A Model-Based Human Reliability Analysis Framework

Ali Mosleh[a] , John A. Forester[b] , Ronald L. Boring[c], Stacey M. L. Hendrickson[c], April M. Whaley[c] , Song-Hua Shen[d], , Dana L. Kelly[c], James Y.H. Chang[d], Vinh N. Dang[e] Johanna H. Oxstrand[f], Erasmia L. Lois[d]

[a]University of Maryland, College Park, MD, USA
[b]Sandia National Laboratories, Albuquerque, NM, USA
[c]Idaho National Laboratory, Idaho Falls, ID, USA
[d]US Nuclear Regulatory Commission, Washington, DC, USA
[e]Vinh N. Dang, Paul Scherrer Institute, Villigen PSI, Switzerland
[f]Vattenfall Ringhals AB, Väröbacka, Sweden

Abstract: In response to a Staff Requirements Memorandum (SRM) to the Advisory Committee on Reactor Safeguards (ACRS), the US Nuclear Regulatory Commission (NRC) has undertaken a research effort to create a consensus approach to human reliability analysis (HRA). This paper provides an overview of the approach being developed. The approach introduces the "crew response tree" (CRT) concept, which depicts the human failure events in a manner parallel to the PRA event tree process, provides a structure for capturing the "context" associated with the human failure events under analysis, and uses the Information Processing Model as a platform to identify potential failures. It incorporates behavioral science knowledge by providing the decompositions of human failures/failure mechanisms/failure factors built from a top-down and bottom-up approach, the latter reflecting those findings from scientific papers that document theories and data of interest. The structure provides a roadmap for incorporating the phenomena with which crews would be dealing, the plant characteristics (e.g., design, indications, procedures, training), and human performance capabilities (awareness, decision, action). In terms of quantification, the approach uses the typical PRA conditional probability expression, which is delineated to a level adequate for associating the probability of a human failure event with conditional probabilities of the associated contexts, failure mechanisms, and the underlying factors (e.g., performance shaping factors). Such mathematical formulation can be used to directly estimate HEPs using various data sources (e.g., expert estimations, anchor values, simulator or historical data), or can be modified to interface with existing quantification approaches. [1]

Keywords: HRA, Failure Mechanism, Cognitive Model, IDA.

1. INTRODUCTION

In a Staff Requirements Memorandum (SRM) (SRM-M061020) to the Advisory Committee on Reactor Safeguards (ACRS), the Commission directed the ACRS to "work with the staff and external stakeholders to evaluate the different human reliability models in an effort to propose a single model for the agency to use or guidance on which model(s) should be used in specific circumstances."[1]

As a first step toward meeting this directive, a survey of various methods and user needs was performed, followed by a workshop of experts in human reliability analysis (HRA) and related domains to discuss the outline of an approach to address the SRM concerns. The workshop in particular identified a set of desirable attributes of a robust HRA method, which could also be used as a set of criteria in evaluating various methods. The attributes are:

[1] The information presented in this paper does not currently represent an agreed-upon NRC staff position. The NRC has neither approved nor disapproved its technical content.

- Content Validity (coverage of plant, crew, cognition, action, errors of commission , errors of omission, etc)
- Explanatory power, "causal model" for error mechanisms and relation to context, theoretical foundations
- Ability to cover HFE dependency and recovery
- Clear definition of "unit of analysis" and level of detail for various applications
- Empirical Validity (of HEPs), e.g., having basis in Operational Data, Simulator Experiments, Other Industries
- Reliability (Reproducibly, Consistency, Inter- and Intra-rater Reliability)
- Traceability/Transparency (ability to reverse engineer analysis)
- Testability (of part or the entire model and analysis)
- Capability for Graded Analysis (screening, scoping, detailed analysis)
- Usability/Practicality

A preliminary assessment of existing methods indicated that none satisfy all of the above criteria. A key observation made during the workshop was that improving the qualitative aspects of HRA methods could increase their robustness and reduce some of the sources in the variability of results we see in the applications of different methods, as well as in cases where the same method is used by different analysts. This observation has been corroborated through a number of other studies, most notably the International HRA Empirical Study recently conducted at Halden [2].

Accordingly, an effort has been undertaken to develop a comprehensive qualitative analysis method that leverages the past experience and advances introduced in the second generation and emerging third generation HRA methods. As such, the method under development is a hybrid of the best features of the existing methods and frameworks. More importantly, as per the requirement articulated in the aforementioned workshop, this method formally incorporates relevant psychological and cognitive theories in its core human performance model, on which the qualitative analysis tools and procedures are built. This approach is intended to be applicable at a minimum for HRA within a full-power internal events probabilistic risk assessment (PRA), as well as for evaluating low-power shutdown (LPSD) operations.

