Version of Record: https://www.sciencedirect.com/science/article/pii/S0951832024000425
Manuscript_ac23e20dc37228f244ee1151f62d8b48

Identifying human failure events for human reliability analysis: a review of gaps and research opportunities

Camille S. Levine[a], Ahmad Al-Douri[a], Vincent Philip Paglioni[a,b], Michelle Bensi[c], Katrina M. Groth[a]

[a] *Systems Risk and Reliability Analysis (SyRRA) Lab, Center for Risk and Reliability, University of Maryland, College Park, USA.*
[b] *Risk, Reliability, & Resiliency Characterization (R3C) Lab, Department of Systems Engineering, Colorado State University, Fort Collins, USA.*
[c] *Department of Civil and Environmental Engineering, University of Maryland, College Park, USA.*

Abstract

The human failure event (HFE) is the element of interest for human reliability analysis (HRA). However, the process of identifying and characterizing HFEs varies significantly between HRA methods, limiting the compatibility and traceability of results. To identify specific areas where this could be rectified, a literature review was conducted to assess the current state of research regarding the identification and underlying definitions of HFEs and their components. Current methods for identifying HFEs and their qualitative analysis are reviewed, as well as recently standardized terminology describing the breakdown of an HFE. Significant advances have been made in characterizing HFEs through macrocognitive functions, and the incorporation of HFE scenarios into larger probabilistic risk assessment (PRA) analyses. Several gaps and opportunities were found, with the most pressing issue being the dearth of HRA methods with both robust HFE characterization and quantification accuracy. Also identified was the lack of scientific basis for current complex, cognitively-based models. Other challenges include the complexity of human cognition, the various failure mechanisms that manifest in crew failure modes, the wide variety of failure scenarios (including ex-control room actions), and treatment of errors of commission. Many of these gaps are currently being pursued by interdisciplinary teams, but there is a significant amount of work to be done to address these shortcomings.

Keywords: human reliability analysis, human failure events, causal models,

Preprint submitted to Reliability Engineering & System Safety

© 2024 published by Elsevier. This manuscript is made available under the Elsevier user license
https://www.elsevier.com/open-access/userlicense/1.0/

1. Introduction

Human reliability analysis (HRA) is a complex and dynamic field with quantitative and qualitative means for assessing and addressing the human potential for contributing to failures in engineered systems. HRA is frequently used as a component in probabilistic risk assessment (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 [1]. Qualitatively, HRA produces knowledge about error causes, potential interventions, and consequences, by identifying and characterizing human failure events (HFEs). This yields valuable insights beyond numerical probabilities. As such, the accuracy and completeness of qualitative HRA 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 [2]. 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 [3]. 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 in 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 humans, hardware, and structures. 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 in HRA [4]. Notably, the HFE is a typical unit of analysis across HRA methods, but there is no established definition or procedure for

identifying HFEs in an HRA. While there are standards available, they are typically focused on PRA, rather than HRA, uses. This flexibility allows analysts to tailor the definition to the needs of the scenario; however, this leads to significant inter-analyst and inter-method variability, particularly for complex HFEs [4, 5].

Existing work has demonstrated that 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]. Other key qualities of a method include traceability, credibility, and reproducibility [6, 7]. However, previous research efforts have struggled with integrating these needs throughout the development of new methods and data sources. By enhancing HRA's causal basis in current research, analysts will be able to more easily justify their modeling choices, which should reduce inter-analyst variability. An HEP quantification method 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. 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. The remainder of this paper is structured as follows. Section 2 provides an overview of HRA and discusses the current conceptualization of the HFE and procedures for HFE identification. Section 3 details the approach used to identify technical gaps in identifying HFEs for HRA. Sections 4 and 5 discuss the gaps identified in the literature and corresponding areas for future research. Section 6 concludes with a summary of the research opportunities in HRA.

2. Overview and Procedures for HFE Identification

2.1. Definition of the Human Failure Event

There have been multiple methods developed that include guidance for identifying HFEs in HRA scenarios, including ATHEANA, SPAR-H, Phoenix,

and IDHEAS [8–11]. 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, as shown in Figure 1.

The first important concept is the HFE itself ("Objective" layer in Figure 1), 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 a randomly occurring operator error rather than the product of the event context [8]. Despite its widespread use, the HFE remained nebulously defined until very recently [4], 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 [12]: "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. For HRA purposes, the HFE is the highest-level failure considered in an HRA model. It is the culmination of the failure process initiated by human actions and must involve one or more failed Major Crew Functions (MCFs, the "Function" layer in Figure 1) [4]. This definition neatly defines the HFE as the combination of a failed objective and the process leading up to that failure. Thus, a thorough understanding of the HFE in HRA is predicated upon the rich causal structure embodied by the other elements in HRA, as discussed in the next section.

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 MCFs, which are the high-level mental or physical activities used to accomplish a goal [13]. The Phoenix HRA method provides several example MCFs for use in HRA [10]. The failure of an MCF (e.g., an incorrect decision or action) may initiate the occurrence of an HFE. MCFs represent a system-specific objective

that, if failed, leads to an HFE [4], building upon the idea of *macrocognitive functions*.

A wide variety of models have been created to categorize macrocognitive functions and describe their interactions, but a majority of these models include several functions such as Detecting/Noticing, Understanding/Sensemaking, Decision Making, Action Taking, and Coordination/Teamwork in their framework [14]. The Information-Decision-Action in Crew Context (IDAC) HRA method [15] 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 (I, D, A), where each phase could be further subdivided into a series of nested cognitive activities that also correspond to Information, Decision, and Action tasks, shown in Figure 2.

