Progress in Nuclear Energy* 112 (Apr. 2019), pp. 75–79. ISSN: 0149-1970. DOI: 10.1016/j.pnucene.2018.12.007.

[182] *Report on the effectiveness of training at operating power reactors for calendar years 2018-2021*. Tech. rep. Rockville, MD: U.S. Nuclear Regulatory Commission, May 2022.

[183] Daniel Kwasi Adjekum and Marcos Fernandez Tous. "Assessing the relationship between organizational management factors and a resilient safety culture in a collegiate aviation program with Safety Management Systems (SMS)." In: *Safety Science* 131 (Nov. 2020), p. 104909. ISSN: 0925-7535. DOI: 10.1016/j.ssci.2020.104909.

[184] Kamarizan Kidam and Markku Hurme. "Analysis of equipment failures as contributors to chemical process accidents." In: *Process Safety and Environmental Protection* 91.1-2 (Jan. 2013), pp. 61–78. ISSN: 0957-5820. DOI: 10.1016/j.psep.2012.02.001.

[185] Yaoqin Lu, Zhe Zhang, Huan Yan, Baoling Rui, and Jiwen Liu. "Effects of Occupational Hazards on Job Stress and Mental Health of Factory Workers and Miners: A Propensity Score Analysis." In: *BioMed Research International* 2020.1 (2020), p. 1754897. ISSN: 2314-6141. DOI: 10.1155/2020/1754897.

[186] William R. Warren. "The Effect of Shift Turnover Strategy and Time Pressure on Aviation Maintenance Technician Performance." PhD thesis. Daytona Beach: Embry-Riddle Aeronautical University, 2011.

[187] Jinkyun Park and Wondea Jung. "The operators' non-compliance behavior to conduct emergency operating procedures—comparing with the work experience and the complexity of procedural steps." In: *Reliability Engineering & System Safety* 82.2 (Nov. 2003), pp. 115–131. ISSN: 0951-8320. DOI: 10.1016/S0951-8320(03)00123-6.

[188] Kristin Fraser, Adam Peets, Ian Walker, Janet Tworek, Michael Paget, Bruce Wright, and Kevin McLaughlin. "The effect of simulator training on clinical skills acquisition, retention and transfer." In: *Medical Education* 43.8 (2009), pp. 784–789. ISSN: 1365-2923. DOI: 10.1111/j.1365-2923.2009.03412.x.

[189] Matthew S. Cain, Joseph E. Dunsmoor, Kevin S. LaBar, and Stephen R. Mitroff. "Anticipatory Anxiety Hinders Detection of a Second Target in Dual-Target Search." In: *Psychological Science* 22.7 (July 2011), pp. 866–871. ISSN: 0956-7976. DOI: 10.1177/0956797611412393.

[190] Walter F. Grether. *Performance and Biodynamic Stress-Influence of Interacting Stresses on Performance*. Tech. rep. Ohio: Aerospace Medical Research Laboratory, Wright- Patterson AFB, 1972.