This paper provides an overview of the approach being developed. The approach introduces the "crew response tree" (CRT) concept, which depicts the human failure events in a manner parallel to the PRA event tree process, provides a structure for identifying the "context" associated with the human failure events under analysis, and uses the Information Processing Model as a platform to identify potential failures. It incorporates behavioral science knowledge by providing the decompositions of human failures/failure mechanisms/failure factors built from a top-down and bottom-up approach, the latter reflecting those findings from scientific papers that document theories and data of interest. The structure provides a roadmap for incorporating the phenomena with which crews would be dealing, the plant characteristics (e.g., design, indications, procedures, training), and human performance capabilities (awareness, decision, action). The structure's aim is to create a template-based guide for a consistent, efficient, and effective analysis. In terms of quantification, the approach uses the typical PRA conditional probability expression, which is delineated to a level adequate for associating the probability of a human failure event with conditional probabilities of the associated contexts, failure mechanisms, and the underlying factors (e.g., performance shaping factors). Such mathematical formulation can be used to directly estimate HEPs using various data sources (e.g., expert estimations, anchor values, simulator or historical data), or can be modified to interface with existing quantification approaches. At this time, the quantification approach is still under exploration. Three companion papers in this conference proceedings [3] [4] [5] discuss various aspects of the methods being developed and simple examples to communicate the concepts, analysis procedures, and technical challenges ahead.

2. PRELIMINARIES

The modelling and analysis scope includes the individual operators, the crew as a whole, and the plant interact dynamically within a physical and organizational environment. There are two crew modeling options, (1) Crew as the "unit of analysis" and (2) Individual Operator as the "unit of analysis" on which a crew response model is built. The first approach is appealing, as it simplifies the modeling of many of the interface issues between the plant and the operators. The proposed approach uses option 1, acknowledging that it entails approximations and abstractions of the more complex reality. This means, for instance, that some team behavioral aspects that emerge from individual behaviors are viewed as crew characteristics.

Some of the key characteristics of the nuclear power plant control room environment that need to be considered in HRAs are:

Information Rich – Operators need to filter information to focus on important parameters/alarms/verbal communication.
Procedurally Driven – Actions are largely guided by written procedures.
Highly regulated – Actions must comply with technical specifications, emergency plans, quality assurance requirements, organizational expectations, etc.
Time Constrained – Some actions must be accomplished quickly, under high workload conditions.
Crew Environment – Typical crew complement includes shift manager, senior operator, two or more reactor operators, and several non-licensed auxiliary operators.

3. MODELING FRAMEWORK

The proposed approach uses two modeling vehicles:

- A process and representational method for analyzing crew-plant interactions, focusing on identifying and quantifying human failure events (HFEs) and recovery
- A model of human response relating the observable human (crew) "failure modes" to "context factors"

The first modeling tool is a forward-branching tree of operator (crew) cognitive activities and actions (Crew Response Tree, CRT). The relation between CRT and typical PRA events tree is shown in Figure 1. The top tree (above the time arrow) is the plant event tree for an initiating event with system failures and human failure events (HFEs). CRT (the tree below the time arrow) is a supporting tree, providing a causal explanation of the HFEs that will appear in the PRA event trees or fault trees (symbolically shown as dashed lines in Figure 3.2). The link is a "bookkeeping" link to help analysts keep track of the relation between the CRT scenarios and event tree scenarios. No formal logical or mathematical link is proposed.

Depending on whether the unit of analysis is the crew or individual operators, the "branch points" of the CRT can include (1) operator action options, (2) operator decision options, (3) crew member interactions, and (4) key plant functional states. Both the ET and the CRT are synchronized (symbolized by the green time arrow in Figure 1), and their dynamic nature is mostly implicit. Both start at the initiating event (for full power applications). CRT is developed by an interdisciplinary team of PRA analysts as it requires knowledge of plant behavior and human response. CRT is an analysis aid to ensure systematic coverage of crew-plant interactions consistently with the scope of the analysis. As such, the CRT identifies the opportunities for and types of HFEs in the context of an accident or plant upset. We note that event trees essentially play a similar role in PRA, although the level of resolution is normally inadequate for HRA analysis. CRTs are also envisioned as an HRA work product, a means of documenting and reporting HRA analysis.

Each sequence of events (decisions, actions, HFE, recovery action, etc.) in the CRT is a pictorial