An HFE may comprise any number of MCFs from any of the I-D-A cognitive phases, depending on the complexity of the objective. For example, if multiple decisions need to be made to achieve the objective, there will be multiple Decision-phase MCFs. The failure of any of these decisions leads to an HFE. These MCFs are derived through task analysis methods, which will be discussed further in the next section.

Regardless of cognitive phase, MCF failures are caused by *crew failure modes* (CFMs), which are specific failure pathways, i.e., how the MCF failed state occurs. CFMs ("Function" layer in Figure 1) are analogous to hardware or software failure modes; for example a pump failure might be due to failure modes such as seizure, leakage, or overspeed. CFMs leading to the failure of an MCF are orthogonal (i.e., uniquely defined) 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 [13].

Human failure mechanisms represent the pathways through which CFMs might be brought about. Mechanisms ("Causal Factor" layer in Figure 1) represent real psychological, or physical processes that explain the underlying reasons for failure and are frequently represented via contextual factors [16]. Human failure mechanisms are analogous to failure mechanisms in hardware or software. For example, pump leakage (failure mode) might be due to failure mechanisms such as corrosion, wear, and/or fatigue.

The context and environment in which these failure mechanisms and failure modes operate are accounted for by *performance influencing factors*

(PIFs). PIFs ("Causal Factor" layer in Figure 1) modify the probability that cognitive mechanisms fail [14], which in turn adjusts the CFM occurrence probabilities, the MCF failure probabilities, and ultimately the *human error probability* (HEP). PIFs may be internal or external to the operator, and may fall broadly into the categories of organization-, team-, person-, situation-, or machine-based causes [17]. 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 paper. 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 PIFs [18].

Recent efforts have sought to standardize lexicographical and mathematical definitions of dependency among HFE components [13], as well as an exhaustive set of relationships that describe all possible dependency relationships between these components [19]. Standard terminology is important in discussions of dependent and causal factors to facilitate compatibility between HRA methods. As discussed above, a common vocabulary also reduces the variability of results, improving a method's traceability and reproducibility by standardizing the analysts' treatment of scenario components [13]. This ensures that HFEs between scenarios are decomposed to the same level, and that this level can be quantified using standard tools available.

2.3. Task Decomposition

Task decomposition (or task analysis) is the process of discretizing a scenario into distinct elements (Figure 1): objectives, functions, tasks, subtasks, and causal factors [13]. Part of this process entails identifying the HFEs in a scenario, for which there are broadly two methods available to analysts. The first involves identifying the potential hardware faults and deducing the human actions that must have caused them (a top-down systems approach). This top-down process is employed in methods like THERP and SPAR-H. The second approach (working from the bottom-up) directly lists out the necessary operator tasks and uses inductive reasoning to determine their potential for system safety degradation [20]. This bottom-up method is employed by EPRI's Human Factors Analysis Methodology (HFAM), which works upwards from a human factors perspective by first performing function allocation to inform a subsequent task analysis [21]. Theoretically, these methods produce similar HFEs, although there are some differences in the resulting HFEs in practice.

There are also concerns with the ability of either method to produce a comprehensive HFE set. For example, the top-down approach may miss errors of commission, which have played a significant role in previous high-profile incidents, e.g., Chernobyl and Three Mile Island [3, 20]. This is because the top-down approach takes a view of operations wherein the focus lies on (correct) actions/decisions in a scenario, for instance as outlined in a procedure. Thus, in the top-down approach, it is much easier to identify an action/decision *not* taken by a human (i.e., an error of omission) than it is to theorize an incorrect action or decision (i.e., an error of commission) [20]. Both methods face difficulties imposed by deficient or absent procedures, which is a concern not only for task analysis (particularly in the case of ex-CR actions [22]) but also for reliable operator performance [23].

Developing a comprehensive HFE set requires a comprehensive task analysis, which often uses expert elicitation and team consensus-building exercises for decomposing a given scenario. Techniques such as observation, operator verbal protocols, expert walkthroughs, and direct interviews and documentation review are used to gather information about the necessary actions operators must take [24]. Additionally, the Critical Incident Technique can be used to identify the progression of near-miss events, and Link Analysis can be used to identify prominent connections within the system [24]. These results can be described through a Hierarchical Task Analysis, Tabular Task Analysis, or timeline analysis for straightforward procedures, or Decision Action Diagrams for less linear decisions that open up multiple action paths for operators [24]. For a more thorough coverage of task analysis topics, the reader is referred to resources such as [25].

Decomposition of documentation, especially operating procedures, is an established technique for conducting task analysis in many HRA methods. Potential cognitive errors are characterized through identification of a critical path through the procedures. Critical procedural steps are any steps addressing the operation of essential equipment, as well as any emergency operating procedures used up until the success criteria of the event. Then, important procedural instructions are extracted from the critical steps, and mapped to an existing taxonomy of primitive tasks derived from simulator data [26–28]. For example, EMBRACE considers five levels of operator tasks, including ultimate plant-level goals, scenario/procedure goals, cognitive/physical actions, interface manipulation tasks, and keystroke level tasks [29–31]. The action level tasks, analogous to MCFs in the framework described in Section 2.2 above, are the main unit of analysis in EMBRACE. In a similar manner,

GOMS-HRA maps action verbs, or procedure level primitives, to a set of task level primitives which can then be analyzed through the HRA method [32]. Comprehensive modeling of necessary operator tasks is integral to accurately implementing dynamic HRA [32]. Both task analysis methods provide considerable guidance for this process, reducing inter-analyst variability, but depend heavily on the existence and quality of the procedures available.

One major aspect of the task analysis that must be determined is the "stop rule" that governs when the information and discretization of the subtasks are sufficient for the purposes of the analysis [33]. However, the specification of a stopping rule imparts subjectivity and limits the ability to reproduce the study. Two categories of stopping rules that have been studied previously include the I-D-A/dependency based rules in the Concurrent Task Analysis (CoTA) method [33], and PIF-guided stopping rules [34].

The CoTA method includes specific rules that stop the task analysis once (1) the function or task being decomposed reveals a pure Information (I), Decision (D), or Action (A) phase task, or (2) when the task represents an interaction with another agent (such as another operator, an autonomous system, or plant hardware), which indicates a task dependency [33]. This method ensures that a given subtask would have a single CFM, meaning that subtasks will have to be grouped to define multiple distinct HFEs. In a nested Information-Decision-Action structure, the pure I, D, or A task represents the smallest nested subtask under consideration (shown in Figure 2). The point of interaction with another agent in a task can include the interaction of the crew with the human-machine interface (HMI) [35].

The CoTA method's stopping rules ensure a robust assessment of the MCFs and CFMs that may occur during a scenario. However, these task-oriented stopping rules may not provide for a sufficient level of detail to assess the PIFs in a scenario. To rectify this situation, a stopping rule guided by the PIF descriptions has been explored to limit the exclusion of PIFs [34]. This stopping rule can ensure that the environmental, organizational, team-related, and personal characteristics of the scenario are included in the resulting model.

The issues discussed above with available task decomposition procedures do not represent an exhaustive list. There are other concerns related to ensuring the comprehensiveness and objectivity of a task decomposition, including lexical differences in the definitions of, e.g., objectives, functions and tasks, as well as the sufficiency and comprehensiveness of the PIF set. A general lexicon for task decomposition is offered in [13], and a comprehensive

PIF set was created by [17].

2.4. Qualitative Identification of HFEs through HRA Methods

As mentioned above, ATHEANA [8] is considered the origin of the HFE as the standard HRA unit of analysis. It defines the HFE as a representation of the failure of a function, system, or component resulting from unsafe human actions that place the plant in a worse condition (relative to the condition following a non-failure). 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 as follows, roughly corresponding to the macrocognitive functions: 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 [9] 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). For instance, one study revealed inter-analyst and inter-method variability even when using pre-defined HFEs, meaning that variability may be even higher when analysts are allowed to define the HFEs themselves [5, 36, 37].

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 an HFE to be linked to the parent PRA method, inter-analyst variability is still highly likely [4, 37]. A newer method, the Integrated Human Event Analysis System for Event and Condition Assessment (IDHEAS-ECA) [38], defines HFEs as the failure of one or more of the following five macrocognitive functions [14]: 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 5.5, which have also been lacking in prior methods.

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," "task analysis," and "dependency." Each one of these phrases was used alongside the phrase "human failure event" with the AND operator, which had the effect of sorting the results by their relevance to this review. Secondary searches for "human failure event" were conducted on 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), were also found to be relevant.

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 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 relevant to external hazards probabilistic risk assessment (XHPRA) [39, 40]. 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.

4. Overview of Identified Gaps

The results from the two activities described in Section 3 led to the synthesis of five major gaps in identifying HFEs in HRA models. Figure 3 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 5 contains a more detailed discussion of each gap. First, the quantitative, checklist-based HRA methods (methods that check for the presence of degraded contextual factors and take the product of the corresponding HEP modifiers) commonly used in PRA studies lack comprehensive HFE identification methods. 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.

5. Gaps and Opportunities for Research

5.1. Incorporate both robust HFE characterization and usable, accurate quantification 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 method of quantifying HEPs. 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. They allow the analyst to derive a reasonable HEP value, at the expense of not being able to identify the full range of relevant failure causes. Both of these ends are extremely important to plant safety: characterization of HEPs allows for the inclusion of the human element in the 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 simple, computationally cheap quantification, such as the THERP, SPAR-H, or HEART methods [9, 41, 42]. 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 method compared to others, the US HRA Empirical Study found that qualitative analysis is a shared weakness across all HRA methods [43]. HRA researchers have sought to address this gap [44], 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 (e.g., PIF)

values to find an HEP, as is done in some current HRA methods. Results can vary by orders of magnitude between methods and even between analysts [43]. 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 quantification case studies employing the Petro-HRA method, a direct offshoot of the SPAR-H method applied to the petroleum industry [45]. 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 [46]. Newer HRA methods tend to involve estimating conditional HEPs and evaluating uncertainties. These are non-trivial tasks, especially for the required scope of most HRAs [47]. It can be cumbersome to perform such calculations by hand, which is another reason why analysts might choose to employ checklist methods instead.

5.2. Expand the basis of HFE structures via scientific studies

Another research opportunity arising from the current state of HRA methods is the limited experimental and theoretical literature used to fully substantiate the causal paths through which human failure occurs. A strong literature basis is integral to the HRA analyst's ability to accurately identify and characterize HFEs. 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 IDAC method [48] 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 [15]. This error identification scheme may double-count errors, and may identify HFEs at mismatched task or objective levels.

IDHEAS [11] 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 [49]. 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 [49]. 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.

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 [16]. 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 [13]. 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 [50, 51].

These mechanisms (denoted in Figure 1 as "MECH") can be observed in describing how PIFs could cause a particular failure mode to arise. For example, an operator may misinterpret data (the failure mode) due to a

failure in *attention* (the failure mechanism), which in turn can be affected by stress, workload, or internal biases. Workload can also affect the operator's perceived severity of the scenario, subsequently causing *bias* towards certain interpretations of the data, which is another failure mechanism. It is clear that there are many causal pathways leading to this failure mode stemming from different personal, environmental, and task-related causes, and these factors interact with each other to increase or decrease the likelihood of failure [52]. Certain PIFs, particularly *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 [16]. By identifying human failure mechanisms, this may provide a rationale for the identification of more important PIFs to human performance [37].