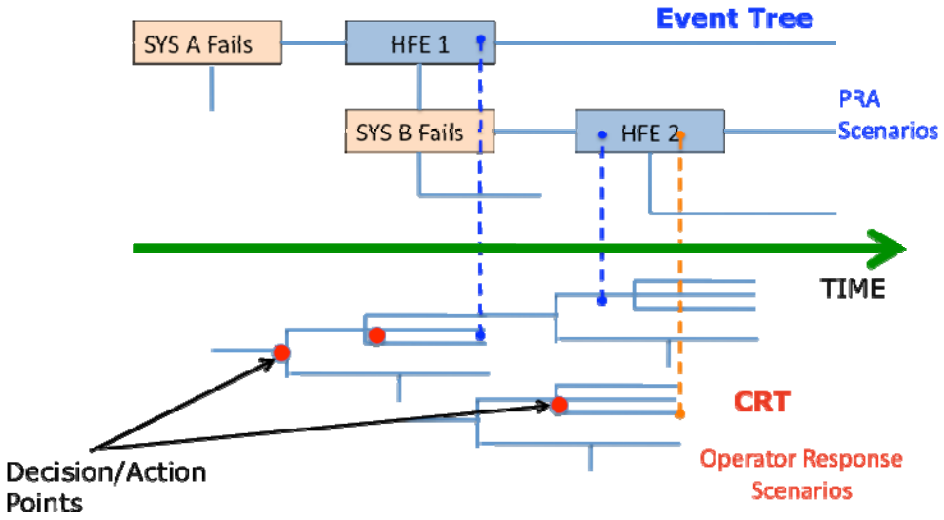

Figure 1 Plant and Crew Interaction Modeling Through CRTs

representation of a possible crew response scenario in the entire accident (from initiation to termination). The formalism can help "standardize" the qualitative analysis of HRAs and provide a common language for different HRA analysis styles, such as PSF-based, or methods that work with narrative or operational stories.

Referring to Figure 2, the "context" of an HFE includes:

- The portion of the specific PRA event tree scenario that leads to the HFE of interest
- The corresponding time from the start of the scenario (yellow highlight)
- The portion of the specific CRT scenario that leads to the HFE of interest (yellow highlight)
- All other relevant plant and crew "factors" not shown on ET and CRT, including:
  - Relevant plant physical parameter displays, alarms, and indicators
  - System and component functional states
  - External conditions, such as harsh environment
  - Other human and organizational performance-influencing/shaping factors

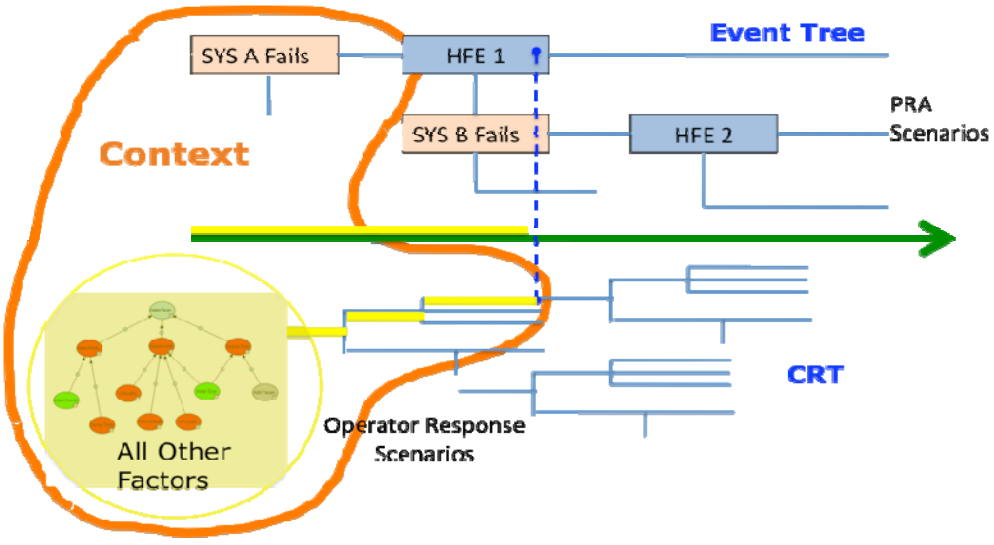

Figure 2. "Context" According to the Proposed Framework

In developing the CRT and using it to describe the various scenarios leading to each HFE, all context factors are considered, and most will have an explicit representation in the CRT or supporting sub-modes (see Section 4) and information.

4. CREW PERFORMANCE MODEL

CRT sequences and branches capture some but not all of the context factors and causes of operator responses. The branches of CRT are mainly at the functional level and do not typically cover the "human failure mechanisms" or their causes. This aspect is delegated to a set of supporting models of crew behavior in the form of causal trees. The overarching human response model adopted in for this purpose is the human information processing cognitive model. An implementation of this modeling concept is the IDA [6] approach, which has provided a basis for developing the proposed framework. Figure 3 is a schematic representation of the main elements of the IDA modeling concept as described in [6] and its key elements in the form of I-D-A dynamic loop for each member of the crew. The figure shows possible lines of interaction (information exchange and actions) between an operator, plant/system, and other crew members. Such interactions are influenced by internal and external factors (e.g., organization and environment). Given incoming information, the crew model generates a response, linking the context to the action through explicit causal models. Failure mechanisms are linked to the possible human failures identified within the CRTs on the basis of the IDA cognitive model.