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 quantitative data, which has the power to either justify or challenge current expert judgment or laboratory experiments [53].

Modeling scenario complexity presents a non-trivial research challenge. Frequently, identification of these deeper error causes necessitates deliberation from multi-disciplinary research teams [30]. To achieve a higher model fidelity, causal pathways beyond the direct consequences of PIFs on overall human error probabilities must be acknowledged [6]. HRA methods such as Phoenix [10] with a strong theoretical and cognitive basis, particularly those including a suitable set of pre-defined CFMs, can enable the analysts to identify relevant human failure mechanisms.

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 [54]. 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 [48]. Each tendency is described and associated with one or two PIFs from the set identified earlier in the method [18]. 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 [55]. 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 [55] either; the output describes operator decision-making strategy as the scenario progresses in time.

Other models, such as CMS-BN [56], 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 timestep. 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.

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) [57]. However, the Fukushima Daiichi disaster in March 2011 has 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 will be critical to obtaining accurate risk profiles of a nuclear power plant site, and a recent review by [58]

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 [59]. St. Germain et al. [60] 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 (ERO) 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 [61] 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 [62]. Additional research activities have recognized the inadequacy of HRA methods in addressing human and organizational factors for multi-unit scenarios for severe accident management. [58] 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

due to external events 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 [63].

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 [64]. It is also particularly difficult to identify scenario-initiating human actions because most prior research has examined pre- and post-initiator human actions [65]. Even for post-initiator events, most current HRA methods do not explicitly identify EOC; rather, the EOC effect is implicitly included in the HEP [51, 66]. 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 [65, 67].

Various EOC search schemes have been developed to address widely used HRA methods' lack of support for EOC identification. The Borssele-global misdiagnosis method [68] begins by assuming that a scenario has occurred, and searches the response procedure for potential erroneous actions. Other methods, like ATHEANA [8], also assume that a scenario has occurred but instead examine necessary system functions and how operators could cause them to fail. Finally, methods such as Commission Errors Search and Assessment (CESA) [64] list both proceduralized and non-proceduralized actions, then identify events linked to important systems, and identify 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 nuclear power plant HRA because of the procedure-dominated nature of the environment [65].

A shortcoming of relying on procedure analysis is that it excludes non-proceduralized actions. EOC are frequently associated with maintenance and abnormal operations [67] which may not have (complete) procedures available. 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 [22]. 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. EOC 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 EOC 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 EOC than a top-down search scheme that prioritizes system functions [69]. The bottom-up human factors approach was found to be suitable for identifying HFEs for ex-control room actions during external hazard events [70]. 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 IDAC 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 [71]. It uses macrocognitive functions and proximate causes of failure (analogous to CFMs) to map generic tasks associated with radiotherapy administration to Groth's PIF hierarchy [17]. 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 [72, 73] support EOC identification via task analysis, although they are developed for HFEs involving information-gathering and action-taking tasks. The IDHEAS method [11] can be used as a more complete catalog of cognitive errors, considering a wider variety of decision-making failure modes [74]. The Phoenix method has a similarly strong catalog of potential CFMs [10]. However, there remains a lack of standardization surrounding the identification of EOC HFEs through this identification of macrocognitive failure points.

6. Conclusions

In this paper, we present a detailed review of the current analysis methods and terminology for identifying and characterizing HFEs in HRA studies. From a set of 470 publications, we identified 57 as the most relevant to HFE identification. We also used ongoing project activities on identifying uncertainties in external hazards PRA to corroborate the literature findings. Through this two-pronged approach, we enumerated five major research opportunities related to HFE identification and characterization. These opportunities are:

1. Incorporation of robust HFE characterization and usable, accurate quantification schemes into HRA methods.
2. Expansion of scientific research basis for HFE structures.
3. Capturing how the complexity of human cognition manifests in failures through the implementation of human failure mechanisms.
4. Expansion of HFE identification and treatment in the context of multi-unit and external hazards PRA studies.
5. Improving the approach to characterizing HFEs stemming from EOC.

This review was limited to publicly-available English-language documents available on Google Scholar, which excluded many international sources and technical reports from the search. We encourage the broader HRA community to both maintain awareness of the research opportunities and exchange information on advances that improve upon the gaps enumerated here. It is only through the resolute efforts of the broad HRA community that we will be able to address each of these gaps.

Ongoing and future research activities should continue to advance the scientific basis of HFE identification and characterization through causally robust models, leveraging human factors literature to capture and model the complexities of human cognition and activity, and expanding applicability to a wider range of risk contexts and CFMs.

7. Acknowledgments

This material is based upon work supported in part by the Department of Energy Office of Nuclear Energy under Award Number DE-NE0008974 and

by the U.S. Nuclear Regulatory Commission, grant number 31310020M0002. Additionally, this material is based upon work supported under a Department of Energy, Office of Nuclear Energy, University Nuclear Leadership Program Graduate Fellowship. The authors would like to thank the A. James and Alice B. Clark Foundation for their support through the Clark Doctoral Fellows Program at the University of Maryland.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

8. CRediT authorship contribution statement

Camille S. Levine: Conceptualization, Methodology, Investigation, Writing- Original Draft, Writing- Review and Editing **Ahmad Al-Douri**: Writing- Original Draft, Writing- Review and Editing, Investigation **Vincent P. Paglioni**: Writing- Original Draft, Writing- Review and Editing, Investigation **Michelle Bensi**: Writing- Review and Editing, Funding acquisition **Katrina M. Groth**: Writing- Original Draft, Supervision, Funding acquisition, Project administration

References

[1] J.-E. Holmberg, M. Liinasuo, Non-PSA applications of HRA, Tech. Rep. 14124_R003, VTT Technical Research Centre of Finland (2017).