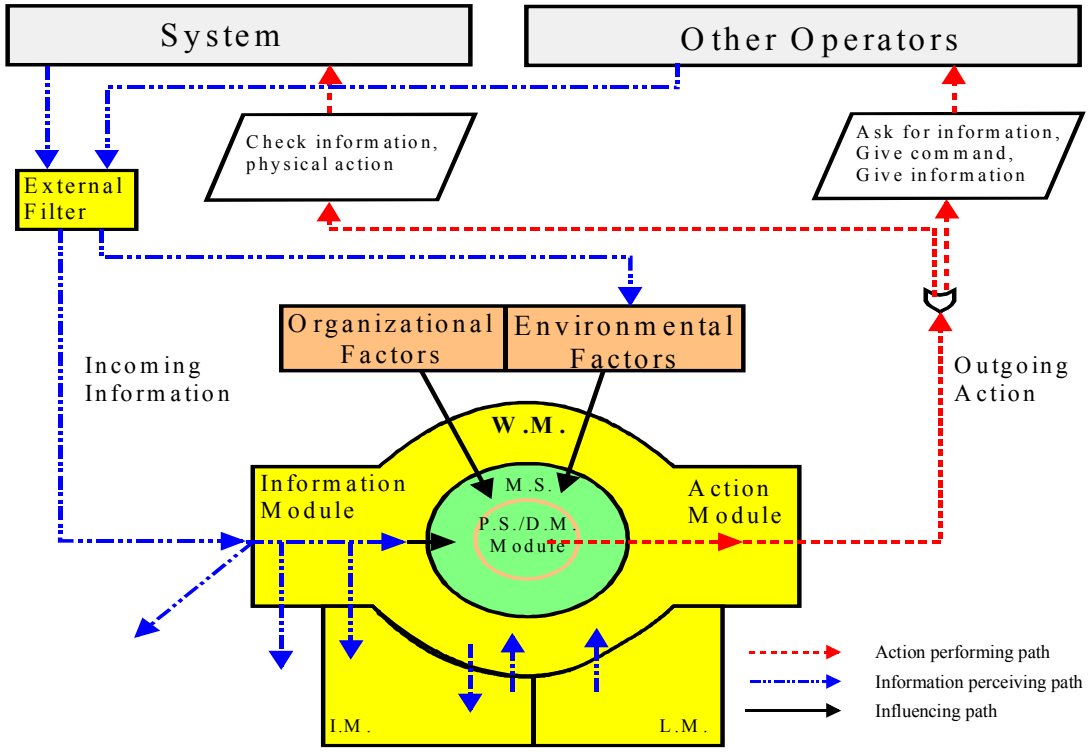

Figure 3 A High-Level View of Operator Dynamic Response Model

The stages of the IDA cognitive model are:

- *Information.* This stage focuses on the operator's perception of the environment and the cues presented to him/her. The information is presented externally. Cognitive processing of the

information is limited to the task of perceiving the information, but limited processing of the information is done at this stage.

- *Diagnosis/Decision.* This stage is internal to the operator. At this phase, the operator uses whatever information was perceived in the previous stage, along with stored memories, knowledge, and experience to develop a mental model of the situation. Following this situational assessment, the operator engages in decision-making strategies to plan the appropriate course of action.
- *Action.* In this final stage, the operator puts the chosen course of action into play.

Within each of these elements, a nested IDA structure may exist [6]; that is, each phase of the IDA model may be decomposed into further IDA structures as needed during task analysis and parsing of different human activities into 'sub-events" or sub-tasks. For instance, I-in-I explains the information being perceived and recognized, D-in-I involves deciding what to do with the perceived information (e.g., discard it or keep it), and A-in-I involves any actions stemming from this decision. For the application described within this paper, only the nested structure for the primary I phase was used.

Error is defined in terms of the operator failing to meet a plant need. The focus is on the safety impact of operator actions, some of which may be identical to HFEs defined in PRA (as top events in the event tree or as basic events in fault trees), and may be viewed as "unsafe acts" (in the ATHEANA method [7]) or as corresponding causes.

Using the information processing model, HFEs (errors defined based on mismatch between the operator's action and a plant need) can be traced through the I, D, A chain (Figure 4). An error could therefore be rooted in (1) action execution failure (A) given correct decision; (2) failure in situation assessment, problem solving and decision making, given correct information (D); or (3) failure in the information-gathering stage. In this view, failures of I, D, or A are "minimal cut-sets" of the human failure events.

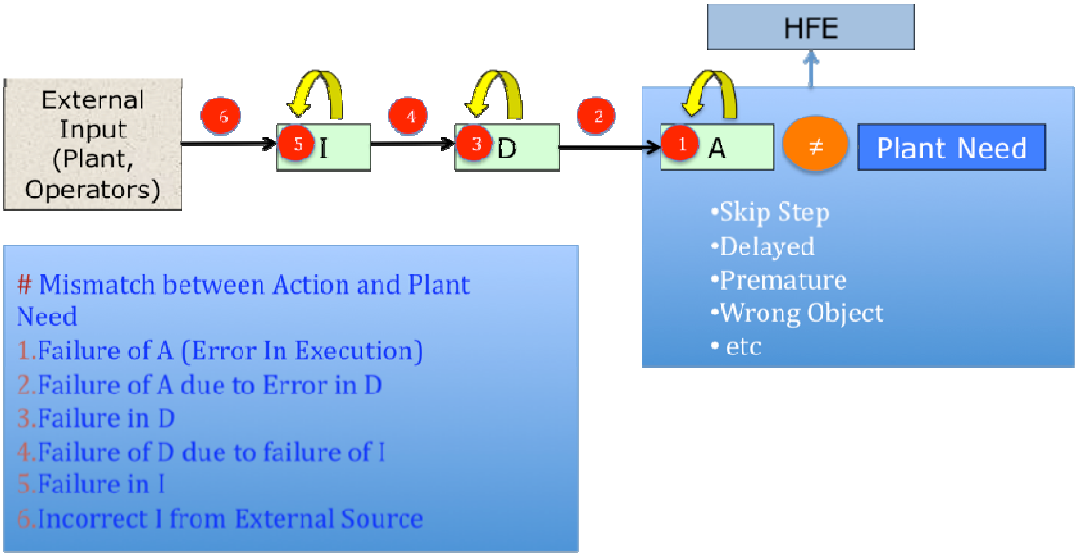

Figure 4 Human Error in IDA Framework

Since errors in the IDA framework are the result of failure in I, D, or A phases, specific models needed to identify corresponding human failure mechanisms. A companion paper [3] discusses the development/identification of the failure mechanisms, which represent the link between the

performance shaping factors (PSFs) and the possible human failures identified within the CRTs. Thus, the failure mechanisms represent a middle layer in the qualitative analysis approach, with the CRTs representing the top layer and the PSFs representing the bottom or lower layer. The mid-layer linkage is an important component, as it ensures that the correct PSFs (and scenario context) are identified for quantifying the probability of the possible human failures.

The failure mechanisms are built into fault trees to be used in the qualitative analysis under development [5]. These fault trees can be linked back to the HFE as described in the CRTs. A series of fault trees have been developed, representing each of the IDA phases. As an example Figures 5 shows the fault tree leading to the failure mechanisms for the phase.

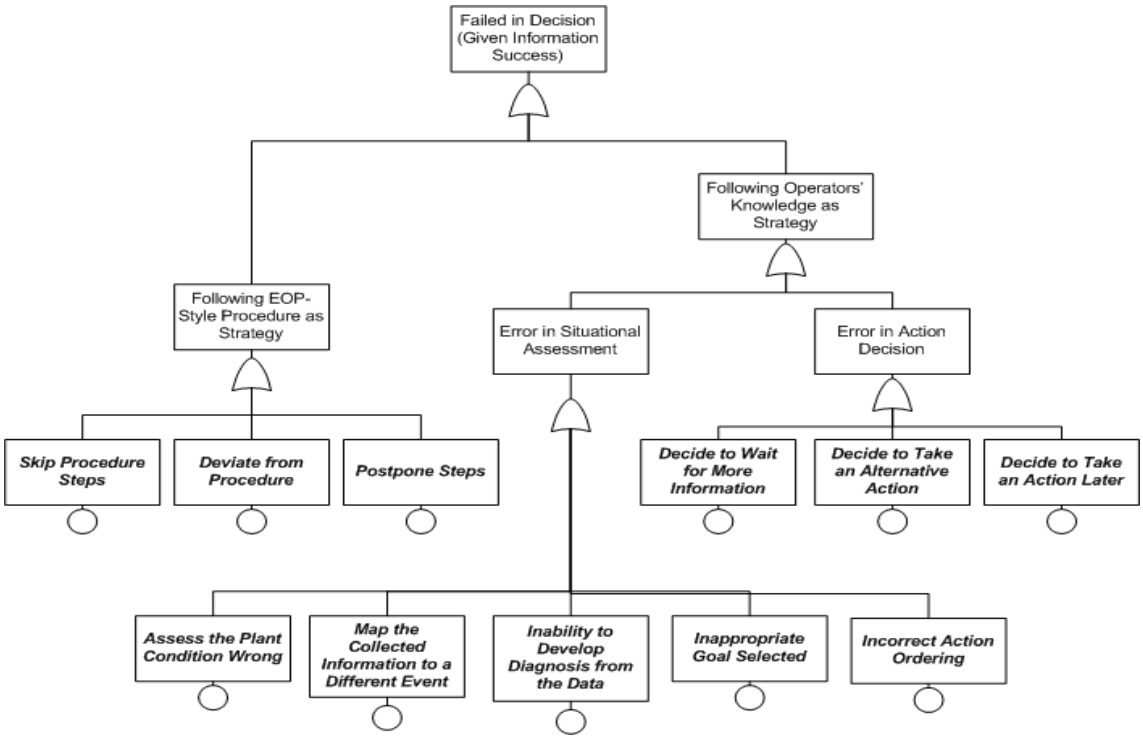

Figure 5. Fault Tree Identifying the Failure Mechanisms for the D Model Stage [4]