[2] S. French, S. Adhikari, C. Bayley, T. Bedford, J. Busby, A. Cliffe, G. Devgun, M. Eid, R. Keshvala, S. J. T. Pollard, E. Soane, D. Tracy, S. Wu, Human Reliability Analysis: A review and critique, Manchester Business School Working Paper 589 (2009).

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

[4] V. P. Paglioni, T. Mortenson, K. 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), Honolulu, 2022.

[5] H. Liao, J. Forester, V. N. Dang, A. Bye, Y. H. J. Chang, E. Lois, Assessment of HRA method predictions against operating crew performance: Part II: Overall simulator data, HRA method predictions, and intra-method comparisons, Reliability Engineering & System Safety 191 (2019) 106510.

[6] K. M. Groth, R. Smith, R. Moradi, A hybrid algorithm for developing third generation HRA methods using simulator data, causal models, and cognitive science, Reliability Engineering and System Safety 191 (2019) 106507. doi:10.1016/j.ress.2019.106507.

[7] A. Mosleh, Y. H. Chang, Model-based human reliability analysis: prospects and requirements, Reliability Engineering & System Safety 83 (2) (2004) 241–253. doi:10.1016/j.ress.2003.09.014.

[8] S. E. Cooper, A. M. Ramey-Smith, J. Wreathall, G. W. Parry, D. C. Bley, W. J. Luckas, J. H. Taylor, M. T. Barriere, NUREG-6350: A Technique for Human Error Analysis (ATHEANA), Tech. rep., US Nuclear Regulatory Commission (1996).

[9] D. Gertman, H. Blackman, J. Marble, J. Byers, C. Smith, NUREG-6883: The SPAR-H Human Reliability Analysis Method, Tech. rep., Idaho National Laboratory, US Nuclear Regulatory Commission, Idaho Falls (2005).

[10] N. J. Ekanem, A. Mosleh, S. H. Shen, Phoenix - A model-based Human Reliability Analysis methodology: Qualitative Analysis Procedure, Reliability Engineering and System Safety 145 (2016) 301–315. doi:10.1016/j.ress.2015.07.009.

[11] J. Xing, J. Chang, J. DeJesus, The general methodology of an Integrated Human Event Analysis System (IDHEAS-G), Tech. Rep. NUREG-2198, US Nuclear Regulatory Commission (May 2021).

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

[13] V. P. Paglioni, K. M. Groth, Dependency definitions for quantitative human reliability analysis, Reliability Engineering & System Safety 220 (2022) 108274. doi:10.1016/j.ress.2021.108274.

[14] A. M. Whaley, J. Xing, R. L. Boring, S. M. L. Henrickson, J. C. Joe, K. L. Le Blanc, S. L. Morrow, Cognitive Basis for Human Reliability Analysis (NUREG - 2114), Tech. Rep. NUREG/CR-2114, US Nuclear Regulatory Commission (Jan. 2016).

[15] Y. H. J. Chang, A. Mosleh, Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1: Overview of the IDAC Model, Reliability Engineering & System Safety 92 (8) (2007) 997–1013. doi:10.1016/j.ress.2006.05.014.

[16] V. P. Paglioni, C. Levine, A. Al-Douri, K. 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), Knoxville, TN, 2023.

[17] K. M. Groth, A. Mosleh, A data-informed PIF hierarchy for model-based Human Reliability Analysis, Reliability Engineering and System Safety 108 (2012) 154–174.

[18] Y. H. Chang, A. Mosleh, Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 2: IDAC performance influencing factors model, Reliability Engineering and System Safety 92 (8) (2007) 1014–1040. doi:10.1016/j.ress.2006.05.010.

[19] V. P. Paglioni, K. M. Groth, Dependency idioms for quantitative Human Reliability Analysis, Nuclear Science and Engineering 0 (0) (2023) 1–23. doi:10.1080/00295639.2023.2250159.

[20] R. L. Boring, J. C. Joe, Task Decomposition in Human Reliability Analysis, in: Probabilistic Safety Assessment and Management, Honolulu, 2014.

[21] M. Presley, M. Gibson, A unified approach: Human factors, human reliability and human performance data for digital systems, in: 13th Nuclear Plant Instrumentation, Control & Human-Machine Interface Technologies (NPIC&HMIT 2023), American Nuclear Society, Knoxville, TN, 2023, pp. 178–187. doi:10.13182/NPICHMIT23-41265.

[22] R. Prasad, G. Coles, A. Dalton, N. Kohn, K. Branch, A. Bittner, R. Taylor, Effects of environmental conditions on manual actions for flood protection and mitigation, Tech. Rep. NUREG/CR-7256, US Nuclear Regulatory Commission (May 2020).

[23] IAEA (International Atomic Energy Agency), Nuclear Power Plant Operating Experience from the IAEA/NEA International Reporting System for Operating Experience 2015-2017, Tech. rep., International Atomic Energy Agency, Vienna (2020).

[24] C. Taylor, Task analysis as a cornerstone technique for human reliability analysis, in: Advances in Human Error, Reliability, Resilience, and Performance: Proceedings of the AHFE 2017 International Conference on Human Error, Reliability, Resilience, and Performance, Springer, Los Angeles, CA, 2018, pp. 86–95.

[25] B. Kirwan, L. K. Ainsworth (Eds.), A Guide To Task Analysis: The Task Analysis Working Group, CRC Press, 1992. doi:10.1201/b16826. URL https://www.taylorfrancis.com/books/9781482263053

[26] Y. Kim, J. Park, W. Jung, A classification scheme of erroneous behaviors for human error probability estimations based on simulator data, Reliability Engineering & System Safety 163 (2017) 1–13. doi:10.1016/j.ress.2017.01.022.