The failure mechanisms will be used to direct the HRA analyst to the appropriate list of PSFs relevant for the HFE in question. The failure mechanisms provide the means by which the PSFs are connected to the HFE. Where possible, this link between failure mechanism and PSFs will be informed by the psychological literature, and where not possible, by expert inference. The PSFs will ultimately be used in the quantification of the human error probability (HEP) for the HFE. The PSFs will be the final element within the failure mechanisms (mid-layer) fault trees and are currently represented by small circles underneath each failure mechanism.

PSFs proposed by Groth and Mosleh [8] are being considered for use in the proposed human performance model. This set was based on an analysis of all HRA methods, IDAC [6], and the Human Event Repository and Analysis system [9]. They are organized in hierarchical PSF structure that is grouped into six categories which are definitionally orthogonal; therefore, relationships between the PSFs can be examined without fear of contamination by overlapping PSFs. The six PSF groupings proposed are:

1. Machine: factors associated with the physical system as it was designed by the manufacturer, including plant hardware, software, human-system interface, and system responses. Machine factors are often static.
2. Situation: factors of the situation that are likely to affect human performance, which are often dynamic. Situation factors include objective task and time load, environmental conditions, and situation and task complexity. Situation factors are often perceived as stressors.
3. Stressors: factors that can act as stressors on the person(nel) involved in the situation, including perceived task and time load, perceived severity, perceived urgency, and perceived responsibility. Stressors interact with Personal factors to produce the *stress* that the involved personnel experience.
4. Person: factors internal to the individual, including attention, psychological and physical abilities, attitude, knowledge and experience, skills, and work conduct.
5. Team: factors associated with the crew or team, including communication, coordination, cohesion, supervision, and role awareness.
6. Organization: factors that are under control of the organization, including organizational programs, safety culture, management, resources (including tools and procedures), and staffing and scheduling.

These six groups are composed of 36 PSFs. Table 4 outlines how these 36 PSFs are distributed across the six PSF groups.

5. MODEL INTEGRATION AND ANALYSIS

CRT sequences and branches capture some but not all of the context factors and causes of operator responses. The branches of CRT are mainly at the functional level and do not typically cover the "human failure mechanisms" or their causes. This aspect is delegated to the supporting mid-layer models (fault trees) which are attached to the some of the branches of the CRT as appropriate (see Figure6). This is analogous to the modeling division between event tree fault trees and fault trees in PRA.

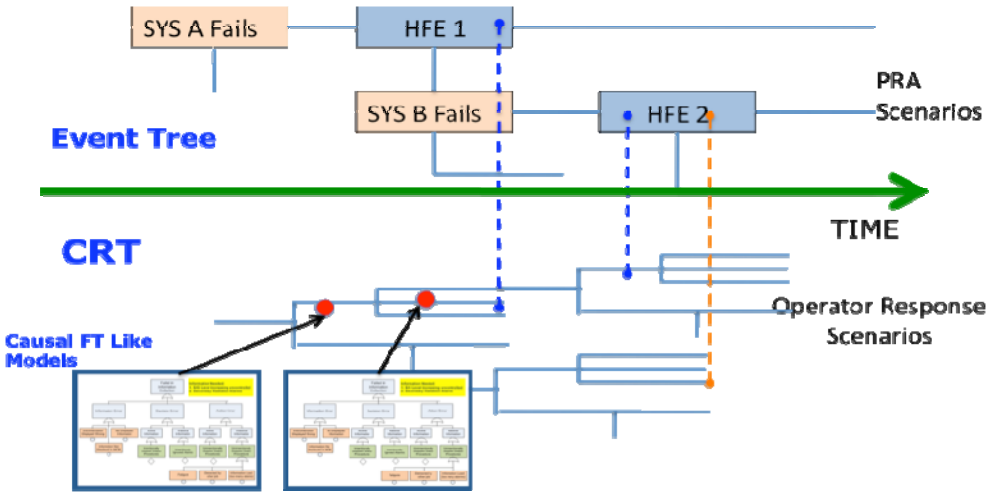

Figure 6 Adding Human Failure Mechanisms and Causal Factors to CRT Branch Points

This causal model of human failure modes and failure mechanisms is connected to a set of performance-shaping factors, as the last layer of causal modeling. The three layers of analysis, as depicted in Figure 7, are (1) Top Layer, the CRT model, (2) Mid-Layer, the human performance model, and (3) performance-influencing factors (i.e., context factors not captured by the first two layers). This layer essentially comprises the relevant PSFs connected to the elements of the mid-layer model.

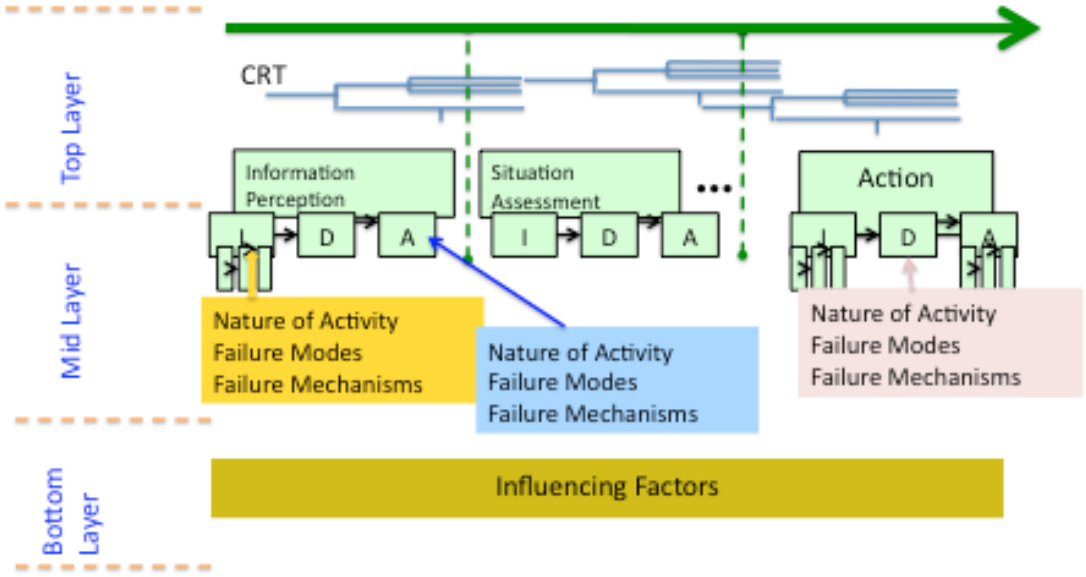

Figure 7 Three Layer Crew Response Modeling

A "linked trees" approach (similar to FT-ET linking) can help identify CRT scenario "causal cut-sets." This will result in a more explicit identification of sources of HFE dependencies. Qualification models can also use such "causal cut-sets" as their starting point in quantifying the HFE probabilities.

6. HUMAN ERROR PROBABILITY ASSESSMENT FRAMEWORK

Since an HFE is the result of one or several sequences of events or conditions (overall context) for a given plant PRA scenario according to the CRT and corresponding linked causal models, the HEP can be calculated as follows (see Figure 8):

$$p(HFE \mid S) = \sum_i \left[p(HFE \mid C_i) \times p(C_i \mid S) \right]$$

where

- S = PRA Scenario (essentially the initiator)
- Ci = Specific "context" i (as defined earlier in CRT) - Each CRT sequence leading to an HFE represents a unique Ci
- Different Cis often have common elements (e.g., common human actions, plant events, and PSFs) as represented by common segments of CRT scenarios

This is essentially the formulation adopted by ATHEANA [7].

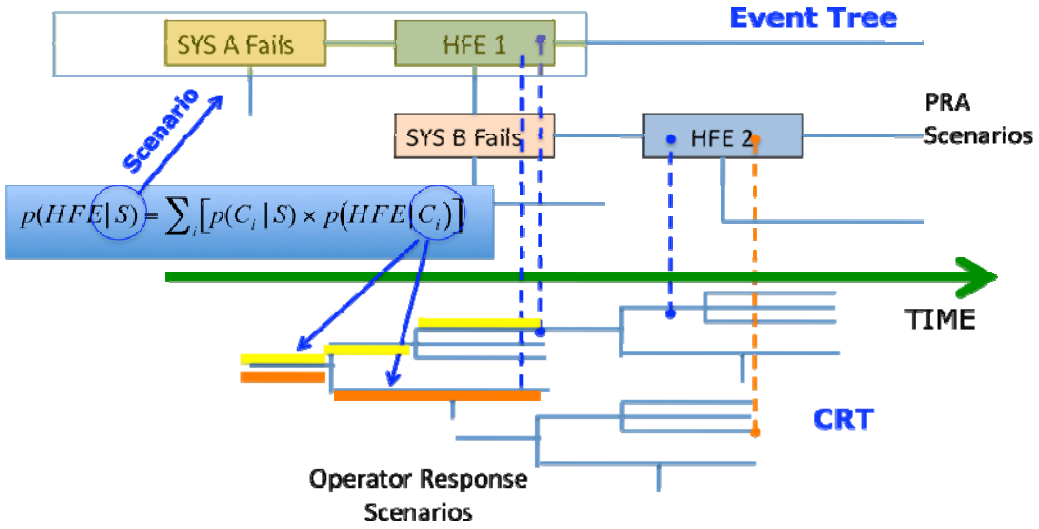

Figure 8 Scenario-Based HEP Quantification Concept

Under some modeling assumptions and abstractions, each Ci can be described via a set of context factors:

$$C_i \equiv \{ F_{i1}, F_{i2}, ..., F_{in} \}$$

where F_{ij} = context factor j for context i. Examples are "a specific crew," "elapsed time in scenario," "a specific PSF," or "a specific operator action." In this case HEP equation can be written as follows.