[27] Y. Kim, J. Kim, J. Park, S. Choi, H. Kim, An HRA method for digital main control rooms - part II: Estimating the failure probability due to

cognitive error, Tech. Rep. KAERI/TR-8065/2020, KAERI, Daejeon, South Korea (Sep. 2020). doi:10.13140/RG.2.2.34243.63524.

[28] Y. Kim, Y. H. J. Chang, J. Park, L. Criscione, SACADA and HuREX part 2: The use of SACADA and HuREX data to estimate human error probabilities, Nuclear Engineering and Technology 54 (3) (2022) 896–908. doi:10.1016/j.net.2021.09.003.

[29] Y. Kim, J. Park, A framework of HRA for computer-based control room, in: Proceedings of the 29th European Safety and Reliability Conference (ESREL), Research Publishing Services, 2019, pp. 180–184. doi:10.3850/978-981-11-2724-3_0197-cd.

[30] Y. Kim, S. Choi, S. Kim, J. Park, Direction of the Human Reliability Analysis method development for computer-based control rooms, in: Enlarged Halden Programme Group Meeting, Sandefjord, Norway, 2019.

[31] Y. Kim, Considerations for generating meaningful HRA data: Lessons learned from HuREX data collection, Nuclear Engineering and Technology (2020). doi:10.1016/j.net.2020.01.034.

[32] R. Boring, M. Rasmussen, T. Ulrich, S. Ewing, D. Mandelli, Task and procedure level primitives for modeling human error, in: International Conference on Applied Human Factors and Ergonomics, Springer, 2017, pp. 30–40.

[33] M. A. Ramos, C. A. Thieme, I. B. Utne, A. Mosleh, Human-system concurrent task analysis for maritime autonomous surface ship operation and safety, Reliability Engineering & System Safety 195 (2020) 106697. doi:10.1016/j.ress.2019.106697.

[34] C. Blackett, S. Øie, N. Paltrinieri, Human reliability in the petroleum industry: A Case study of the Petro-HRA method, in: Proceedings of the 2016 European Safety and Reliability Conference (ESREL 2016), Glasgow, Scotland, 2016.

[35] G.-O. Strand, C. Haskins, On linking of task analysis in the HRA procedure: The case of HRA in offshore drilling activities, Safety 4 (3) (2018) 39. doi:10.3390/safety4030039.

[36] H. Liao, J. Forester, V. N. Dang, A. Bye, Y. H. J. Chang, E. Lois, Assessment of HRA method predictions against operating crew performance: Part I: Study background, design and methodology, Reliability Engineering & System Safety 191 (2019) 106509.

[37] H. Liao, J. Forester, V. N. Dang, A. Bye, Y. H. J. Chang, E. Lois, Assessment of HRA method predictions against operating crew performance: Part III: Conclusions and achievements, Reliability Engineering & System Safety 191 (2019) 106511. doi:10.1016/j.ress.2019.106511.

[38] J. Xing, J. Chang, J. DeJesus Segarra, Integrated Human Event Analysis System for Event and Condition Assessment (IDHEAS-ECA), Tech. Rep. NUREG-2256, US Nuclear Regulatory Commission (Oct. 2022).

[39] M. Bensi, K. M. Groth, Z. Wu, R. Schneider, S. Mohammadi, T. Liu, A. Al-Douri, C. 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, Honolulu, HI, 2022.

[40] M. Bensi, Z. Wu, K. Groth, Z. Ma, R. Schneider, T. Liu, K. Faraji Najarkolaie, A. Al-Douri, C. 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), Knoxville, TN, 2023.

[41] A. Swain, H. Guttmann, NUREG/CR-1278: Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, Tech. rep., Sandia National Laboratories, US Nuclear Regulatory Commission (1983).

[42] J. Williams, HEART–A proposed method for assessing and reducing human error, in: Ninth Advances in Reliability Technology Symposium, Bradford, UK, 1986.

[43] J. Marble, H. Liao, M. Presley, J. Forester, A. Bye, V. Dang, E. Lois, Results and insights derived from the intra-method comparisons of the US HRA empirical study, Proceedngs of the 11th Probabilistic Safety Assessment and Management/European Safety and Reliability 2012 (2012) 12.

[44] S. L. Hendrickson, G. Parry, J. A. Forester, V. N. Dang, A. M. Whaley, S. Lewis, E. Lois, J. Xing, Towards an improved HRA method., Tech. Rep. SAND2012-1319C, Sandia National Laboratories (Feb. 2012).

[45] C. Blackett, S. Øie, K. Gould, Lessons learned from applying a new HRA method for the petroleum industry, Reliability Engineering & System Safety (Oct. 2018). doi:10.1016/j.ress.2018.10.001.

[46] V. Garg, G. Vinod, V. Kant, Auto-CREAM: Software application for evaluation of HEP with basic and extended CREAM for PSA studies, Reliability Engineering & System Safety 236 (2023) 109318. doi:10.1016/j.ress.2023.109318.

[47] J. Kim, W. Jung, J. 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, 2007, pp. 314–321. doi:10.1109/HFPP.2007.4413226.

[48] Y. H. J. Chang, A. 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, Reliability Engineering & System Safety 92 (8) (2007) 1061–1075. doi:10.1016/j.ress.2006.05.011.

[49] K. Zwirglmaier, D. Straub, K. M. Groth, Capturing cognitive causal paths in human reliability analysis with Bayesian network models, Reliability Engineering and System Safety 158 (October 2016) (2017) 117–129. doi:10.1016/j.ress.2016.10.010.