$$p(HFE \mid S) = \sum_{i=1}^{I} p(HFE \mid F_{i1}, F_{i2}, ..., F_{in}) \times p(F_{i1}, F_{i2}, ..., F_{in} \mid S)$$

We note that oftentimes $F_{ij}=F_j$ for several i (e.g., same level of time pressure or same crew).

Different formulations of the equation are necessary, depending on, for example, the level of HFE decomposition and whether we characterize context in terms of a set of factors, "proximate cause," "story," "scenario," etc.

The above equations are useful in providing a **conceptual** link between the qualitative and quantitative portions of HRA. It can shown that under certain set of assumption and simplifications these equations can lead to quantification formulas used by a number of the existing HRA methods

7. SUMMARY

This paper provides an overview of a new hybrid HRA approach being developed as part of response to a Staff Requirements Memorandum (SRM) to the Advisory Committee on Reactor Safeguards (ACRS) of US Nuclear Regulatory Commission (NRC). The approach introduces the "crew response tree" (CRT) concept, which provides a structure for capturing the "context" associated with the human failure events under analysis, and uses the Information Processing Model as a platform to identify potential failures. It incorporates behavioral science knowledge by providing the decompositions of human failures/failure mechanisms/failure factors built from a top-down and bottom-up approach, the latter reflecting those findings from scientific papers that document theories and data of interest. Ongoing work includes addressing technical gaps in models, development and testing of guidance for use of the qualitative analysis methodology, and possible short term and long terms solution to the issues related to error probability assessment.

Acknowledgements and Disclaimers

This work was funded by the U.S. Nuclear Regulatory Commission (USNRC) at Sandia National Laboratories (Sandia) and Idaho National Laboratory (INL). Sandia is a multi-program laboratory operated by Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000. INL is a multi-program laboratory operated by Battelle Energy Alliance, for the United States Department of Energy. The University of Maryland and Paul Scherrer Institute participated under subcontract from Sandia. Vattenfall Ringhals AB participated independently through a grant from the Nordic Nuclear Safety (NKS) Research Council. This work of authorship 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. This paper was prepared in part by employees of the USNRC. It presents information that does not currently represent an agreed-upon staff position. NRC has neither approved nor disapproved its technical content.

References

[1] US Nuclear Regulatory Commission, *Staff Requirements—Meeting with Advisory Committee on Reactor Safeguards, SRM M061020*, US Nuclear Regulatory Commission, November 8, 2006, Washington, DC.

[2] E. Lois et al, "International HRA Empirical Study –Phase 1 Report", NUREG/IA-0216, November 2009

[3] S. Hendrickson, A.M. Whaley, R.L. Boring, J.Y.H. Chang, S.H. Shen, A. Mosleh, J.H. Oxstrand, J.A. Forester, D.L. Kelly, E.L. Lois, "*A Mid-Layer Model for Human Reliability Analysis: Understanding the Cognitive Causes of Human Failure Events*", Proc. of the 10th Int. Conf. on Probabilistic Safety Assessment and Management (PSAM10) (this conference).

[4] S.H. Shen, A. Mosleh, D. L. Kelly, R. L. Boring "*Example Application of Model-Based HRA Approach"*, Proc. of the 10th Int. Conf. on Probabilistic Safety Assessment and Management (PSAM10) (this conference).

[5] V. N. Dang, J. A. Forester, A. Mosleh, *"Developing a New HRA Quantification Approach from Best Methods and Practices"*, Proc. of the 10th Int. Conf. on Probabilistic Safety Assessment and Management (PSAM10) (this conference).

[6] Chang, Y. H. J. & Mosleh, A. (2007). Cognitive modelling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 1: Overview of the IDAC model. *Reliability Engineering and System Safety, 92*, 997-1013.

[7] U.S. NRC. "*ATHEANA User's Guide - Final Report*", NUREG-1880, U.S. Nuclear Regulatory Commission, Washington DC, USA, 2007

[8] Groth, K. M. & Mosleh, A. (2009, September). *A data-informed model of performance shaping factors and their interdependencies for use in human reliability analysis*. Paper presented at the ESREL Annual Conference, Prague, Czech Republic.

[9] Hallbert, B., Whaley, A., Boring, R., McCabe, P., & Chang, Y. (2007, November). Human Event Repository and Analysis (HERA): The HERA Coding Manual and Quality Assurance. *NUREG/CR-6093, Vol. 2, INL/EXT-07-12387.* Washington, DC: U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research.