[50] R. L. Boring, M. Rasmussen, T. A. Ulrich, N. 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, Los Angeles, CA, 2018.

[51] Y. Kim, S. Y. Choi, J. Park, J. Kim, Empirical study on human error probability of procedure-extraneous behaviors, Reliability Engineering & System Safety 227 (2022) 108727. doi:10.1016/j.ress.2022.108727.

[52] C. S. Levine, A. Al-Douri, K. 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), Knoxville, TN, 2023.

[53] K. M. Groth, A. Mosleh, Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model, Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 226 (4) (2012) 361–379. doi:10.1177/1748006X11428107.

[54] B. Polat, A model-based human reliability analysis methodology for the physical protection of nuclear facilities, Ph.D. thesis, North Carolina State University, Raleigh, NC (2022).

[55] Y. H. J. Chang, A. Mosleh, Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 5: Dynamic probabilistic simulation of the IDAC model, Reliability Engineering & System Safety 92 (8) (2007) 1076–1101. doi:10.1016/j.ress.2006.05.012.

[56] Y. Zhao, C. Smidts, CMS-BN: A cognitive modeling and simulation environment for human performance assessment, part 1 — methodology, Reliability Engineering & System Safety 213 (2021) 107776. doi:10.1016/j.ress.2021.107776.

[57] A. Mosleh, PRA: a perspective on strengths, current limitations, and possible improvements, Nuclear Engineering and technology 46 (1) (2014) 1–10.

[58] T. Zhou, M. Modarres, E. L. Droguett, Multi-unit nuclear power plant probabilistic risk assessment: A comprehensive survey, Reliability Engineering & System Safety 213 (2021) 107782.

[59] Y. Akl, K. Fleming, Summary report of the international workshop on multi-unit probabilistic safety assessment, Tech. rep., Canadian Nuclear Safety Commission, Ottawa, Canada (2015).

[60] S. St Germain, R. Boring, G. Banaseanu, Y. Akl, H. Chatri, Multi-unit considerations for human reliability analysis, Tech. rep., Idaho National Laboratory, Idaho Falls, ID (2017).

[61] J. Park, A. M. Arigi, J. Kim, Treatment of human and organizational factors for multi-unit HRA: Application of SPAR-H method, Annals of Nuclear Energy 132 (2019) 656–678. doi:10.1016/j.anucene.2019.06.053.

[62] A. M. Arigi, G. Park, J. Kim, Dependency analysis method for human failure events in multi-unit probabilistic safety assessments, Reliability Engineering and System Safety 203 (December 2019) (2020) 107112. doi:10.1016/j.ress.2020.107112.

[63] R. Bisio, A. Fernandes, C. 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, Venice, Italy, 2020.

[64] B. Reer, V. N. Dang, S. Hirschberg, The CESA method and its application in a plant-specific pilot study on errors of commission, Reliability Engineering & System Safety 83 (2) (2004) 187–205. doi:10.1016/j.ress.2003.09.010.

[65] A. M. Arigi, J. Park, H.-K. Lim, J. Kim, Analysis of human-induced initiating events in a multi-unit loss of offsite power scenario, Journal of Nuclear Science and Technology 57 (1) (2020) 121–132. doi:10.1080/00223131.2019.1661305.

[66] X. He, C. Andersson, A. Olsson, J. Ljungbjörk, A. Karlsson, L. Nordlof, K. Gustafsson, L. Tunturivuori, P. Hellström, Errors of commission in HRA – NPSAG Phase 1 project, in: Probabilistic Safety Assessment and Management (PSAM) 14, Los Angeles, 2018.

[67] Y. Kim, J. Kim, Identification of human-induced initiating events in the low power and shutdown operation using the Commission Error Search and Assessment method, Nuclear Engineering and Technology 47 (2) (2015) 187–195. doi:10.1016/j.net.2014.12.006.

[68] J. Julius, E. Jorgenson, G. W. Parry, 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, Reliability Engineering & System Safety 50 (2) (1995) 189–201. doi:10.1016/0951-8320(95)00075-D.

[69] R. L. Boring, Top-down and bottom-up definitions of human failure events in Human Reliability Analysis, Proceedings of the Human Factors and Ergonomics Society Annual Meeting 58 (1) (2014) 563–567.

[70] A. Al-Douri, C. Levine, K. M. Groth, Identifying human failure events (HFEs) for external hazard Probabilistic Risk Assessment, Reliability Engineering & System Safety 235 (2023). doi:https://doi.org/10.1016/j.ress.2023.109236.

[71] D. Pandya, L. Podofillini, F. Emert, A. J. Lomax, V. N. Dang, Developing the foundations of a cognition-based human reliability analysis model via mapping task types and performance-influencing factors: Application to radiotherapy, Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 232 (1) (2018) 3–37. doi:10.1177/1748006X17731903.

[72] D. 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, Knoxville, TN, 1986, pp. 184–193.

[73] S. Shorrock, B. Kirwan, TRACEr: A technique for the retrospective of cognitive errors in ATM, Engineering Psychology and Cognitive Ergonomics: Volume Three-Transportation Systems, Medical Ergonomics and Training (1999).

[74] C. Blackett, J. E. Farbrot, S. Oie, M. Fernander, The Petro-HRA guideline, Rev. 1, Vol. 2, Tech. Rep. IFE/E-2022/002, Institute for Energy Technology (Jan. 2022).

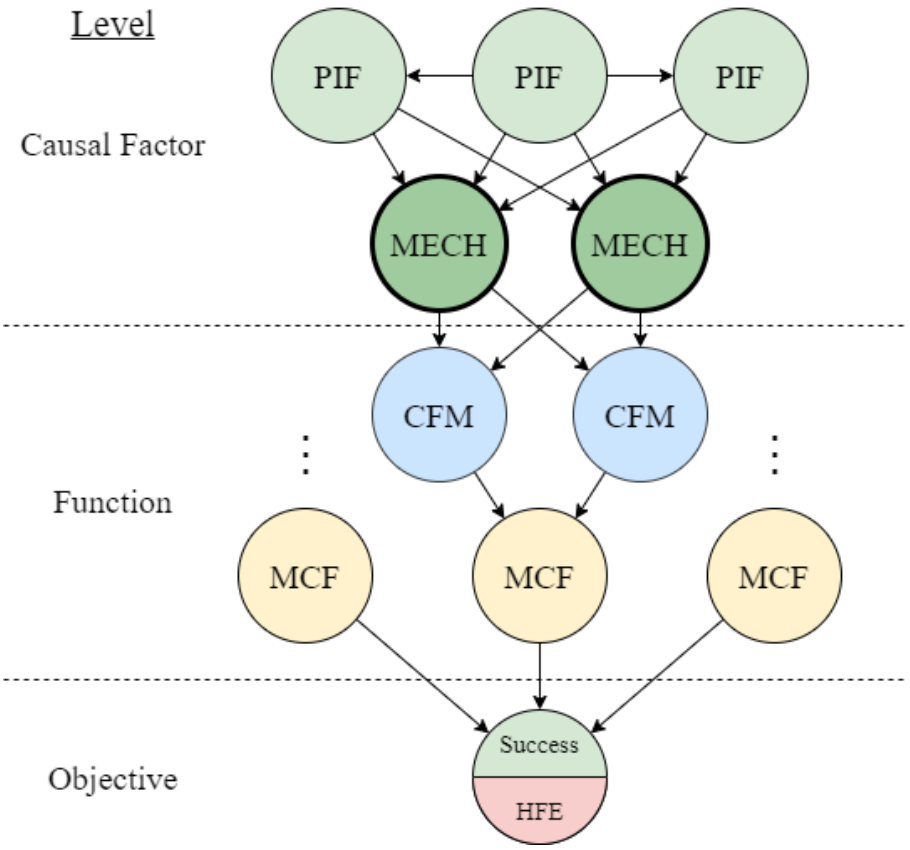

Figure 1: Cognitive structure of the HFE shown in a Bayesian Network framework, connecting performance influencing factors (PIFs) as causes of crew failure mechanisms (MECH), in turn causing crew failure modes (CFMs), which lead to failure of major crew functions (MCFs), which define whether crews fail key objectives (HFEs).

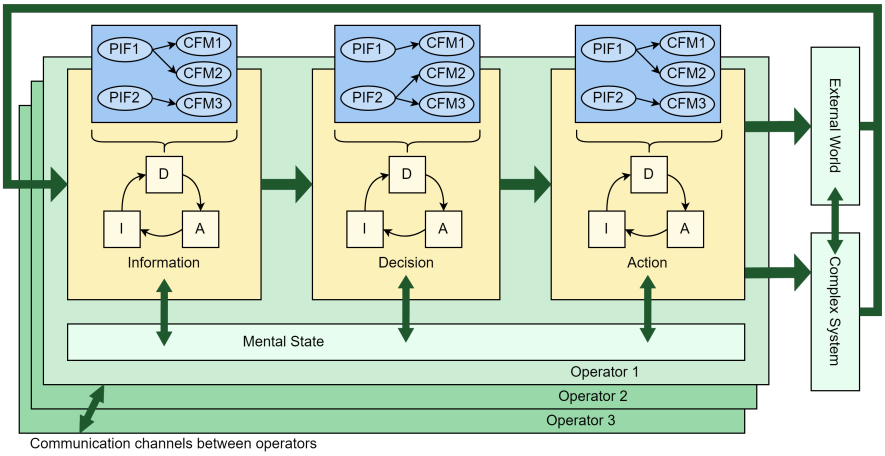

Figure 2: Nested structure of Information-Decision-Action macrocognitive functions.

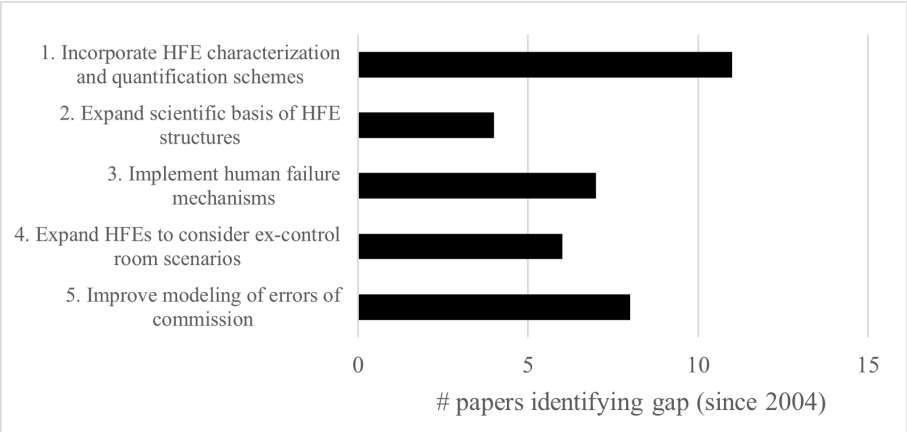

Figure 3: Number of relevant publications to the five gaps in HFE identification and characterization for HRA.

- Current state of literature on HFE identification and characterization is presented
- Identified opportunity for robust, usable HFE characterization and quantification
- Need for expanding scientific and reliability engineering basis of HRA models
- Opportunity to expand HFE characterization to additional failure domains, scenarios

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: