On the Evaluation of Human Error Probabilities for Post-Initiating Events

by

Mary R. Presley

SUBMITTED TO THE DEPARTMENT OF NUCLEAR SCIENCE AND ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREES OF

MASTER OF SCIENCE IN NUCLEAR SCIENCE AND ENGINEERING AND BACHELOR OF SCIENCE IN NUCLEAR SCIENCE AND ENGINEERING AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2006
[September 2006]

© 2006 Massachusetts Institute of Technology. All Rights Reserved.

Signature of Author _____
Mary R. Presley
Department of Nuclear Science and Engineering
August 29, 2006

Certified by _____
George E. Apostolakis
Professor of Nuclear Science and Engineering
Professor of Engineering Systems
Thesis Supervisor

Accepted by _____
Michael W. Golay
Professor of Nuclear Science and Engineering
Thesis Reader

Accepted by _____
Jeffrey A. Coderre
Associate Professor of Nuclear Science and Engineering
Chairman, Departmental Committee on Graduate Students

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
OCT 12 2007
LIBRARIES

ARCHIVES

On the Evaluation of Human Error Probabilities for Post-Initiating Events

by

Mary R. Presley

Submitted to the Department of Nuclear Science and Engineering
on August 30, 2006, in partial fulfillment of the requirements for the
Degrees of Master of Science and Bachelor of Science in
Nuclear Science and Engineering

ABSTRACT

Quantification of human error probabilities (HEPs) for the purpose of human reliability assessment (HRA) is very complex. Because of this complexity, the state of the art includes a variety of HRA models, each with its own objectives, scope and quantification method. In addition to varying methods of quantification, each model is replete with its own terminology and categorizations, therefore making comparison across models exceedingly difficult. This paper demonstrates the capabilities and limitations of two prominent HRA models: the Electric Power Research Institute (EPRI) HRA Calculator (using the HRC/ORE and Cause Based Decision Tree methods), used widely in industry, and A Technique for Human Error Analysis (ATHEANA), developed by the US Nuclear Regulatory Commission. This demonstration includes a brief description of the two models, a comparison of what they incorporate in HEP quantification, a "translation" of terminologies, and examples of their capabilities via the Halden Task Complexity experiments. Possible ways to incorporate learning from simulator experiments, such as those at Halden, to improve the quantification methods are also addressed.

The primary difference between ATHEANA and the EPRI HRA Calculator is in their objectives. EPRI's objective is to provide a method that is not overly resource intensive and can be used by a PRA analyst without significant HRA experience. Consequently, EPRI quantifies HEPs using time reliability curves (TRCs) and cause based decision trees (CBDT). ATHEANA attempts to find contexts where operators are likely to fail without recovery and quantify the associated HEP. This includes finding how operators can further degrade the plant condition while still believing their actions are correct. ATHEANA quantifies HEPs through an expert judgment elicitation process.

ATHEANA and the EPRI Calculator are very similar in the contexts they consider in HEP calculation: both factor in the accident sequence context, performance shaping factors (PSFs), and cognitive factors into HEP calculation. However, stemming from the difference in objectives, there is a difference in how deeply into a human action each model probes. ATHEANA employs a HRA team (including a HRA expert, operations personnel and a thermo-hydraulics expert) to examine a broad set of PSFs and contexts. It also expands the accident sequences to include the consequences of a misdiagnosis beyond simple failures in implementing the procedures (what will the operator likely do next given a specific misdiagnosis?) To limit the resource burden, the

EPRI Calculator is prescriptive and limits the PSFs and cognitive factors for consideration thus enhancing consistency among analysts and reducing needed resources. However, CBDT and ATHEANA have the same approach to evaluating the cognitive context.

The Halden Task Complexity experiments looked at different factors that would increase the probability of human failures such as the effects of time pressure/information load and masked events. EPRI and ATHEANA could use the design of the Halden experiments as a model for future simulations because they produced results that showed important differences in crew performance under certain conditions. Both models can also use the Halden experiments and results to sensitize the experts and analysts to the real effects of an error forcing context.

Acknowledgements

I would like to thank my advisor, Professor George Apostolakis, for his patience, wisdom, and guidance in getting me through the thesis process. I would like to also thank Professor Michael Golay for being my reader.

I would like to extend a special thank you to Larry Rau from Seabrook Station and Jeffery Julius from Scientech. They were both an invaluable guide and resource to this project. They tirelessly answered my questions, provided extra material to consider, and generally supported this project every step of the way. Thank you.

I would like to thank Erasmia Lois from the US Nuclear Regulatory Commission for suggesting the topic.

Finally, I would like to thank my family for their endless love, patience and support.

Table of Contents

Table of Figures

Table of Tables

Acronyms

AFW	Auxiliary Feedwater
AOP	Anticipated Operating Procedure
ASEP	Accident Sequence Evaluation Program
ATHEANA	A Technique for Human Event Analysis
ATWS	Anticipated Transient Without Scram
BWR	Boiling Water Reactor
CBDT	Cause-Based Decision Tree
CCDF	Complimentary Cumulative Distribution Function
CCP	Centrifugal Charging Pump
CREAM	Cognitive Reliability and Error Analysis Method
DETAM	Dynamic Event Tree Analysis Method
EBAOP	Event-Base AOP
EFC	Error Forcing Context
EOC	Error of Commission
EOO	Error of Omission
EOP	Emergency Operating Procedure
EPRI	Electric Power Research Institute
FLIM	Failure Likelihood Index Methodology
HCR	Human Cognitive Reliability
HEP	Human Error Probability
HERA	Human Event Repository and Analysis
HFE	Human Failure Event
HI	Human Interaction
HORAAM	Human and Organizational Reliability Aspects in Accident Management
HRA	Human Reliability Analysis
IE	Initiating Event
IRTU	Intelligent Remote Terminal Unit
MAUD	Multi-Attribute Utility Decomposition
MERMOS	Méthode d'Evaluation de la Rélisation des Missions Opérateur pour la Sûreté (French)
MMI	Man-Machine Interface
NPP	Nuclear Power Plant
NRC	Nuclear Regulatory Commission
OPERA	Operator Performance and Reliability Analysis
ORE	Operator Reliability Experiments
PRA	Probabilistic Risk Assessment

PSF	Performance Shaping Factors
PWR	Pressurized Water Reactor
RHR	Residual Heat Removal
RIDM	Risk-Informed Decision Making
RWST	Reactor Water Storage Tank
SBEOP	Symptom-Based EOP
SD	Significance Determination
SHARP	Systematic Human Action Reliability Procedure (EPRI HRA framework)
SLIM	Success Likelihood Index Methodology
SP	Suppression Pool
SPAR-H	Standardized Plant Analysis Risk Human Reliability Analysis
THERP	Technique for Human Error Rate Prediction
TRC	Time Reliability Correlation
UA	Unsafe Action
VCT	Volume Control Tank

1 Introduction

The quantification of human error probabilities (HEPs) is a challenging endeavor because they are sensitive to a wide range of contexts that go beyond hardware states to include psychological factors, crew training, cognition, event timing, and much more. Unlike hardware, humans can intentionally or unintentionally "fail" in ways that can potentially degrade the plant state beyond a simple failure. Reflecting this complexity, there are many HRA methods/frameworks in existence that attempt to analyze and quantify human error, including, but not limited to: THERP,[1] ASEP,[2] HEART,[3] OPERA,[4] DETAM,[5] FLIM,[6] SLIM/MAUD,[7] SPAR-H,[8] HERA,[9] HORAAM,[10] HRMS,[11] CREAM,[12] MERMOS,[13] ATHEANA,[14] and the EPRI HRA Calculator[15] (which includes SHARP1,[16-17] CBDT[15] and HCR/ORE[15] among other things).

In trying to capture the various aspects of human failure, quantification methods have ranged from look-up tables (THERP) to time reliability curves using statistical data (HCR/ORE) to pure expert opinion elicitation (ATHEANA). There is no widely accepted method for quantification of HEPs in the US nuclear industry. In addition to varying methods of quantification, each method is selective in the type of human failure events (HFEs) it models and is replete with its own terminology and categorizations, making it difficult to directly compare methods or switch back and forth.

Prompted by this patchwork state of HRA, a workshop meeting on HRA methods was recently held in New Orleans (at the PSAM 8 conference) to identify concrete

measures necessary to advance the field. Several action items came from this workshop, two of which are addressed here: 1) to demonstrate what HRA "can do" by determining how various methods interpret and quantify human failure; and 2) to determine the specific type of (simulator) data needed to support these HRA methods.[18]

It is the intent of this paper to "explore the capabilities and limitations"[18] of various quantification methods for post-initiating event (post-IE) HEPs. This scope is limited to the HCR/ORE and CBDT methods, heavily used in industry via the EPRI HRA Calculator, and ATHEANA, developed by the NRC. We will also incorporate learning from the Halden[19] simulator experiments to strengthen our understanding of each method and their limitations. Finally, we shall examine these methods with respect to the following requirements for an advanced HRA method: yields transparent results, not overly resource intensive, incorporates effects of organizational factors, gives special attention to cognitive errors, provides a well-defined quantification procedure and incorporates data and feedback into the model.[20]

In order to accomplish this, Sections 2 and 3 provide a brief overview of ATHEANA and the EPRI HRA Calculator, respectively, including definitions of method-specific terminology. Section 4 begins the comparison and capabilities determination process by describing exactly how each method interprets human failure and incorporates: available time, response time, performance shaping factors (PSFs), cues, cognition and misdiagnosis. Using data from the Halden simulator experiments, Section 5 illustrates the quantification process using the various methods. Section 6 provides a

discussion on simulator data incorporation, advanced HRA methods, and the capabilities and limitations of each method. Finally, Section 6 concludes with a discussion on data incorporation into ATHEANA and the EPRI HRA Calculator.

2 ATHEANA – A Brief Overview and Important Terminology

ATHEANA attempts to identify significant human failure events (HFEs) and quantify both intentional and unintentional errors of commission (EOCs) and errors of omission (EOOs), as well as their resulting consequences. For each human failure event – or functional failure caused by a human – this method attempts to identify important contexts, called error forcing contexts (EFCs), that may lead operators to carry out an inappropriate action. ATHEANA's quantification process is geared towards finding the probability of failure within these contexts. While quantification is the primary focus of this paper, the search process of ATHEANA is presented here as the context for the consensus-based expert opinion quantification method. It is this link between search and quantification that makes ATHEANA unique.

Figure 1 provides a graphical summary of the ATHEANA process; these eight steps will be reviewed briefly here.

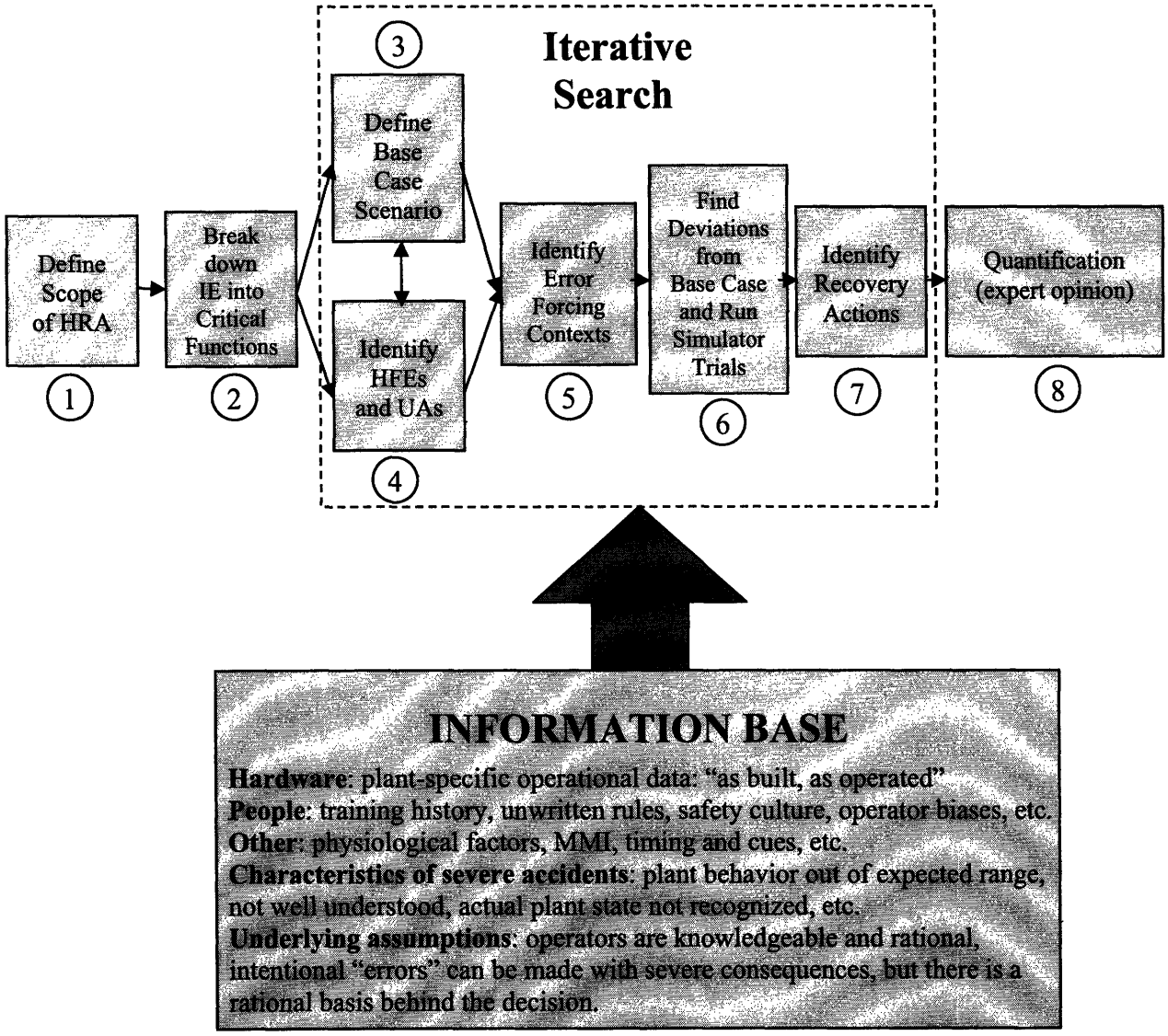


Figure 1: Summary Flow Chart of the ATHEANA Process

2.1 Steps 1-4: Identifying Human Failure Events

The first portion of the ATHEANA framework is geared towards finding modes of human failure that are significant enough to include in the plant's PRA. The first step requires the HRA team (a multi-disciplinary team comprised of experts such as HRA analysts, PRA analysts, operators, trainers, and thermo-hydraulics specialists) to carefully

define the issue to be analyzed and determine the level of detail required to do an "adequate" job. The HRA team must decide what must be modeled: everything, only specific functional failures, only certain initiating events? Quantitative screening can be helpful in determining the necessary scope; also, if the HRA team does not have specific issues in mind at the outset, NUREG-1624[14] suggests a method for identifying candidate HFEs.

Once the important initiators are selected, and their corresponding event trees identified, the HRA team proceeds to prioritize the plant functions, systems, and equipment required to respond to the accident initiators. With these critical items identified, candidate HFEs can be more readily identified by examining the initiating event (IE) event tree and then systematically evaluating the event tree branch points for possible human-caused functional failures. An example of this search process will be presented below.

With the scope and issue clearly defined, the HRA team moves on to describe the base-case scenario, or the expected evolution of the accident scenario. Part of defining the base case includes describing and understanding the human performance context of the scenario. The following is a list of suggested components of the base case description, taken from the ATHEANA user's guide:[21]

- A list of possible causes of the initiating event(s)
- A brief, general description of the expected sequence of events (as in PRA event trees), starting before reactor trip
- A description of the assumed initial conditions of the plant

- A familiarization/description of the expected plant conditions for the accident sequence
- A specification of the *expected sequence timing* of plant status changes
- A description of the *expected trajectories, over time, of key parameters* indicating plant status and a specification of the status of indications and other *cues* that are expected as the sequence evolves
- Any assumptions of expected plant behavior, system/equipment/indicator responses, and operator response
- A discussion of the procedures expected to be used for the given situation
- A description of *key operator actions, and their timing*, expected during the scenario progression

Concurrently with the base case definition, the HRA team identifies and defines the possible human failure events and the corresponding unsafe actions (UAs, where one or more unsafe action makes up a HFE). In order to identify the HFEs, a systematic process, building on the IE event trees from step 1, is followed by determining:[21]

1. whether the function is necessary or undesired
2. the system(s) or equipment that perform the function
3. the pre-initiator status of the system(s) or equipment
4. the functional success criteria for the system(s) or equipment
5. the functional failure modes of the system(s) or equipment
6. how the operator interacts with the equipment and deciding if EOCs, EOOs, or both are relevant.

Both the ATHEANA user's guide[21] and NUREG-1642[14] provide tables of example UAs for generalized equipment functional failure modes to help the HRA team through this step.

2.2 Steps 5-7: The Error Forcing Context

Recognizing that many severe accidents share common attributes, such as plant behavior being out of the expected range, plant behavior not being well understood by operators, plant procedures not helpful/appropriate and actual plant state not being recognized by operators,[21] ATHEANA attempts to define EFCs that exhibit these attributes. These error forcing contexts are comprised of a combination of plant state (hardware) and other performance shaping factors (PSFs). Again, the goal of identifying these EFCs is to find regimes where operators believe an inappropriate action is the "right" thing to do.

In order to identify the PSFs, plant-specific background data must be considered, such as: formal procedures, crew characteristics/dynamics, ergonomics, informal rules and biases. An example of an informal rule is "beat the automatic system when practical," as opposed to "wait for the automatic system before taking action."[14] The former rule might carry with it a greater chance for an EOC to occur. Plant biases include frequency and recency biases that arise from plant operation and simulator training (operators may have an expectation of how a scenario will unfold).[21] The goal of this search is to find scenarios that might prove "troublesome" to operators and produce an error-forcing context. Simulator exercises may prove helpful in this step in that they allow the HRA analyst to observe how the operators behave and think. They can also serve to test theories of operator response.[21] With the base case defined, and having an idea of what the important PSFs might be, the HRA team searches for and defines

potential deviations from the base case. These are credible scenarios that include the identified EFC. Section 2.4 will provide an example of a deviation scenario and its accompanying EFC.

In addition to the identified EFCs, the analysts must take into account other complicating factors (what ATHEANA calls "aleatory" PSFs), as well as recovery factors, as part of the context of quantification. These "aleatory" PSFs are random factors that could alter the HEP, such as: time of day, "aggressiveness" of the crew, crew dynamics, other hardware failures, indicator failures, etc. These PSFs will be used to assess the aleatory uncertainty of the nominal case – by having experts judge the effect of the best and worse case scenarios with respect to these parameters.

To prevent an unrealistically conservative estimate, recovery factors are included in the analysis. In search for potential recovery actions and evaluation of their feasibility, there are five steps outlined in the ATHEANA user's guide:[21]

1. Define possible recovery actions given a HFE/UA has occurred
2. Consider *time available* for diagnosis and execution of potential recovery actions
3. Identify recovery cues
    a. Timing of recovery cues
    b. How compelling these cues are
        i. Do they strongly alert operators to a need for recovery?
        ii. Is there sufficient information to identify the most applicable recovery action?
4. Identify additional resources for aid in recovery (i.e., more staff) and associated timing

5. Assess the strength of recovery *cues and timing* with respect to EFCs – is there a "high" or "low" likelihood of successful recovery? Deviation scenarios with high likelihood of recovery need no further analysis or quantification. In this step, there are some suggested factors to consider when assessing feasibility of recovery:[21]
    a. Dependencies between the initial error and recovery actions that would make recovery unlikely
    b. Initial mindset (or diagnosis) of the situation may be hard to break
    c. Distractions or attention to other activities could cause new cues to be overlooked
    d. Operators may delay recovery action because there is a negative consequence to taking the action; this is especially relevant when plant hardware providing an alternative recovery is "almost" repaired.

NUREG-1624[14] gives further guidance on assessing recovery actions by type of failure: "thinking" (mistakes and circumventions) or "doing" (slips and lapses). For thinking failures, this involves assessing how the operator could persist in believing their UA is the "correct" action, using the same process and information from step 5. "Doing" failures, as suspected, are more straightforward. First the team must decide whether the slip/lapse is recoverable at all: was plant hardware irreparably damaged? Was it so damaged that the time for recovery is greater than the time available? If recoverable, then the team must determine whether the slip/lapse can induce a mistake. If so, then it should be further analyzed, if not, then it can be dropped from further analysis.

2.3 Step 8: Quantification

The search process for EFCs ends when the team feels assured that the EFC is sufficiently well defined and that both the frequency of the context and the conditional probability of the UA in that context can be estimated with an appropriate degree of confidence. To test the adequacy of EFCs, the HRA team can do simulator tests, compare EFCs with past operational experience and with human performance checklists (see NUREG/CR-6208[22]).

The rigorous way to quantify the probability of a human failure event (P(HFE)) is:

$$P(HFE|S) = \sum_{all\_i,j} P(EFC_i,S) * P(UA_j|EFC_i,S) * P(\overline{R}|EFC_i,UA_j,S) \qquad [1]$$

where S refers to the PRA accident scenario, and $P(\overline{R}|EFC_i,UA_j,S)$ is the probability of non-recovery given an unsafe action has occurred in an error forcing context for that scenario. The probabilities are then summed over all UA/EFC combinations.

This rigorous method, however, requires too much resolution to be a feasible method of quantification, and so the HRA team may choose the following method:[21]

$$P(HFE|S) = \sum_{all\_i,j} P(EFC_i|S) * P(UA_j|EFC_i,S) \qquad [2]$$

In this case, recovery is factored implicitly into $P(UA_j|EFC_i,S)$, as explained below. For this formulation, the HRA team must use an expert opinion elicitation approach to quantification.[23]

An error forcing context is comprised of two parts: plant state (hardware) and performance shaping factors. For quantification purposes, $P(EFC_i|S)$ is taken directly from the PRA,[23] and represents the plant state portion of the EFC. $P(PSF_i|S)$ is not included explicitly in the formulation because these are characteristics that cannot be explicitly quantified. So, these PSFs are implicitly taken into account in quantifying $P(UA_j|EFC_i,S)$ through the expert opinion elicitation process.

The term "error *forcing* context" can be misleading because it implies that the conditional probability $P(UA_j|EFC_i,S)$ should be near unity. This, however, is not the case. EFCs are contexts which increase the likelihood of error, and in some cases "trigger" error, but do not generally *force* an error.

The probability of an UA for a specific EFC is taken from the consensus expert opinion elicitation process described below. Due to the way the expert opinion elicitation process is structured, aleatory uncertainty, recovery, and dependencies are all holistically incorporated into $P(UA_j|EFC_i,S)$. This judgment-based quantification consists of six steps:[23]

1. Discuss HFE and influences, identify specific EFC and "aleatory" PSFs

2. "Calibrate" experts
3. Elicit an estimated curve
4. Each expert presents his estimated curve to the group of experts
5. Open discussion amongst experts
6. Arrive at consensus curve

1. So that the experts understand exactly what they are quantifying, they should discuss and fully understand the scenario, including: the definition of the HFE in question, the plant state (part of the EFC), and the relevant PSFs (the other part of the EFC). To prevent the experts from being overwhelmed, only the most relevant PSFs to the scenario at hand should be taken into account. The importance of scenario specificity cannot be over emphasized. Most methods that incorporate PSFs do so by generically applying an adjustment factor (i.e., increase the failure probability by a factor of 2 if there is time pressure). ATHEANA recognizes that in some situations these factors may not have a large impact; for example, time pressure may not impact operators in a very familiar situation on which they get trained frequently.

This is how ATHEANA incorporates recovery actions and the PSF portion of the EFC. Aspects such as scenario timing and relevant cues are woven into the description of the HFE and become the context within which the experts judge the probability of a UA.

2. The next step is to calibrate the experts so they have a more intuitive understanding of what a probability really is (i.e., 1 failure in 10 trials is 0.1). Here, they are encouraged to think about failures as a number of x failures in n trials instead of directly estimating a probability.

3. Now, each expert should come up with a 7-point estimation of $P(UA_j|EFC_i,S)$, including the 1st, 10th, 25th, 50th (median), 75th, 90th, and 99th percentiles. The experts should start by setting the 1st and 99th percentiles to be the probabilities for the best and worst case scenarios, respectively. This is where the "aleatory" PSFs come in – the best scenario is when there are no adverse PSFs, and the worst is when all the PSFs are in play. For example, "aggressiveness" of crew and crew strategy have been found to be a source of variation in the time-to-completion for different tasks.[19] In exercising his/her judgment, an expert would then think about an effective crew and imagine them in the best circumstances for the 1st percentile and imagine a particularly ineffective crew with communication difficulties and imagine them in the worst circumstance for the 99th percentile estimate.

There have been extensive studies demonstrating that expert opinion is replete with potential biases, and that experts have difficulty consistently assessing the extremes of a range. Reference 23 describes some of these biases, and suggests methods to alleviate these effects.

4-6. After the experts have their curves, they present them to the group and the group deliberates until a consensus is reached. Part of the reason behind a consensus-based approach is to avoid unintentional bias. The epistemic uncertainty for this method would be a family of curves, one for each expert's opinion – see Table 1.

Table 1: Example Expert Opinion Elicitation Results for Failure to Isolate a Stuck-Open Atmospheric Dump Valve within 30 Minutes of the Initiating Event[24]

Expert	Percentiles						
	1st	10th	25th	50th	75th	90th	99th
#1	0.01	0.03	0.05	0.08	0.4	0.8	1.0
#2	0.001	0.003	0.008	0.02	0.07	0.1	0.8
#3	0.001	0.01	0.03	0.06	0.4	0.6	0.9
#4	0.005	0.01	0.02	0.033	0.1	0.6	0.8
Consensus	**0.004**	**0.01**	**0.03**	**0.05**	**0.2**	**0.5**	**0.9**

2.4 Example of the Search Process

The following example is taken from a PWR plant.[14] The search process is presented as a series of steps that the HRA team took for this limited trial of ATHEANA. Each step was accompanied by a set of guiding questions to aid the team in identifying important scenarios. For brevity, these questions are not presented here – most of them can be found in the Sections above, and the rest can be found in NUREG-1624.[14] The HRA initially chose three initiating events to quantify analyze: MLOCA, LOSP and ATWS. Only the MLOCA IE will be examined here. Figure 2 is the simplified event tree for the MLOCA, and Table 2 is a list of top events and their descriptions.

1. First the HRA team selected the initiating event of interest (MLOCA) and prioritized the functional requirements as represented by the nodes of the event tree (Figure 2). These functions/priorities were: - Makeup: Medium Priority

- Heat Removal: Medium Priority

- Long-Term Heat Removal: High Priority

2. The team then examined the safety functions required, defined their success criteria, and identified their failure modes. For each failure mode, the team asked "how can the operators produce the effects characterized by the failure modes identified?" From this process, they found two unsafe actions of particular significance, as described in NUREG-1624:[14]

- Operators stop pump (function: makeup; system: high-pressure injection)
- Operators operate pump outside design parameters (function: long-term cooling; system: residual heat removal (RHR) system)

3. Addressing these unsafe acts, the team searched for and defined important error forcing contexts. Here the team identified a credible accident sequence for each EFC to simulate in order to test the strength of the EFC. These sequences are:[14]

Both sequences involve a MLOCA where system repressurization is not possible, and continuous high-head injection is required to keep the core cool and covered. Furthermore, in these sequences it is unclear whether the steam generators act as a heat source or provide a heat sink because the break is the primary method of heat removal, and the primary system pressure is less than the pressure of the secondary system.

HFE #1 – Inappropriate Termination of Makeup

The error forcing context of this scenario is a deceptive failure of the RCS pressure indicators. In this simulation, one RCS pressure indicator was under repair and the second failed stuck during operation at 550 psig. This was intended to make the operators believe that the indicator is functioning normally, when in fact it indicated greater sub-cooling than reality. This misinformation would prompt the operators to shut off the pumps early (as directed by the procedure). In this case, core damage would ensue if recovery of injection was not restored in a timely manner.

HFE #2 – Inappropriate Depletion of Resources

In this scenario, there is increased RWST depletion due to containment spray system activation during the LOCA. If the RWST "empty" alarm sounds, high-head pumps should be stopped until reconfiguration is complete, or pump cavitation will occur leading to core damage unless the pressure can be reduced and low-pressure injection is initiated. To "trigger" the operator error of 'failing to stop the high-head pump,' the RWST alarm was made inoperable due to IRTU maintenance.

The conclusion of this limited test search was the identification of at least one strong EFC. The failed RCS pressure indicators indeed caused the simulation crew to prematurely stop the pumps even though the potential of a failed indicator was recognized (but not verbalized) by one operator. However, the inoperable RWST

"empty" alarm did not prove to be a significant EFC at all – the crew paid sufficient attention to the RWST level throughout the simulation, and the Work Control Supervisor recognized that the IRTU maintenance could fail the alarm. It was also brought to the attention of the trainers that in unfamiliar or tricky situations, operators might not adhere to the strong tendencies developed through training like the "think it, say it" rule. Neither scenario was considered significant enough to retain in the demonstration plant's HRA.[25]

Table 2: MLOCA Event Tree Top-Event Summary[26]

Top Event ID	Title	Description
RW	RWST	RWST failure - no inventory for RCS makeup.
ALT	Alternate Cooling	Alternate cooling to the charging pumps.
CSA	Charging Pump	Centrifugal Charging Pump Train A failure.
CSB	Charging Pump	Centrifugal Charging Pump Train B failure.
SIA	SI Pump	Safety Injection Pump Train A failure.
SIB	SI Pump	Safety Injection Pump Train B failure.
EF	EFW	EFW failure - motor and turbine-driven pumps.
OD	Operator - Depressurizes RCS	Operator failure to depressurize RCS for RHR injection, given failure of HPI.
RA	RWST Valve	RWST isolation valve Train A failure to remain open - RHR and CBS Train A suction path.
RB	RWST Valve	RWST isolation valve Train B failure to remain open - RHR and CBS Train B suction path.
L1	RHR Miniflow	RHR Train A failure in miniflow recirculation.
L2	RHR Miniflow	RHR Train B failure in miniflow recirculation.

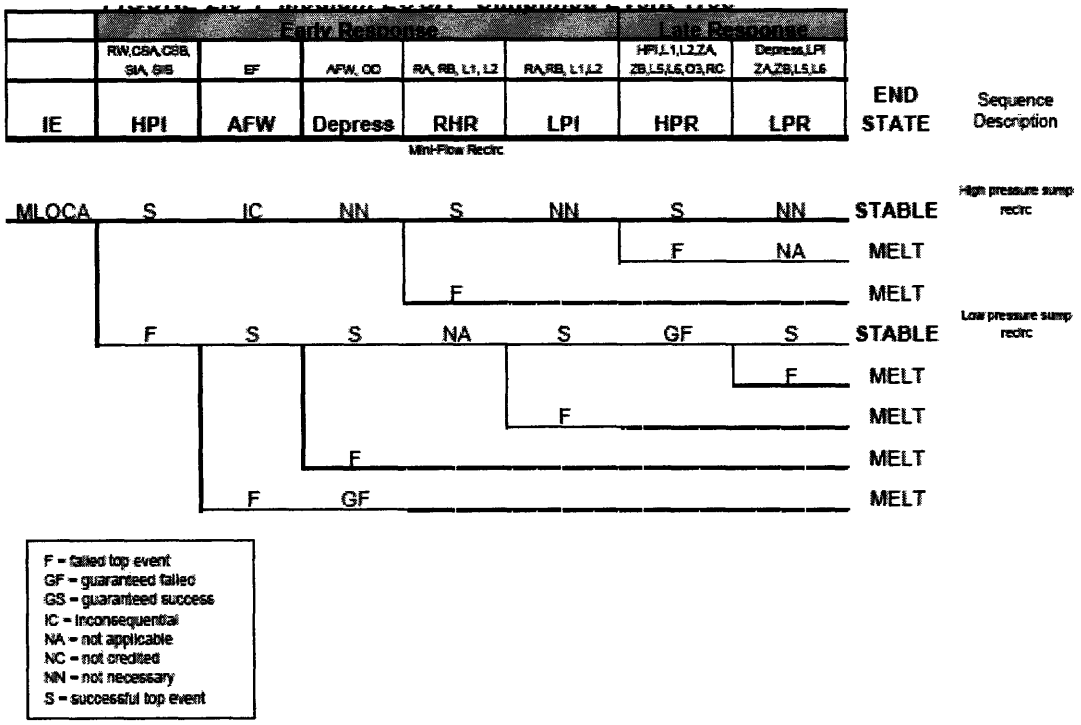


Figure 2: Simplified MLOCA Event Tree[26]

3 EPRI HRA Calculator – A Brief Overview and Important Terminology

The EPRI HRA Calculator combines the SHARP1 framework with five quantification methods (HCR/ORE, CBDT, THERP execution analysis, THERP annunciator response and SPAR-H) to create an HRA methodology that is easy to use, consistent, transparent, and non-resource intensive. Unlike ATHEANA, the EPRI Calculator does not attempt to break down human failure events into specific unsafe actions, or even into specific contexts. Rather, it takes a more generic approach, using time reliability correlations (TRCs) and generic decision trees. Recognizing that many plants do not have an HRA expert on their staff, this method is specifically designed to be

usable by a general PRA practitioner with some training. Figure 3 below provides a graphical summary of the EPRI HRA Calculator; these six steps will be reviewed briefly here.

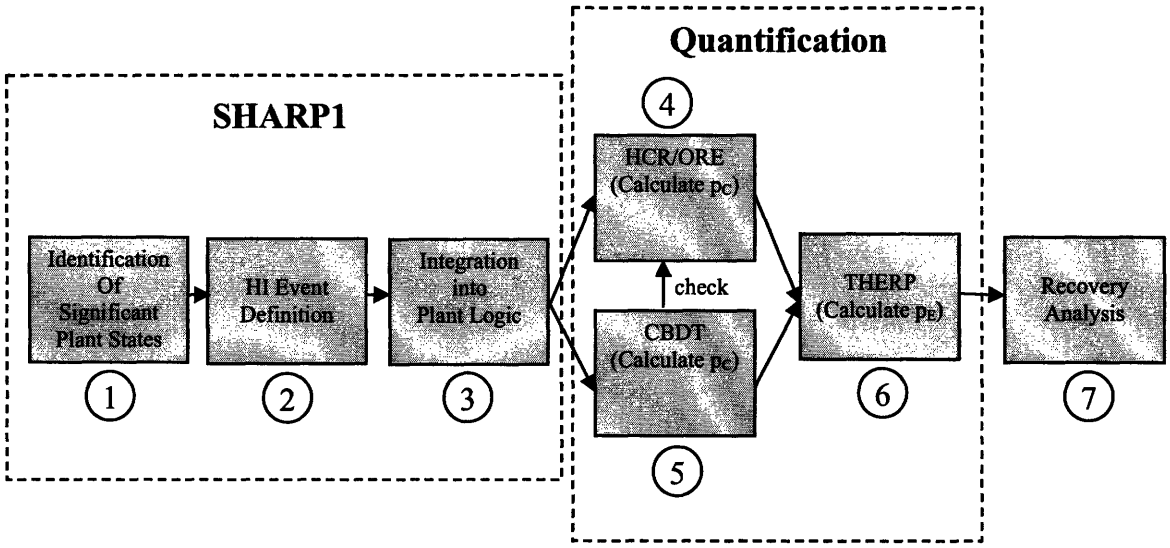


Figure 3: EPRI HRA Calculator (Incomplete) Summary Flow Chart

3.1 Basic Terminology: HI Types, Cue-Response Structures, and Timing

To aid in the definition of human interactions (HIs), SHARP1 defines three broad types of human interactions:[16]

Type A – Pre-initiating event HI.

Type B – Initiating event related HI

Type C – Post-initiating event HI

The type C His can be further broken down into:

Type CP – post-IE HI covered in plant procedures (in the plant logic model)

Type CR – post-IE HI "recovery" actions, not included in the plant model; these may include recovery of a failed system or an attempt at an alternate approach that is not covered in the emergency operating procedures (EOPs).

This paper is only concerned with post-IE HFEs, therefore only Type C HIs are relevant here. Type CR HIs are highly scenario specific and are not conducive to incorporation in the plant logic model – rather, CR events are defined during quantification and recovery analysis, and are modeled as correction factors for individual scenarios. The following steps, then, are primarily concerned with type CP human interactions.

In 1989, EPRI carried out a set of simulator experiments, the Operator Reliability Experiments (ORE)[27,28], in order to gather data to aide in HFE quantification. This program was meant to validate and support the HCR[29] TRCs which rested upon the cognitive categorizations of skill, rule, and knowledge based behavior. The experiments, however, did not support this grouping. Rather, groupings based on cue-response structures were identified by the ORE project as relevant for quantification. These cue-response structures, taken from EPRI TR-100259,[15] are presented here:

CP1: Response following a change in the plant damage state that is indicated by an alarm or value monitored parameter (e.g., response to a spurious pressurizer spray operation in a PWR).

CP2: Response following an event that gives rise to a primary cue (as in CP1) that has to be achieved when a parameter is exceeded or can be seen not to be maintainable below a certain value (e.g., initiate RHR when the suppression pool (SP) temperature exceeds 95°F in a BWR). These HIs involved a waiting period after the primary cue in order to reach a determined plant state.

CP3: Response following an event that gives rise to a primary cue (as in CP1) that has to achieved before some plant parameter reaches a critical value

(e.g., initiating SLCS before SP temperature reaches 110^{oF} in a BWR). This critical value can be regarded as a soft prompt, or secondary cue.

CP4: Performing a step in a procedure that is being followed as a result of a plant disturbance (e.g., inhibiting ADS before lowering level in a BWR, in response to an ATWS). The cue in this case is generally associated with completing the previous step.

CP5: Maintaining a variable parameter below, at, or within specific limits (e.g., controlling the level in a steam generator to prevent overfill or dryout). This is a control action.

Only CP1 – CP3 HIs can be quantified using the HCR/ORE process. The timing information for those cue-response structures is given below in Figure 4, where $T_{1/2}$ is the median crew response time to initiate the appropriate action, T_m is the time required to execute the appropriate action (the "manipulation" time), T_w is the time available to diagnose and initiate the appropriate action, and T_{sw} is the total time window between an initial cue (the time origin) and irreversible plant damage.

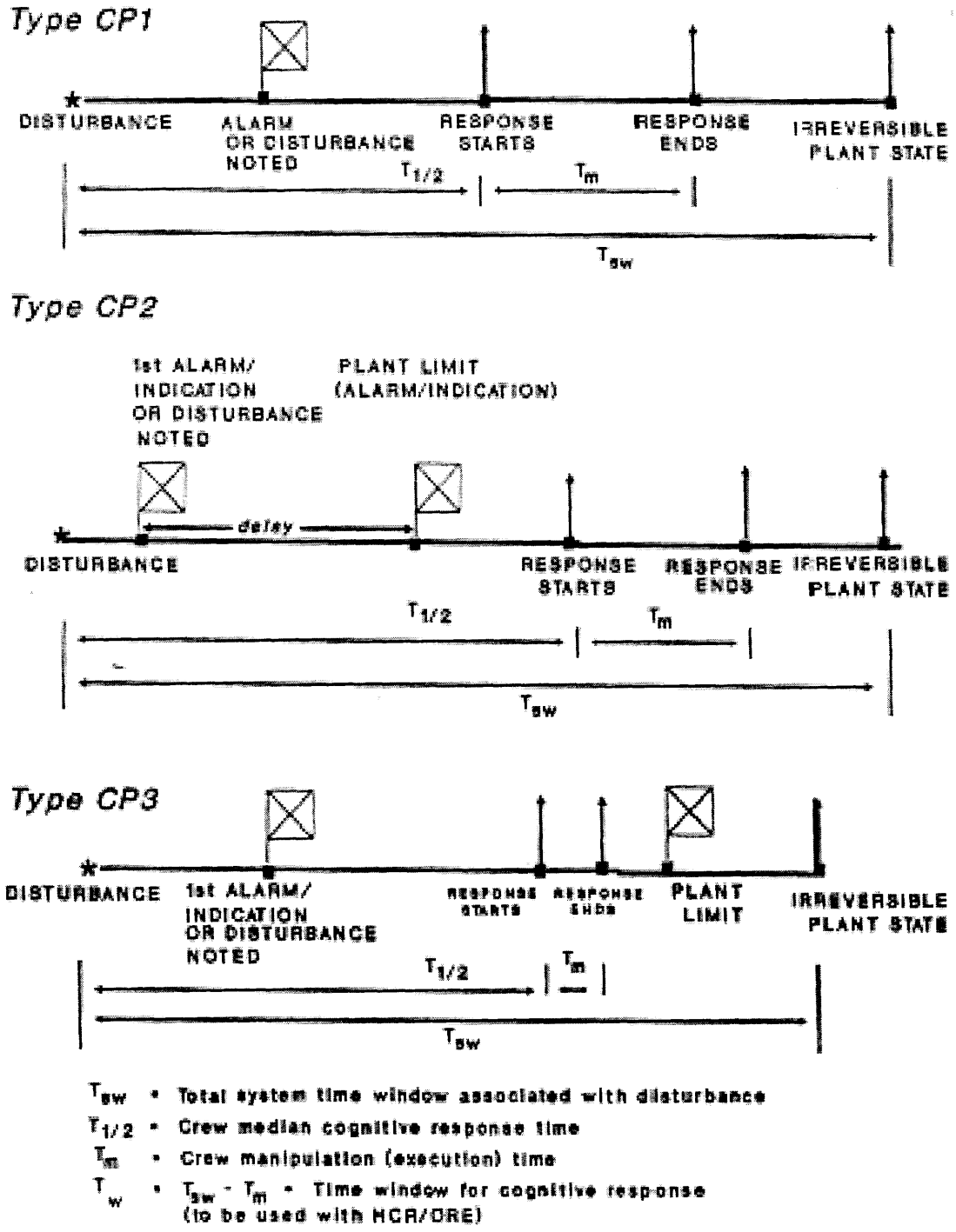


Figure 4: Cue-Response Structure Timeline[15]

3.2 Steps 1-3: SHARP1 and Event Definition

The first step is the identification of significant plant states. In order to do this, the HRA team must limit the scope and context of the analysis. This includes defining initiating event groups and documenting possible plant responses to each group, including success criteria definitions for each function in the event tree and identification of proper EOPs.[17] After the scope and context of the analysis is set, the team can proceed to qualitatively screen the interactions. This is done by identifying significant plant states and functions that are crucial to accident mitigation.

Once the team has identified the significant plant states, they should then strive to understand how operators, guided by emergency operating procedures (EOPs), respond – or should respond – to these plant states. This includes identifying failures in following the EOPs that can lead to unique and significant evolutions of the initial scenario.[17] This step in the event definition stage is where the team can dive into the details of specific scenarios. While there is no specific guidance on how this breakdown and impact assessment should be done, SHARP1 refers to a variant of NUREG/CR-3177 as a possible procedure. This variant involves identifying critical values of key plant parameters associated with EOP response points and evaluating HIs via these parameters. However the team decides to carry out this step, they should be thorough and include such components as:[15]

- examining why the HI is required
- understanding how the HI is carried out
- identifying scenario-specific performance shaping factors

- identifying and understanding dependencies
- understanding failure consequences on the plant
- understanding failure consequences on subsequent operator actions
- identifying possible effects of training on operator actions (similar to what ATHEANA calls identifying "unwritten rules")
- defining time sequence of the accident progression
- defining the cue-response structure of the HI

Now the team has assessed each failure mode and their dependencies, they can start to formally incorporate these modes and dependencies into the plant logic models. Again, here there is a cautionary note: failure modes should be modeled in only as much detail as necessary to capture the proper dependencies – too much detail at this level will make quantification significantly more difficult. Once this step is complete, the team should double check that the overall plant model is self-consistent, all assumptions are documented and well understood, and HI basic events are clearly defined and ready for quantification.

3.3 Steps 4-6: Quantification

The EPRI quantification method is based on dividing the human failure event into a failure to initiate the proper action (p_C) and failure to execute, given a correct initiation (p_E):

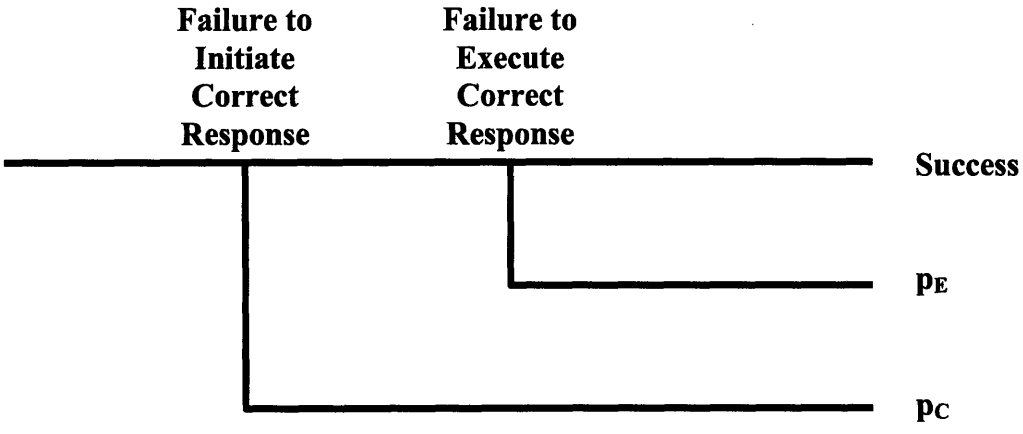


Figure 5: Generalized Event Tree for Calculating HEPs[15]

The probability p_E is quantified using THERP, where look-up tables for simple manipulation actions based on non-nuclear data, along with PSF correction factors, are used to find the probability of failure. The probability p_C, however, is more difficult to estimate. The first choice for estimating p_c is to simply look it up using a TRC – a curve, as in Figure 6, that correlates non-response probability to available time (T_w, $T_{w'}$). This time-reliability approach is called the HCR/ORE method, and is further described in Section 3.2.1. For a short available time (T_w), this method works quite well. However, the data used to create these curves fails to include those points where the operators misdiagnosed the situation and were on the "wrong path." For these cases, the extrapolated curve, seen as the dotted line, is not an accurate assessment of the HEP. For long times ($T_{w'}$), the actual probably of non-response would behave asymptotically, with a minimum that reflects a failure of the operator to properly diagnose the correct action, as seen in the figure below. Therefore, the HCR/ORE method is only useful for some

situations. Other situations require an alternate method; that alternate approach is generally the cause-based decision tree (CBDT) method described in Section 3.2.2.

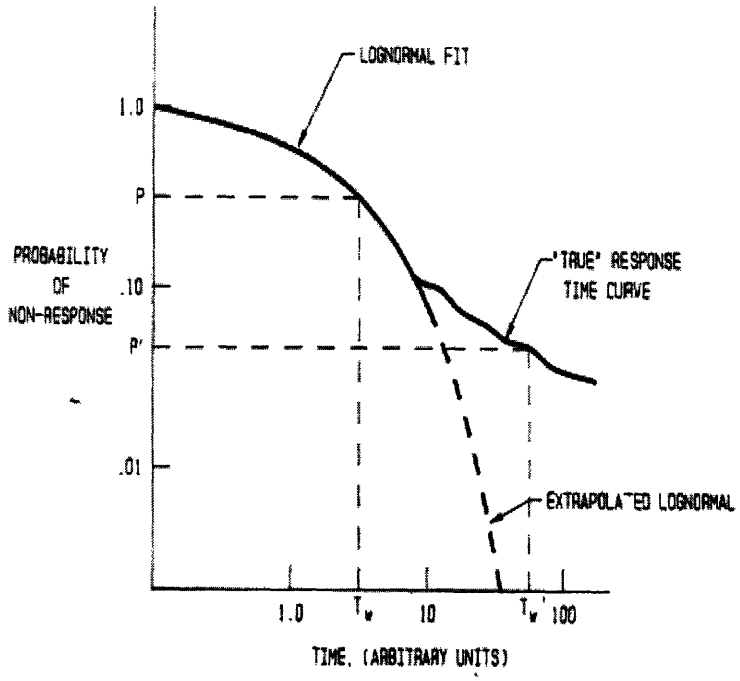


Figure 6: Conceptual Representation of the p_c Distribution as a Function of Available Time (T_w, $T_{w'}$).[15]

3.2.1 HCR/ORE

In addition to validating the cue-response structures, the ORE data also demonstrated that the lognormal distribution was a good approximation for HEP quantification, and so the HCR/ORE correlation was developed:[15]

$$p_c = \Pr(t_r > T_w) = 1 - \Phi\left[\frac{\ln(T_w / T_{1/2})}{\sigma}\right] \qquad [3]$$

where $T_{1/2}$ is the median response time, σ is the logarithmic standard deviation of normalized time, t_r is the response time, T_w is the available time, and $\Phi()$ is the standard normal cumulative distribution. This correlation is demonstrated in Figure 7 below.

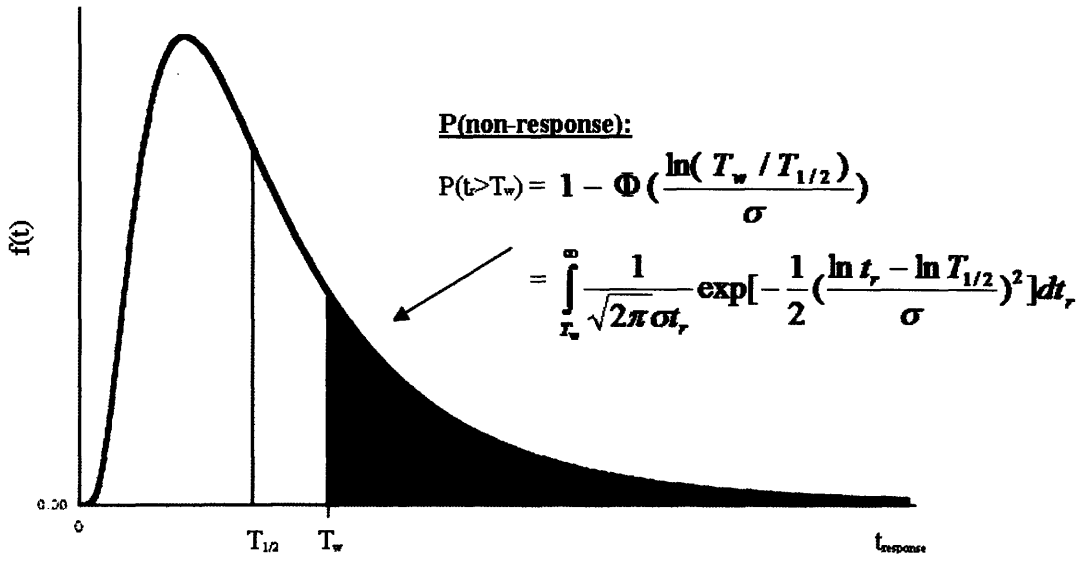


Figure 7: HCR/ORE Correlation, Lognormal Distribution of Response Time

Figure 8 is the lognormal complementary cumulative distributions for various cue-response structures, where normalized time is defined to be:

$$T_{normalized} = T_{real}/T_{1/2} \qquad [4]$$

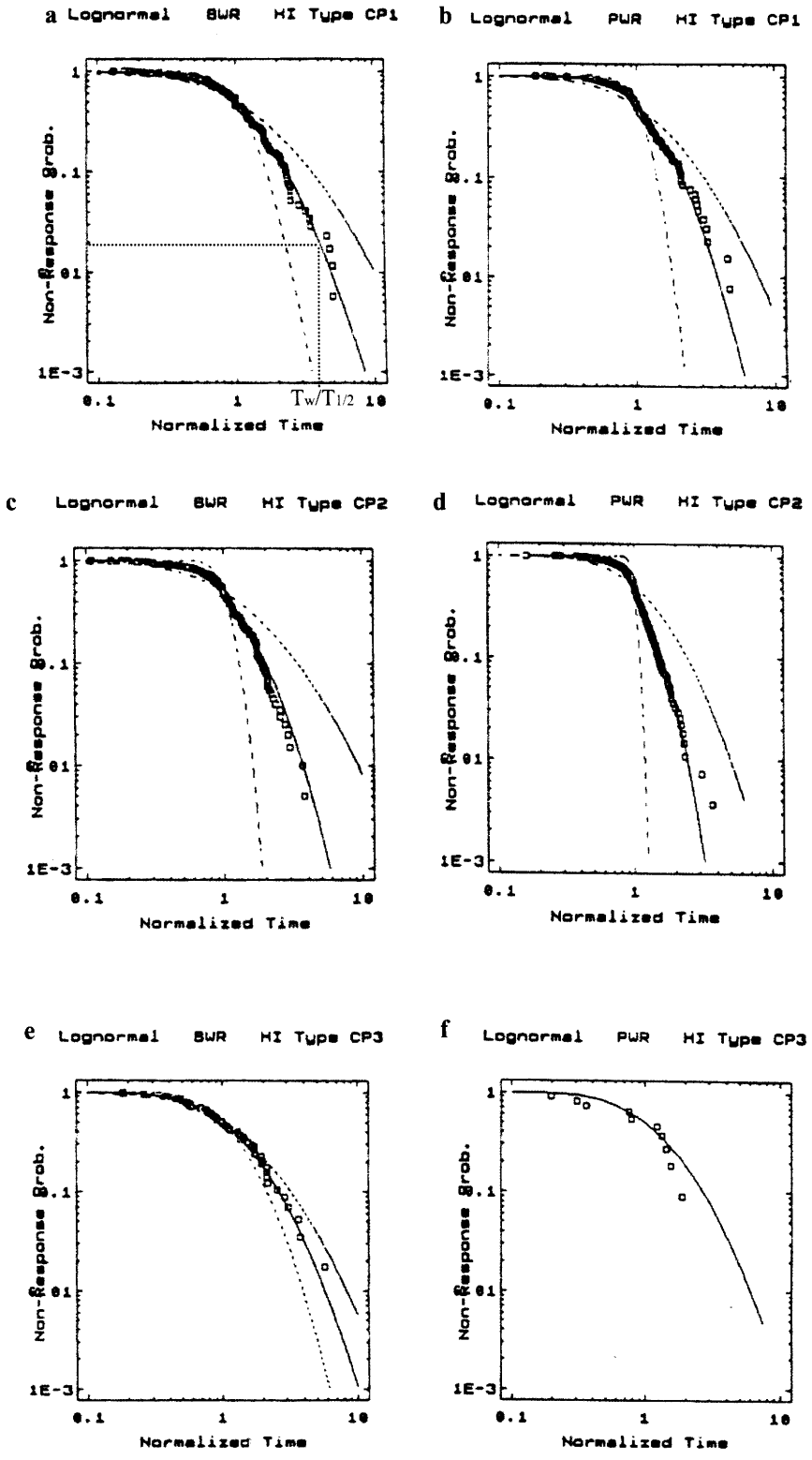


Figure 8: BWR and PWR HCR/ORE Non-Response Probability Curves for Various Cue-Response Structures; a) BWR CP-1; b) PWR CP-1; c) BWR CP-2; d) PWR CP-2; e) BWR CP-3; f) PWR CP-3. (Dotted lines are 90% uncertainty bounds on curve fitting; insufficient data to generate uncertainty bounds for PWR CP-3).30

Figure 8a demonstrates how the HRA team can apply these curves to simply estimate p_C. First they determine which cue-response structure is appropriate and find the $T_{1/2}$ and σ for that curve. Then, for a given scenario, they determine the time window of the total system ($T_{sw'}$) of a given HI from thermo-hydraulic codes (like MAAP). This system window must be adjusted (T_{sw}) to fit the HI event description (i.e., MAPP may give the total time until an event, but if the HI event begins with an alarm or parameter value, the time origin must also begin at the point that the alarm is sounded or the parameter value is reached). To use the curve the team must calculate the normalized time window:[15]

$$T_{w\_normalized} = (T_{sw} - T_m)/T_{1/2} = T_w/T_{1/2} \quad [5]$$

again, T_m is the time needed to actually execute the necessary action, or the "manipulation time." Using the appropriate cue-response curve, the team can then just look up p_C on the graph for a given normalized time window (see Figure 8a), or calculate it using Equation 3.

This method is only valid in the ranges where operating and simulator data is available – *extrapolation of these curves may produce unrealistically low estimates*, and CBDTs should be used instead for these cases. It is generally good practice to perform a CBDT analysis in addition to using HCR/ORE and use the highest (reasonable) HEP to be conservative. Furthermore, the probabilities taken from the HCR/ORE correlation are only as good as the inputs to the correlation: $T_{1/2}$ and σ. Sigma is generally taken from the ORE curves for a given cue-response structure. $T_{1/2}$ should be obtained from plant-specific, HI specific data, such as simulator experiments or operator/trainer judgment.

For the latter case, it is recognized that operations personnel may not have a good grasp for the time required for more complicated actions. The time ranges can, in these cases, be indirectly estimated by having the personnel identify ranges of key parameters within which they operators might act – the times could then be obtained from the thermo-hydraulics code. The median time would then be the middle of the given range.

3.2.2 *Cause-Based Decision Tree (CBDT)*

The CBDT method is used to find HEPs for situations where TRCs are not applicable, situations such as: CP-4 and CP-5 HIs, HIs where $T_w >> T_{1/2}$ (ample diagnostic time) and other areas where the HCR/ORE method is determined to be unrealistically low. This method is based on a decision tree decomposition of a HFE into situation specific failure mechanisms, associated PSFs and possible recovery modes. Each HI interaction is decomposed into two high-level failure modes, each of which are in turn broken down into four failure mechanisms. These modes and mechanisms are defined in EPRI TR-100259:[15]

Mode 1: Failures of the Plant Information-Operator Interface

a) The required data are physically not available to the control room operators.
b) The data are available, but are not attended to.
c) The data are available, but are misread or miscommunicated.
d) The available information is misleading.

Mode 2: Failure in the Procedure-Crew Interface

e) The relevant step in the procedure is skipped.
f) An error is made in interpreting the instructions.
g) An error is made in interpreting the diagnostic logic.
h) The crew decides to deliberately violate the procedure.

A decision tree is created for each failure mechanism (see Figure a-h). The nodes for each tree are PSFs which were predetermined to be important. The definitions for each PSF and any addition guidance provided to analysts on how to assess each PSF is provided in EPRI TR-100259:[15]

a) Availability of Information:

1. *Indicator Available in CR* – Is the indicator in the Control Room?
2. *CR Indicator Accurate* – Are the indications available accurate?
3. *Warn/Alt. Procedure* – If the displayed information is perceived to be unreliable, do the procedures direct the operator to alternate sources of information? Do they warn the operator the indication might be inaccurate?
4. *Training on Indicator* – Has the crew received training in interpreting or obtaining the required information under conditions similar to those prevailing in this scenario?

b) Failure of Attention:

1. *Low v. High Workload* – Do to the cues critical to the HI occur at a time of high workload or distraction? [Workload or distraction leading to a lapse of attention (omission of an intended check) is the basic failure mechanism for this mechanism.]
2. *Check v. Monitor* – Is the operator required to perform a one-time check of a parameter, or is he required to monitor it until some specified value? "Monitor" leads to a greater failure probability than "check" because the

operator might miss (exceed) the specified value if he does not check the parameter frequently enough.

3. *Front v. Back Panel* – Is the indicator displayed on the front or back panel of the main control area? Does the operator have to leave the control area to read the indicator?
4. *Alarmed v. Not Alarmed* – Is the critical value of the cue signaled by an annunciator? If the alarm comes in long before the value of interest is reached, it will likely be silenced and therefore ineffective.

c) Misread/Miscommunicated Data:

1. *Indicator Easy to Locate* – Is layout, demarcation, and labeling of the control boards such that it is easy to locate the required indicator?
2. *Good/Bad Indicator* – Is the MMI good or bad? Is it conducive to errors in reading the display?
3. *Formal Comms.* – Is a formal or semi-formal communication protocol (i.e., 3-way communication) used for transmitting values? Is the value always identified with its associated parameter?

d) Information Misleading:

1. *All Cues as Stated* – Are cues/parameter values as stated in the procedure? For example, if high steamline radiation is given as one of the criteria for a decision or action, at the time for the given action, the steamline radiation indicator would read high, not normal. The "no" branch is used if an indicator is *not* obviously failed but would not give the anticipated value (i.e., if the steamline was isolated).
2. *Warning of Differences* – Does the procedure itself provide a warning that a cue may not be as expected, or provide instructions on how to proceed if the cue states are not as anticipated?
3. *Specific Training* – Have operators received specific training in which the cue configuration was the same as the situation of interest where the correct interpretation of the procedure for the degraded cue state was emphasized?

4. *General Training* – Have the operators received general training that should allow them to recognize that the cue information is not correct in the circumstances? That is, is it something that every licensed operator is expected to know? For the steamline example, the answer would be "yes" because isolations are common; for instrument abnormalities that only occur under a very special set of circumstances, the answer would be "no" unless the operators had received specific training. Operators cannot be expected to reason from their general knowledge of instrumentation to the behavior of a specific indicator in a situation where they are not forewarned and there are other demands for their time and attention.

e) Skip a Step in the Procedure :

1. *Obvious v. Hidden* – Is the relevant instruction a separate, stand-alone numbered step or is it easily overlooked? A "hidden" instruction might be on of several steps in a paragraph, in a note or caution, on the back of page, etc.
2. *Single v. Multiple* – At the time of the HI, is the procedure reader using more than one flowchart procedure?
3. *Graphically Distinct* – Does the step stand out on the page? This effect is diluted if there are several things on the page which stand out.
4. *Placekeeping Aid* – Are placekeeping aids, such as checking off completed steps, used by all crews?

f) Misinterpret Instruction :

1. *Standard Wording* – Does the step use unfamiliar or ambiguous nomenclature or grammatical structure? Does it require any explanation?
2. *All Required Information* – Does the step present all information required to identify the actions directed and their objectives?
3. *Training on Step* – Has the crew received training on the correct interpretation of this step under conditions similar to those in the given HI?

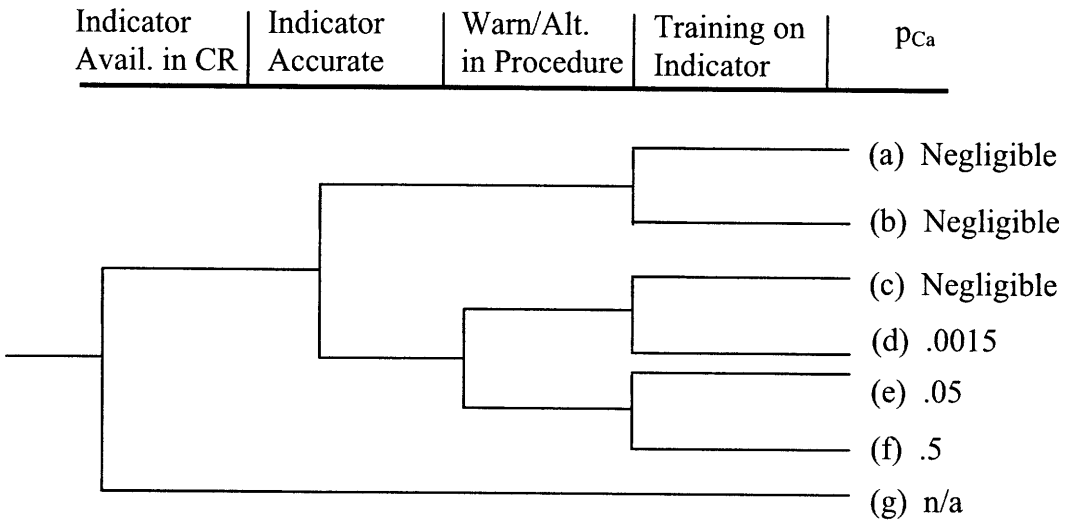
g) Misinterpret the Decision Logic :

1. *"NOT" Statement* – does the step contain the word "not"?
2. *AND or OR Statement* – Does the procedure step present diagnostic logic in which more than one condition is combined to determine the outcome?
3. *Both AND & OR* – Does the step contain a complex logic involving a combination of ANDed and ORed terms?
4. *Practiced Scenarios* – Has the crew practiced executing this step in a scenario similar to this one in a simulator?

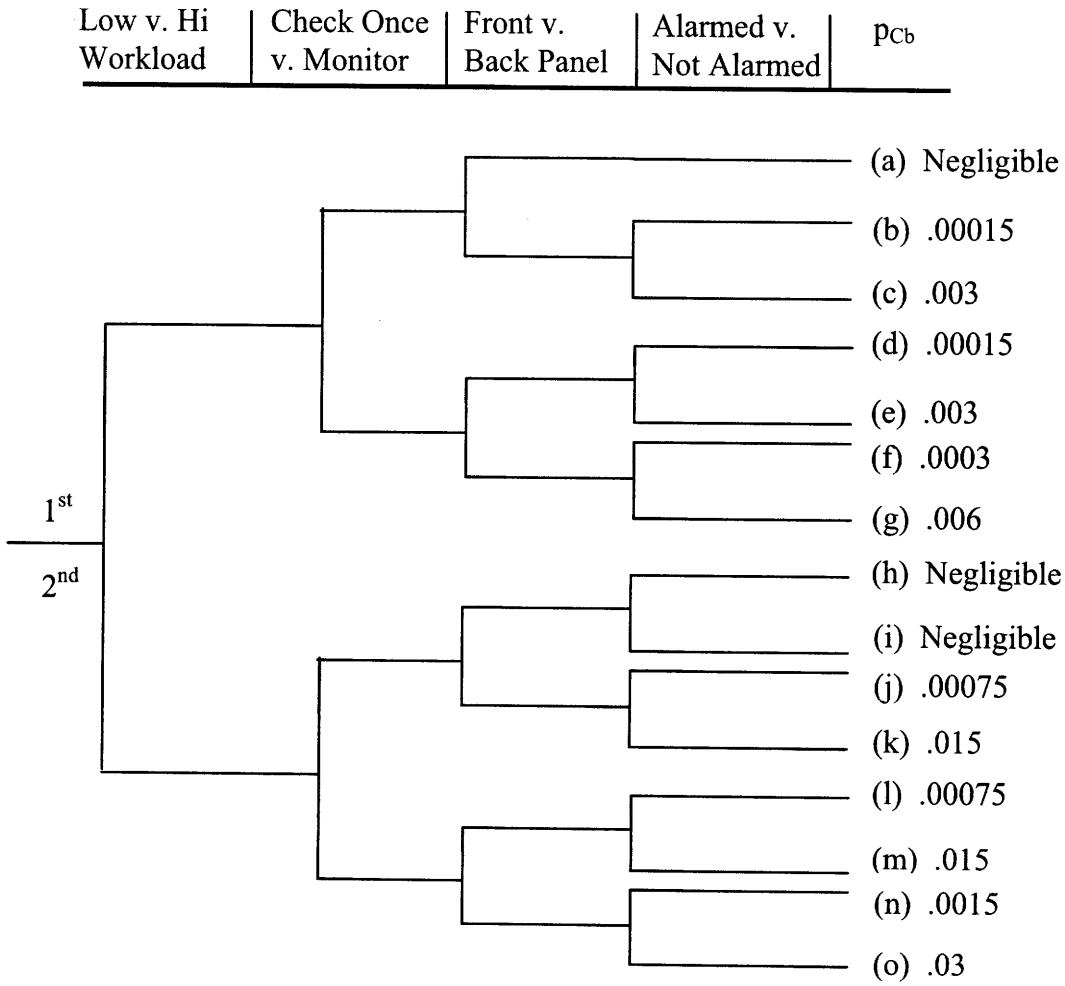
h) Deliberate Violation (*NOTE: this tree is rarely used in practice)

1. *Belief in Adequacy of Instruction* – Does the crew believe that the instructions presented are appropriate to the situation (even in spite of any potential adverse consequences)? Do they have confidence in the effectiveness of the procedure for dealing with the current situation? In practice this may come down to: have they tried it in the simulator and found that it worked?
2. *Adverse Consequences if Comply* – Will literal compliance produce undesirable effects, such as release of radioactivity, damage to the plant, unavailability of needed systems or violation of standing orders? In the current regulatory climate, a crew must have *strong motivation* for deliberately violating a procedure.
3. *Reasonable Alternatives* – Are there any fairly obvious alternatives, such as partial compliance or use of different systems, that appear to accomplish some or all of the goals of the step without the adverse consequences?
4. *Policy of "Verbatim" Compliance* – Does the utility have and enforce a strict policy of verbatim compliance with EOPs and other procedures?

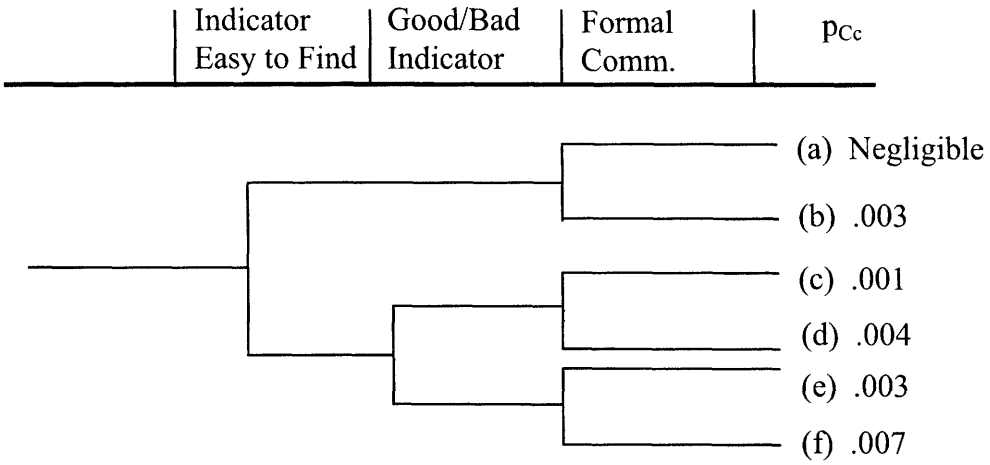
a) Data not Available:



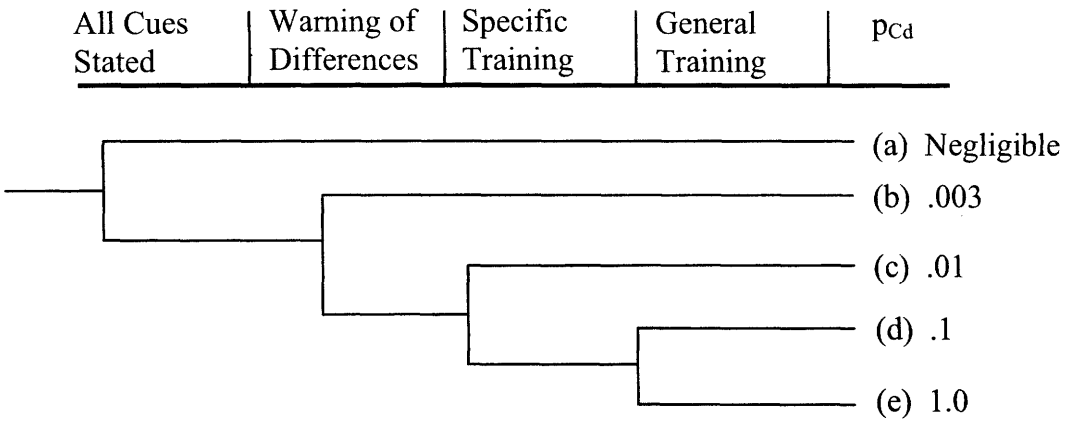
b) Failure of Attention:



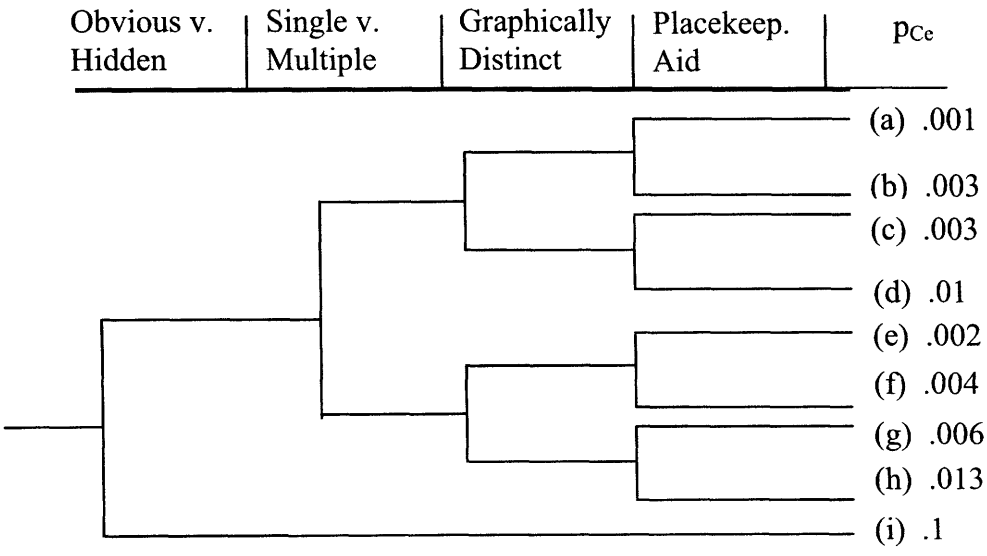
c) Misread/Miscommunicated Data:



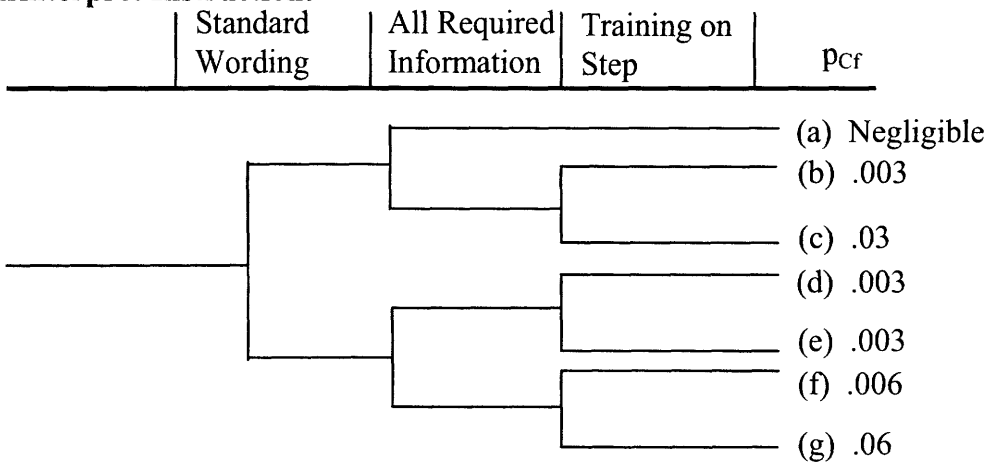
d) Information Misleading:



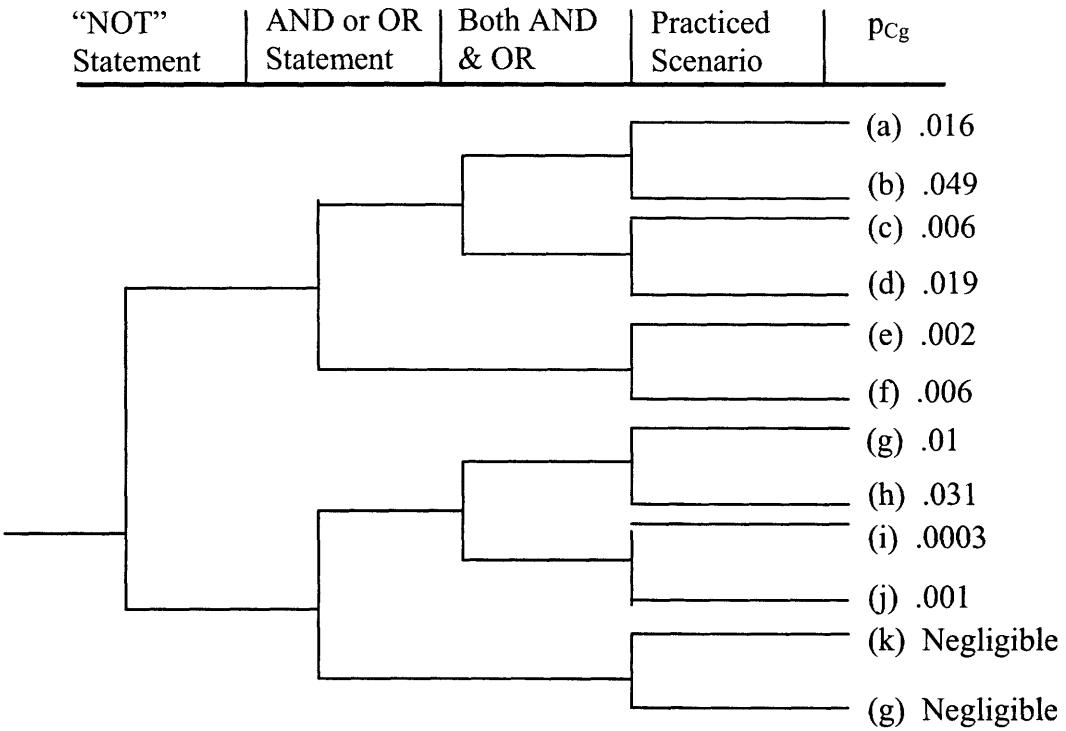
e) Skip a Step in the Procedure:



f) Misinterpret Instruction:



g) Misinterpret Decision Logic:



e) Deliberate Violation:

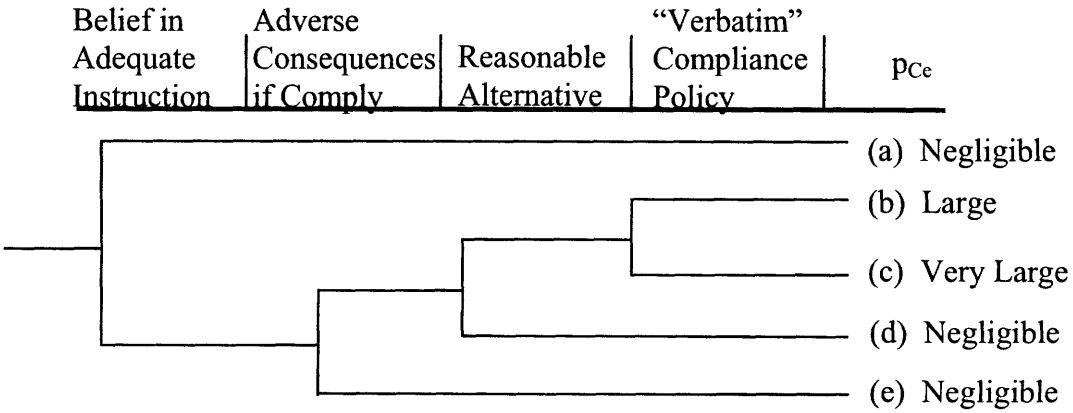


Figure 9: CBDT Failure Mode Decision Trees, a-h[15]

How then does an analyst get from the decision trees to a HEP? For each failure mechanism, the analyst chooses the branch that most closely describes the HI being quantified; the probability for that branch is the probability of failure due to that mechanism. The total HEP is then calculated according to the following equation:[15]

$$p_c = \sum_{i=1,2} \sum_j p_{ij} p_{nr}^{ji} \qquad [6]$$

where p_{ij} is the probability of mechanism j of mode i occurring, and p_{nr}^{ij} is the associated non-recovery probability for that mechanism.

Again, the probabilities p_{ij} come from generic decision trees for a given mechanism. These default probabilities are generally taken from THERP, but can be adjusted by the HRA team. Some of the THERP values have been altered because of feedback from the ORE and other reviews of THERP.[31] The probability p_{ij} is then taken from the branch which represents the expected plant conditions. Similarly, recovery factors can also be taken from THERP or estimated by the HRA team, and are influenced primarily by opportunity for review by another operator and time available (Table 3 and Table 4). To avoid over-crediting recovery, credit is only given for one recovery mechanism; however, override values may be used if credit for multiple recoveries can be justified.[31]

Table 3: Available Recovery Factors for a Given Recovery Time[32]

Recovery Factor	Time Effective
Other (Extra) Crew	At any time that there are crew members over and above the minimum complement present in the CR and not assigned to other tasks
STA	10 to 15 minutes after reactor trip.
ERF/TSC	1 hour after reactor trip – if constituted
Shift Change	6 hours after reactor trip given 8 hour shifts 9 hours after reactor trip given 12 hour shifts

Table 4: Example Recovery Checklist with Probability for Recovery (modified)[15]

HI Failure Mode		Review Type Available				
Tree	**Branch**	**Self-Review**	**Extra Crew**	**STA Review**	**Shift Change**	**ERF Review**
a	a	No Credit	0.5	No Credit	0.5	0.5
b	d	X	No Credit	X	X	X
c	f	No Credit	No Credit	X	X	X
d	c	No Credit	0.5	X	X	0.1
e	a	X	0.5	No Credit	X	X
f	i	No Credit	0.5	X	X	X
g	b	No Credit	0.5	X	X	X
h	c	No Credit	X	X	No Credit	No Credit

4 Juxtaposing ATHEANA and the EPRI HRA Calculator

The intent of this Section is to create an understanding of what each method accounts for in quantification and how that accounting is done. Specifically, we will look at the terminology, scope, time available, PSFs, response time variation, cognitive factors, and recovery. Below is a summary of the basic characteristics of ATHEANA and the EPRI Calculator (as illustrated by Figure 10):

Summary Comparison:

- Contexts Considered in Quantifying HEPs:
    - ATHEANA and the EPRI Calculator are very similar in the contexts they consider in HEP calculation; both factor in aleatory and epistemic PSFs (see definitions below), the accident sequence (including plant state/hardware failures) and cognitive factors into HEP calculation
    - ATHEANA takes a look at a broader set of PSFs and contexts. It also expands the accident sequence to include the consequences of a misdiagnosis beyond simple failure of the procedure (what will the operator likely do next given a specific misdiagnosis?)
    - The EPRI Calculator is prescriptive and limits the PSFs and cognitive factors the analyst must consider, thus enhancing consistency among analysts and reducing the level of resources needed. However, CBDT and ATHEANA still have the same approach to evaluating cognitive context.
- ATHEANA and the EPRI Calculator differ in objectives:
    - EPRI: To quantify HEPs in a consistent, transparent and reproducible fashion. And to provide a method that is not overly resource intensive and can be used by a PRA analyst without significant HRA expertise.
    - ATHEANA: To find contexts where operators are likely to fail without recovery, and quantify the associated HEP. This includes inquiring into how

operators can further degrade the plant condition while still believing their actions are correct.

- Quantification:
  - ATHEANA uses expert opinions after the error forcing context is described.
  - Depending on the type of HI – short or long available time, significant diagnosis – the EPRI Calculator defines the context based on the appropriate method. Quantification is then done using a time reliability curve or cause based decision trees.

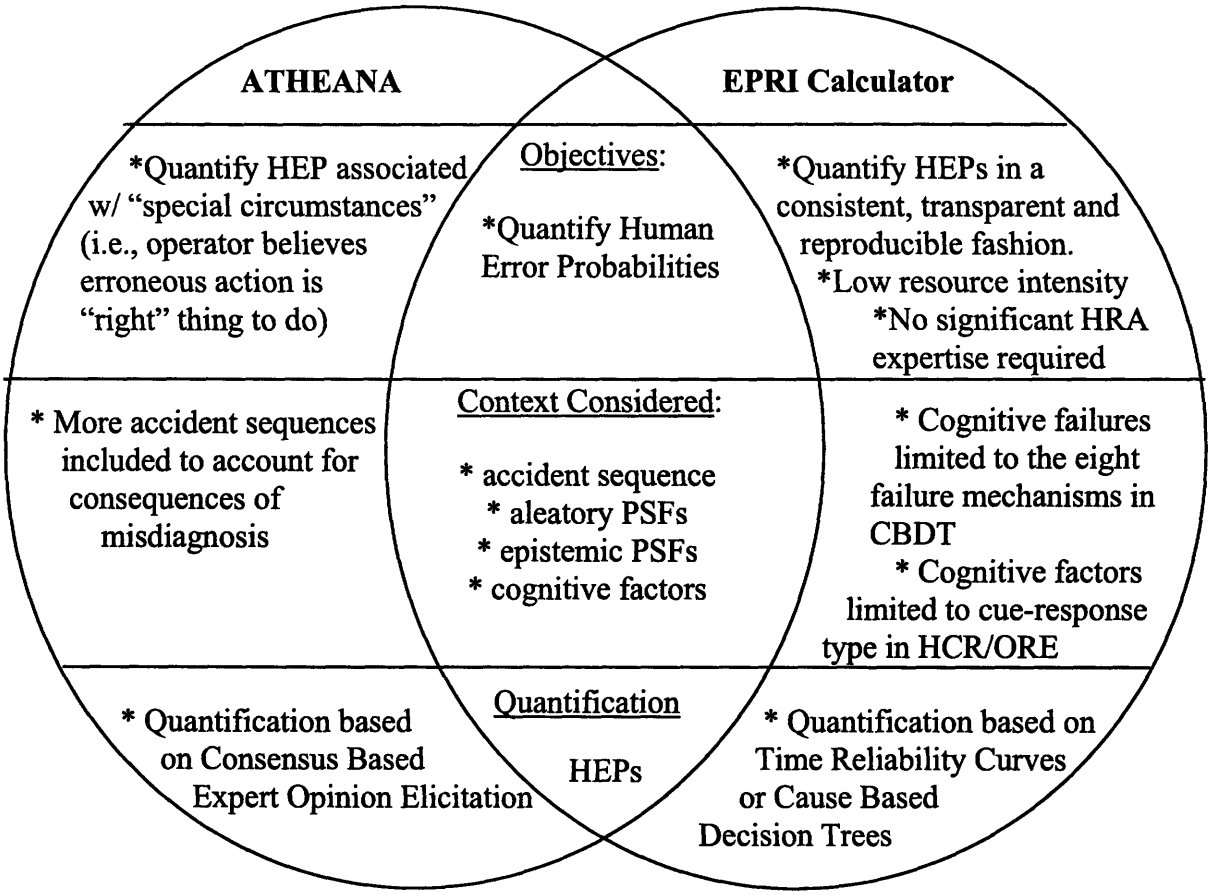


Figure 10: Venn Diagram of Basic Characteristics of ATHEANA and the EPRI Calculator

4.1 Terminology

Human Failure Event (HFE): This is a general term from ATHEANA that refers to points on the IE event tree where operators can take action. For example, a HFE would be "Operator Initiates Boron Injection given ATWS," but the particular manipulations required to perform this step are not generally considered HFEs, they are encompassed within a HFE. Also, when calculating P(HFE), the various evolutions of the accident scenario are taken into account through various EFCs.

Human Interaction (HI): Although there is no clear definition given in the EPRI documentation, it seems a HI is essentially the same as an HFE. Because the EPRI Calculator is geared towards plants with symptom-based EOPs, post-IE HIs are *procedure-based* actions with associated cues and decision rules.[15] The only apparent difference between HFEs and HIs is: HFEs incorporate all accident scenario evolutions (contexts), and HIs refer to either a specific scenario evolution or the average (expected) scenario evolution.

Error Forcing Context (EFC) and Failure Mechanisms a-h: The EFC is part of ATHEANA terminology used to identify contexts of a given HFE that might present a challenge to the operator, therefore increasing his likelihood of failure. An EFC in ATHEANA encompasses plant state and performance shaping factors (including training of the operators). The EPRI Calculator does not search for EFC in the same depth as ATHEANA, nor does the quantification of the EFC appear in the EPRI methodology.

Rather, EFCs only appear in the CBDT method, and are only defined with respect to the eight failure mechanisms defined in Section 3.2.2.

The ORE data implicitly address some PSFs, such as variation in crew experience levels, time pressure, workload levels, and minimum staffing. However, because these experiments were taken from training scenarios, they do not address specific EFCs which "trigger" human error (i.e., contexts which involve additional hardware failures). The HCR/ORE method is not recommended for situations where diagnosis is considered to be important, and so HIs where the analyst has identified significant EFCs should be quantified using CBDT instead.

Aleatory and Epistemic PSFs : ATHEANA considers two types of PSFs for quantification: aleatory and epistemic. Aleatory uncertainty is uncertainty that arises from that which cannot be known (randomness); aleatory PSFs would be those which cannot be known *a priori* (i.e., the time of day that the accident occurs). Epistemic uncertainty arises from incomplete information, and can be eliminated if more information is available;[33] an epistemic PSF would be one that can be known *a priori* (i.e., the safety culture of the plant or crew training level for a given HFE). The aleatory and epistemic PSFs are both plant and scenario specific.

4.2 General Approach and Scope

The EPRI Calculator was designed to be an easy to use, start-to-finish HRA tool for the average PRA analyst. For that reason, the quantification portion of this method was simplified as much as reasonably possible to reduce the amount of judgment the analyst must exercise. When possible, the HCR/ORE method is used, which only requires the analyst to determine $T_{1/2}$ (from training simulations) and the cue-response structure (a relatively easy judgment to make). In cases where this is too simple – where diagnosis comes into play – a slightly more judgment-intensive method is required: CBDT. CBDT requires the analyst to understand the plant state, procedures and cognitive factors for a given HI; however, this too is very prescriptive. There are eight failure mechanisms, each with predetermined PSFs and default probabilities, which the analyst has to evaluate.

Unlike the quantification process, the search process that EPRI uses to identify HIs and their appropriate contexts is not as formalized. SHARP1 presents "fundamental objectives" for the search process and tells the analyst to be aware of EOCs, but does not provide an active, in-depth search method for EFCs. Likewise, the CBDT quantification process includes misdiagnosis (see Figure 9d), but does not provide the framework to analyze what kind of misdiagnosis might occur and what the resulting consequence of this misdiagnosis might be (beyond the failure to complete that specific step in the EOP).

ATHEANA was designed primarily as a supplemental HRA method for existing HRAs, to analyze specific events that the HRA analyst feels are particularly important and have not previously been included in PRA models.[14] Unlike EPRI, ATHEANA is concerned with the depth of analysis rather than breadth. As visually apparent from Figure **1**, the search process is critical to ATHEANA. This search process requires not only an expert HRA analyst, but an entire team of experts (including operations personnel and thermo hydraulics experts). Instead of limiting context definition of the HFE to certain PSFs, ATHEANA approaches context identification through the eyes of the operator – what can make the operator think an inappropriate action is the "right" thing to do? To do this, ATHEANA turns to the behavioral sciences and previous severe accidents to gain perspective. It is with this specific misdiagnosis in mind that the HRA team approaches quantification. This makes a difference because, for a given EOC, it accounts for the effects of not performing a required step, but continues to ask – given this diagnosis, what actions would the operator reasonably take? What is the effect of those actions on the plant? This is the same for slips and lapses – they are only important if they induce a mistake. As expected, however, not all this information can be included in the quantification. The EFC, including everything, is whittled away to what the analyst believes to be the most important part of the EFC. This is then presented to experts who come up with a probability of failure for that EFC.

4.3 Available Time

Now that the general approaches have been addressed, let us look at the specifics, starting with available time. Both methods are required to factor in available time into quantification by defining the timing of an accident scenario as part of the HFE/HI description. EPRI then explicitly incorporates time available to initiate an action into quantification via the HCR/ORE curves. The available initiation time is a fixed number. For example, given an ATWS, for a PWR there is a 10 minute time window to initiate boron injection. However, before boron injection is initiated, manual actuation of the control rods must be attempted. The time available, then, for boron injection depends upon the time it takes the operators to get through the attempted manual actuation and get to the boron injection step.[25] The time available (T_w) used in quantification of 'failure to inject boron' would be adjusted to account for the average time it takes to get to the boron injection step (this is done by changing the time origin or adding a time delay, see Figure 13 and discussion in Section 5.1.2).

ATHEANA incorporates available time into quantification as part of the context of the scenario, as one of the several factors that experts must take into account in their evaluation of $P(UA_j|EFC_i,S)$.

4.4 Performance Shaping Factors and Response Time Variation

When using the HCR/ORE method (for HIs with short available times and little diagnosis required), epistemic and aleatory PSFs are considered only in their impact on the response time. The ORE curves take a statistical approach and treat both contributions as aleatory. The ORE data for a given curve includes different scenarios of a particular cue-response type, with different crews, operating under different PSFs[34]. (Note: the ORE did not attempt to address all PSFs, just the normal variation in PSFs that would be seen in various training scenarios, i.e., experience level of operators, workload, time pressure, hardware unavailability, ect., but *not* time of day or specific EFCs. For more details on the ORE, see References 27 and 30). The standard deviation, σ, should capture the various effects of the aleatory PSFs and some of the epistemic uncertainties (these would not be plant specific). The effects of epistemic PSFs are more firmly incorporated through the estimation of $T_{1/2}$ for a given HI because they are taken from plant specific data (simulation or expert opinion from those well acquainted with their specific plant conditions/culture). Aleatory uncertainty is also captured in $T_{1/2}$.

For ATHEANA, variation in response time is not explicitly accounted for – the PSFs themselves are examined instead. The effects of epistemic PSFs, in conjunction with available time and other contextual elements, are holistically accounted for by the experts by the way the think about the HFE or UA. However, the experts, by virtue of being experts, are expected to have a grasp of response times for different scenarios (or, if not timing explicitly, as mentioned in the EPRI method, experts would have a good

grasp of what actions are taken within certain plant parameter values, and would evaluate the probability of successes based on that). Also, context definition entails review of past operational experience, training records and running a simulator trial for an identified EFC. Aleatory PSFs provide the basis for the probability distribution of P(HFE) as described in Section 2.1.

When the time available is large with respect to the median response time, the EPRI method, using CBDT, only takes into account a set of pre-determined PSFs for each failure mode (Figure 9). The probability is then evaluated using THERP values or the analyst's own judgment. EPRI recommends that the analyst provide his own estimates of each factor in the various failure mode event trees. However, conversations with a HRA practitioner using the EPRI method[25] suggest that, in reality, the default THERP values are generally used for quantification unless there are extraordinary circumstances that require adjustment; this analyst chose this method because he valued consistency more than precision. Response time is incorporated in the THERP values and in the probability of recovery (via time available for recovery).

4.5 Cognitive Factors

The only cognitive factors the HCR/ORE portion of the EPRI calculator accounts for are the cue-response structures described in Section 3.1. This cue-response determination implicitly accounts for some of the informal rules of different plants. For example, if the procedure indicates "do x if temperature y or less cannot be maintained,"

this can be either a CP2 or CP3 HI based on operator training. If operators are trained to *wait until* the temperature exceeds y, then it would be a CP2 behavior; if operators were trained to do the action *before* the temperature exceeds y, then it would be a CP3 categorization.[15] This method does not account for misdiagnosis.

For more difficult tasks, or tasks which have a large available time, the CBDT considers cognition much the same way ATHEANA does – by trying to understand the scenario from an operator's perspective. In this case, though, the cues and cognitive responses are modeled in a generic way through the failure mechanisms described in Section 3.2.2. These cognitive failure mechanisms are based on two categories of failures: failure to obtain required information for decision or failure to make the correct decision.[15] Misdiagnosis is incorporated through failure mechanism d – information misleading.

ATHEANA's biggest strength is identifying contexts where the crew is likely to misdiagnose. This is done by systematically assessing when plant behavior is out of the expected range; when plant behavior is not well understood; when indications of actual plant state are not recognized; or when prepared plans and procedures are inappropriate or unhelpful.[21] Part of this EFC definition is identifying cues and effects of missing or misleading cues; however, this is not done with formal cue-response categorizations. Instead, the HRA analyst must have some familiarity with the behavioral sciences – or at least to the extent which it is presented in the ATHEANA manual.[14] The tutorial on behavioral sciences presented in NUREG-1624 defines and gives examples of several

types of cognitive errors, including: monitoring/detection failures, situation assessment failures and response planning failures. These categories are very similar to the failure mechanisms addressed in CBDT, but ATHEANA does not limit the analysis to certain PSFs. ATHEANA also provides some general guidance on how to identify these cognitive failures, however, unlike CBDT, there is no prescriptive link between cognitive failure mechanism and probability of failure.

In ATHEANA, before quantification, the HRA team tests the strength of the EFC performing a simulator study to see whether the context actually misleads the crew. The simulator lessons are then used as part of the expert judgment elicitation.

4.6 Recovery

As stated in Section 2.2, ATHEANA does not explicitly calculate the probability of recovery given an unsafe action has occurred. Recovery factors are assessed primarily as a screening tool – HFEs with a "high" likelihood of recovery should be dropped from further assessment, this includes slips and lapses, leaving mostly cognitive errors.[14] The point of the ATHEANA method is to find regimes where EFCs are strong enough that, given an unsafe action (i.e., given that the operators misdiagnose and perform an inappropriate action), recovery becomes highly unlikely. In this way, likelihood of recovery is assessed as part of $P(UA_j|EFC_i,S)$. If the analysts are not comfortable with this, they can explicitly calculate recovery, using THERP. However, timing now

becomes a bigger issue: it is difficult to assess whether a crew will realize that they have misdiagnosed and correct themselves in a short time frame.[14]

In the CBDT method of EPRI, credit for recovery is also quantified via THERP tables, and is based on time and a checklist of revisitation factors such as: availability of extra crew, shift change, etc. See Table 3 and Table 4 for details on the recovery actions credited. For each failure mode (with the chosen branch that represents the plant-specific HI at hand) the HRA analyst can check off the types of review which occur for a given mode of failure for that HI (Table 3). The numbers that appear in Table 4 are then taken from THERP tables (see THERP Figure 12-4) based on the available recovery time input ($T_{available\_for\_recovery} = Tw - T_{1/2}$).

For the HCR/ORE method, recovery is not addressed because the graph assumes correct diagnosis and correct initiation.

5 Simulator Studies: Lessons from Halden and Other Uses

The OECD Halden Reactor Project Task Complexity Experiment (2003/2004) provides data from simulator trials geared towards assessing complex tasks. We shall examine a sample of these trials and explore what they can tell us about quantifying human error probabilities.

The Halden experiments took place at the Norwegian HAMMLAB facilities, using the HAMBO simulator to simulate a Swedish BWR. There were seven crews that participated in the study, each crew consisting of three licensed operators (a turbine operator, a reactor operator, and a shift supervisor). These crews were borrowed from the reference plant, and briefed on the differences between the Swedish plant and the simulator. Finally, the procedures for anticipated operational occurrences were event-based.[19]

5.1 Time Pressure and Information Load Experiment

The Task Complexity Experiment included five scenarios, two of which we will examine here. The first scenario was an experiment intended to examine the effects of time pressure and information load on the operators. The accident sequence was an incomplete scram, where the operators had to initiate Boron injection. There were four variations of this sequence: base case, high time pressure, high information load, and both time pressure and information load. Time pressure is simulated by adding additional tasks that had to be done, thus shortening the available time window. Information load is simulated as additional tasks that the operators must attend to but do not require manipulations, therefore taking the operator's attention from the main task, but not a significant amount of time.[19] The actual conditions for the four variations are taken from the Halden report HWR-758,[19] and presented here:

Base Case:

Scram cause:

Isolation of feedwater system due to leakage in main feedwater (MFW) system.

Plant Condition at Time of Scram:

- Operation at full power
- Ordinary maintenance in safety system train C – emergency pumps and diesel generator blocked. (*note: this is the only pre-existing condition that influences the operators' actions and scenario; 2 of 4 trains in the safety system must be operational for plant success*)
- One of six main cooling water pumps has a failure: pump is in operation, but current measurement indicates 0%. The current measurement is input to the logic for main condenser.
- Fast transfer for train D in manual mode due to failure

Initiating Event:

Given a failure in feedwater pump A, stand-by feedwater pump B should start to compensate for failure of A to supply full flow. Simulated failure in pump B led to inadequate compensation, leading to a feedwater isolation signal (IM). This signal stops all feedwater pumps and closes the feedwater and steam lines.

At this point, automatic scram is supposed to occur. For this BWR, the scram valves connect to a pressurized Nitrogen tank. When the scram valves are opened, the pressurized Nitrogen will cause the rods to be driven into the core. Two of the eighteen scram valves fail. One valve fails "unavailable" and the second fails "stuck closed."

Without these valves working, 18 control rods will not be inserted into the core. There are also 12 control rods which are "stuck." This causes an incomplete scram (reactor not subcritical), and Boron injection needs to be manually initiated to reduce reactor power. There are no procedures for this action.

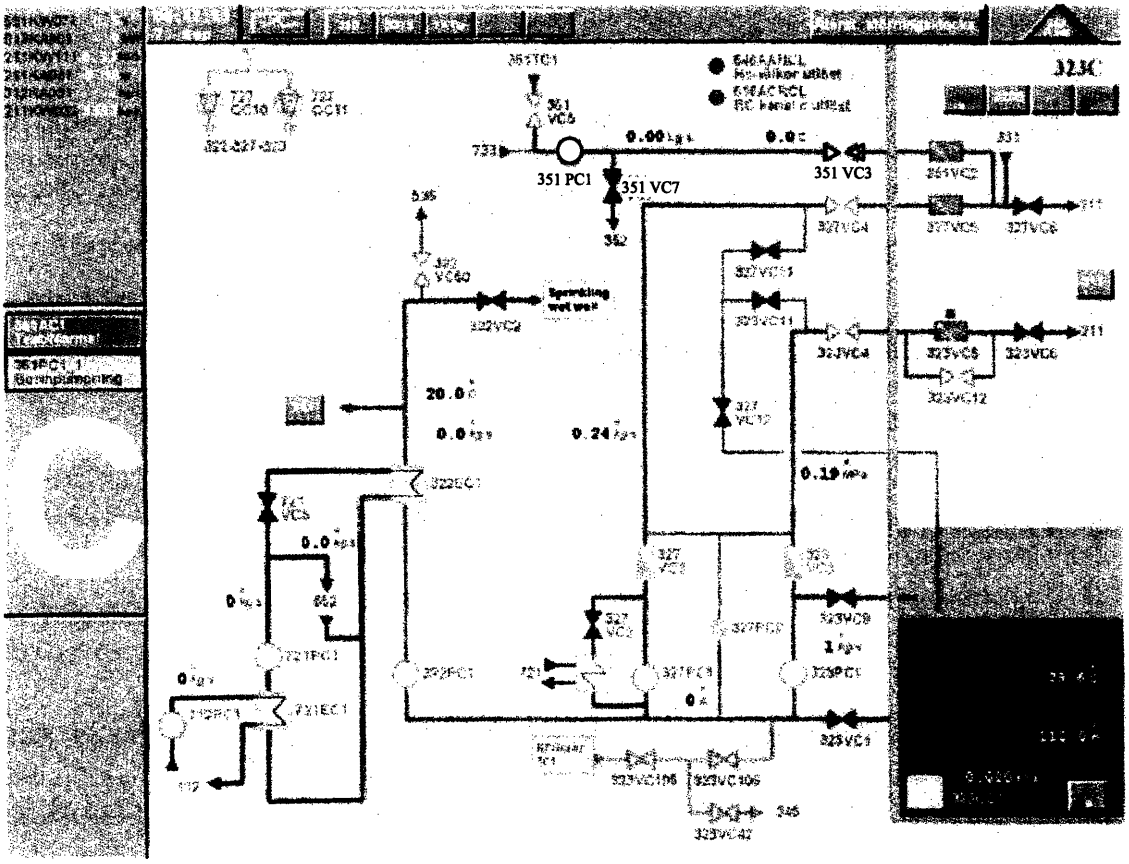


Figure 11: Boron System Train C Schematic[19]

Figure 11 depicts the Boron system train C. In order to initiate the Boron system, a push button must be pressed for one of two trains (trains C or D). It is also possible to maneuver the components individually and start the Boron system that way. To do this,

the operators must: close valve 351 VC7, open 351 VC5, open 351 VC3 and start pump 351 PC1 in Figure 11.

The auxiliary feedwater system (AFW) has four trains: trains C and D start pumping when the water level in the reactor vessel reaches 3.3 meters, and trains A and B start pumping when the water level in the reactor vessel reaches 2.0 meters. Train C is unavailable due to maintenance, and train D fails to automatically initiate, but can be manually initiated by closing a valve. The tricky part about this scenario is that, even though trains A and B are functional, the water level will initially decrease because they do not start until the water level is low.

Time Variation:

- AFW train A electrical failure, needs help from maintenance personnel
- AFW train D cooling pump fails, must be started before manually closing valve
- Pressure relief valve has to be closed

Information Load:

- Feedwater tank indicator level failure (one of five level indicators, but on large screen); would lead operators to believe there was a large LOCA
- High vibration on reactor re-circulation pump appears/disappears frequently
- High temperature in intermediate cooling system for reactor clean-up circuit alarm appears/disappears frequently

5.1.1 Results

Table 5 below presents the results for the time until Boron system initiation. For times greater than 5 minutes, a break down of the crew actions/thought process was given in HWR-758.[19] The reason for delayed times in these cases, under all scenario variations, was mostly due to prioritization. Before addressing the need for and initiation of boration, these crews spent from four to ten minutes trying first to insert the stuck control rods (including opening the stuck scram valve), start the AFW pumps, and open the pressure relief valves. Only one crew (crew C in the high time pressure/high information load scenario) had a delay because they had difficulty performing the action necessary to initiate the Boron system. Taking 10 minutes to be the time available (taken from the PWR analysis below[25]), the frequentist failure probability of this set of trials would be 1/28 (3.6E-02) if one includes the base case trials, or 1/21 (4.8E-02) if one only looks at trials under pressure.

Table 5: Time (min.) from scram until Boron system manual initiation. Bold values are those higher than 6:15, i.e., one standard deviation from the mean (the mean time was 3:47 and the standard deviation 2:28).[19]

Scenario	Crew A	Crew B	Crew C	Crew D	Crew E	Crew F	Crew G	Average
Base Case	2:45	0:49	3:27	3:59	5:10	0:49	5:56	3:16
Time Pressure	3:18	1:04	4:27	1:19	3:04	2:01	**11:43**	3:51
Info. Load	2:27	5:17	4:14	1:11	3:18	3:59	3:40	3:26
TP and IL	2:09	2:28	**6:31**	1:52	**6:20**	4:04	**8:55**	4:37

5.1.2 Quantifying this Scenario Using the EPRI HRA Calculator

For this quantification we used the EPRI HRA Calculator on a PWR that had a comparable accident scenario in its PRA – failure to initiate Boron injection given an ATWS. The major differences between this plant and the simulator are: partial scram is not part of the PRA for the PWR, only failure to scram; boration is a rare event for BWR operators, whereas it is routine for PWR operators; and the PWR plant uses symptom based operating procedures, whereas the simulation used event-based procedures. While it is recognized that the numerical results of quantification using the PWR model is meaningless in comparison to the Halden data, the process of quantification and trying to incorporate time pressure/information load into the results is what is valuable. The results from the EPRI HRA Calculator are presented merely for interest.

5.1.2.1 Quantification: HCR/ORE

Initiation of Boron injection is an immediate action step, and is classified as a type CP1 action.[25] Following the ATWS initiating event, an operator immediately takes action to begin boration. This HI is made up of four steps, as can be seen in Figure 12:

1) Placing the charging system flow controller (CS-FK-121) in manual and charge at the maximum rate (action: turning 2 knobs, one for manual and one for regulation demand)
2) Aligning the Centrifugal Charging Pump (CCP) suction to the RWST (action: opening a valve)
3) Isolating the CCP from the Volume Control Tank (VCT) (action: several valve manipulations)

4) Continuing boration until adequate shutdown margin reached

The timeline for this HI can be seen in Figure 13. The total time window in 600 seconds (10 minutes). The time origin is the signal to scram – the time window was not adjusted to include the time necessary to manually drop control rods because this was seen to be a concurrent action with boration (one operator would be working on boration while another would be working on the control rods). Therefore, the time delay is zero seconds. $T_{1/2}$ was taken to be 120 seconds (2 minutes) despite much shorter times demonstrated in training. [The sigma was taken to be 0.4.]

After $T_{1/2}$ and sigma were determined, the calculated p_c was 2.6E-04, and the p_e was found to be 2.6E-02 (with recovery), yielding a total failure probability of 2.6E-02.

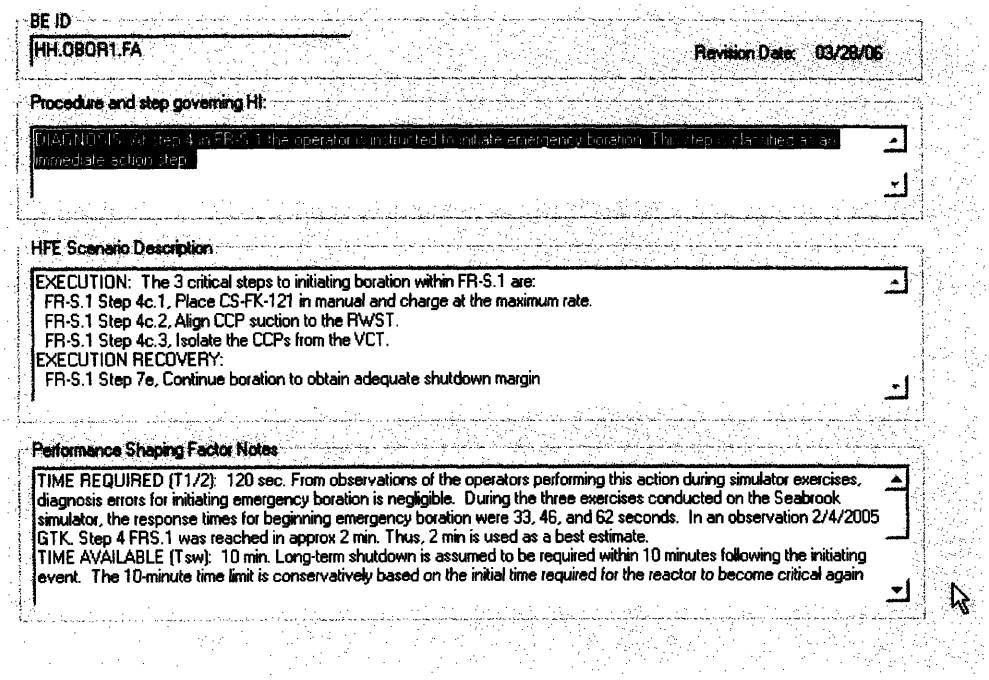


Figure 12: HI Event Description for ATWS Boron Injection in a PWR, EPRI HRA Calculator Screen[25]

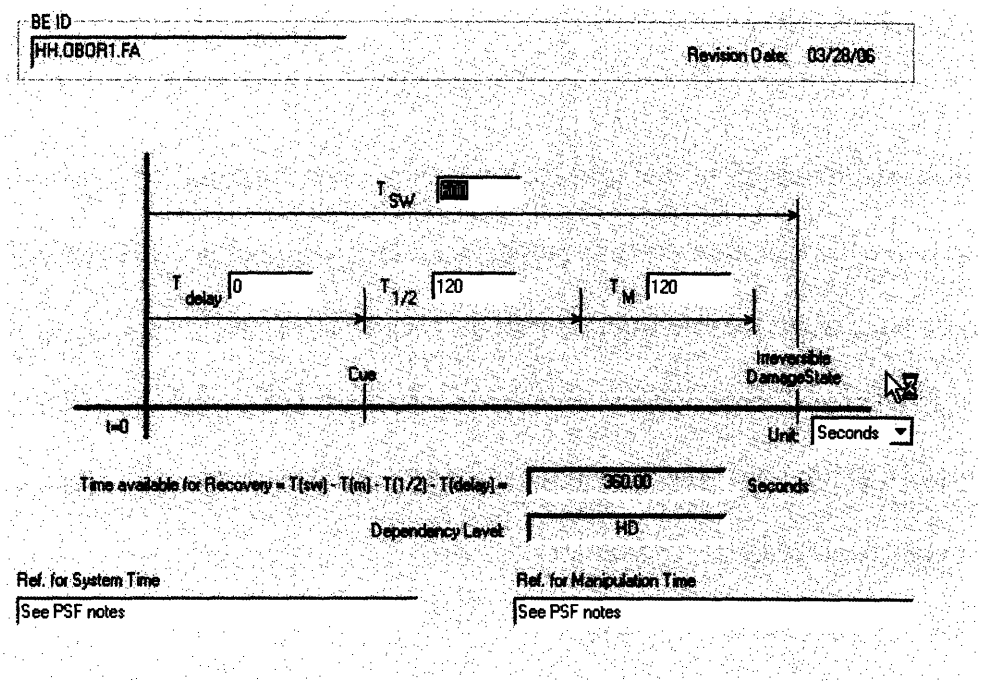


Figure 13: Emergency Boration Timeline[25]

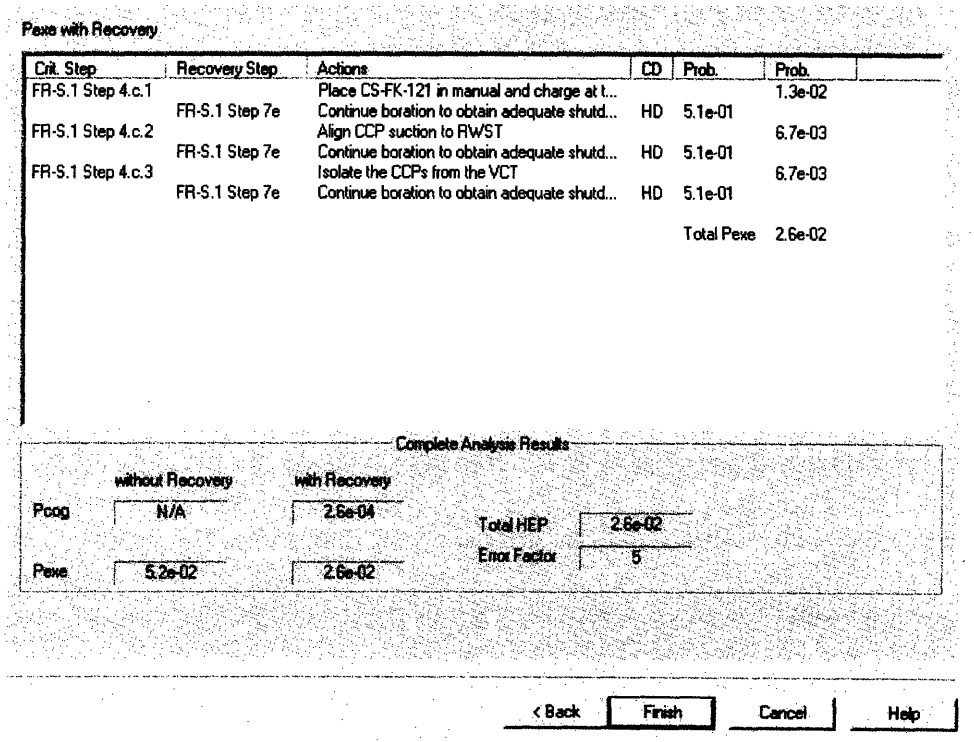


Figure 14: Probability of Execution and Summary of HCR/ORE Results for Boron Initiation Following ATWS[25]

5.1.2.2 Quantification: CBDT

An HRA analyst might be inclined to choose the CBDT method to model this action in order to see the effects of the time pressure and high information load explicitly, which cannot be done using the HCR/ORE method. Using CBDT, the event description (Figure 12) and timeline (Figure 13) are the same as the HCR/ORE method. After the event and timeline are defined, the analyst chooses the appropriate branches for each failure mode event tree (Figure 15 - Figure 22).

Right away we see one of the limitations of this method – for each failure mode, the analysts must choose the 'average' branch that represents their plant. This suggests that certain situations in which high time pressure or high information load may exist are not adequately captured. The figures below illustrate the nominal case of this PWR's plant operation. In this case, the only place where failure contributors are present are in trees e (Figure 19) and f (Figure 20). In e, "skip a step in the procedure," the PWR EOP is written such that this series of steps are not graphically distinct (i.e., it is not the first on a page, or formatted to stand out). In f, "misinterpret instructions," the procedure was determined to be slightly ambiguous because the nomenclature of different valves is similar.[25] These contributor determinations are based on definitions for predetermined PSF categories on the decision tree. Recovery factors were assessed for e and f, and it was determined that the recovery time of the event is of the time frame that only self review and extra crew review is credited (Table 3). The credit given for this potential recovery was to reduce the failure due to e and f by 0.5 each (see Figure 24).

The probability of execution for the CBDT method is the same as for HCR/ORE, 2.6E-02 (with recovery). The total probability for failure to inject Boron in the nominal case is 2.9E-02.

For this PWR, how would the nominal case change to account for the high time pressure and information load seen in the Halden scenario? In tree b, failure of attention, branch h (with High Workload) would be chosen instead of branch a. This change would not affect the overall probability. No other changes would be made, and so the total probability would remain unchanged, 2.9E-02.

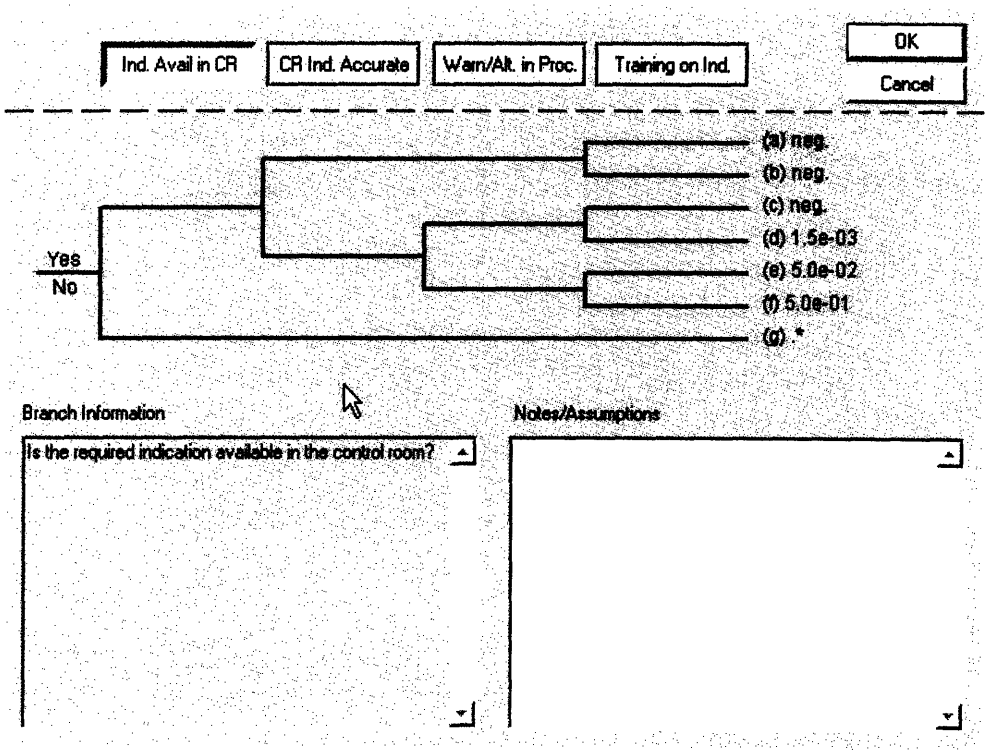


Figure 15: Decision Tree for Failure Mode a, Availability of Information[25]

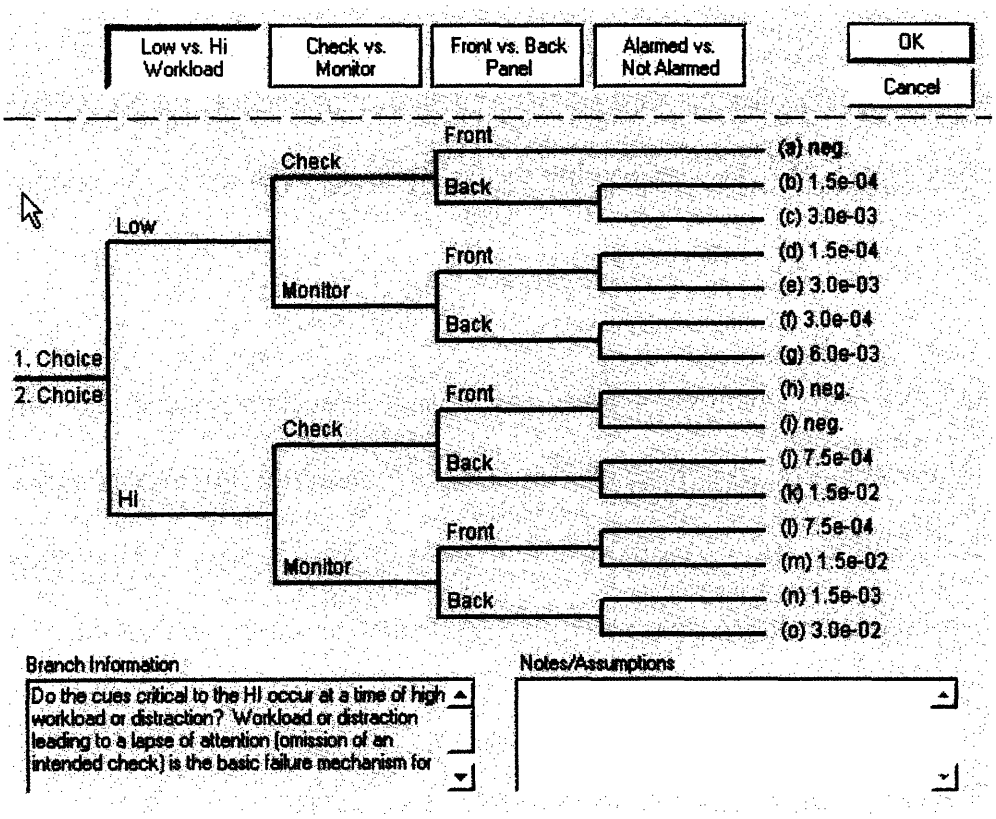


Figure 16: Decision Tree for Failure Mode b, Failure of Attention[25]

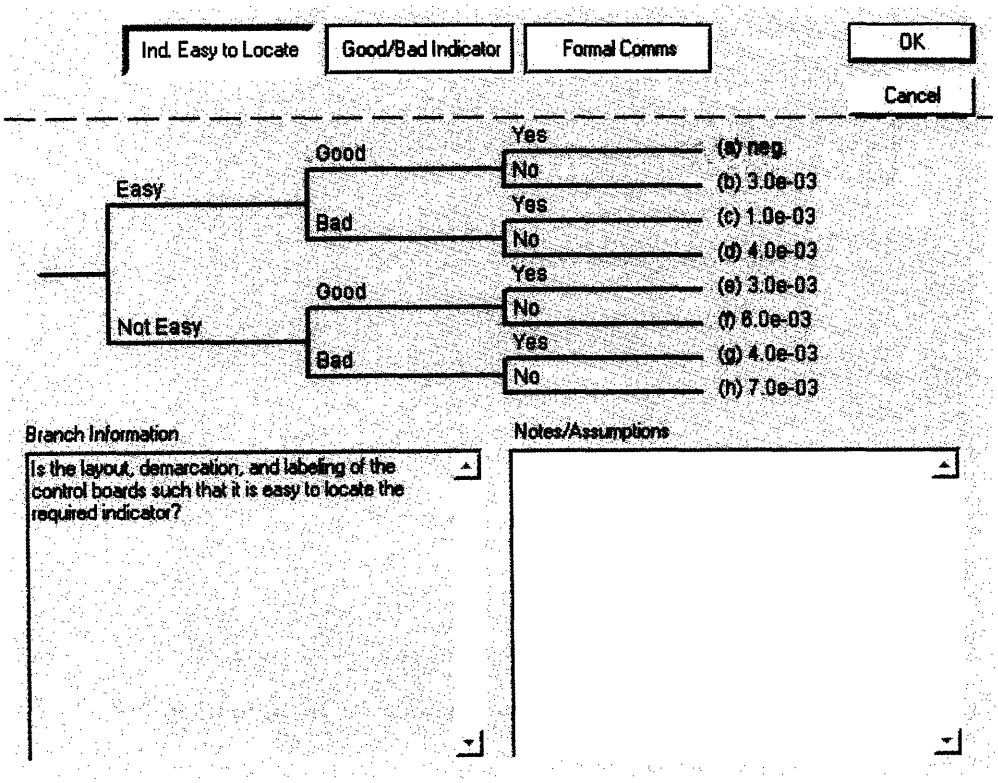


Figure 17: Decision Tree for Failure Mode c, Misread/Mis-Communicated Data[25]

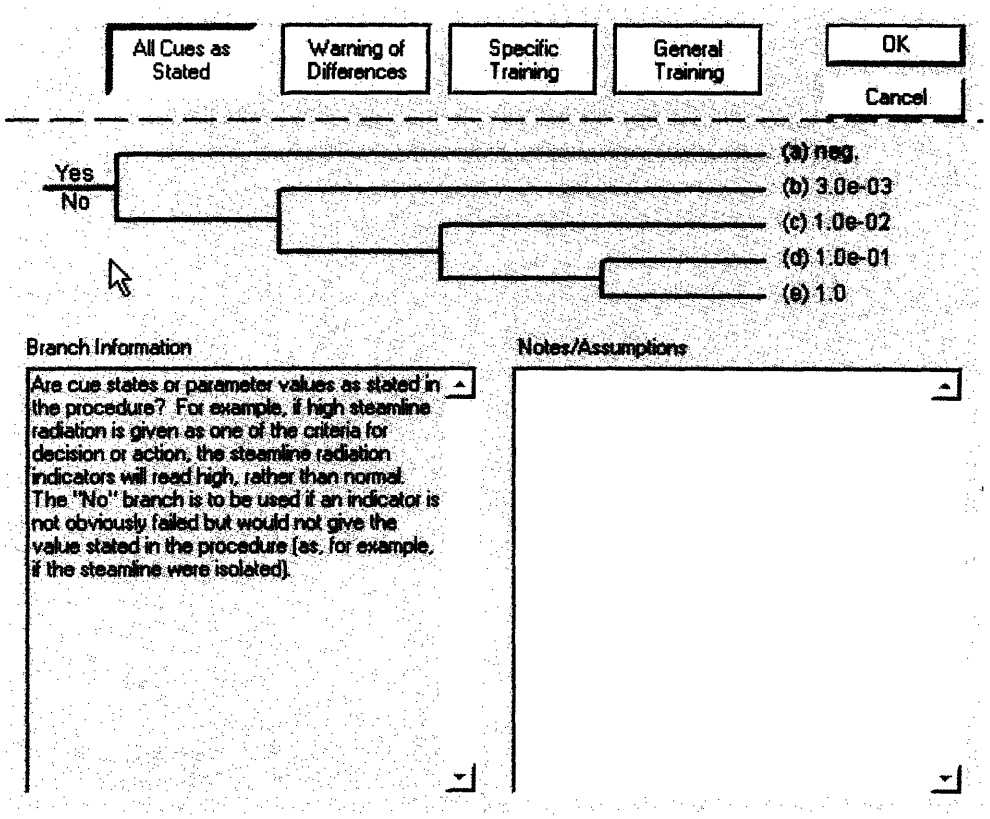


Figure 18: Decision Tree for Failure Mode d, Information Misleading[25]

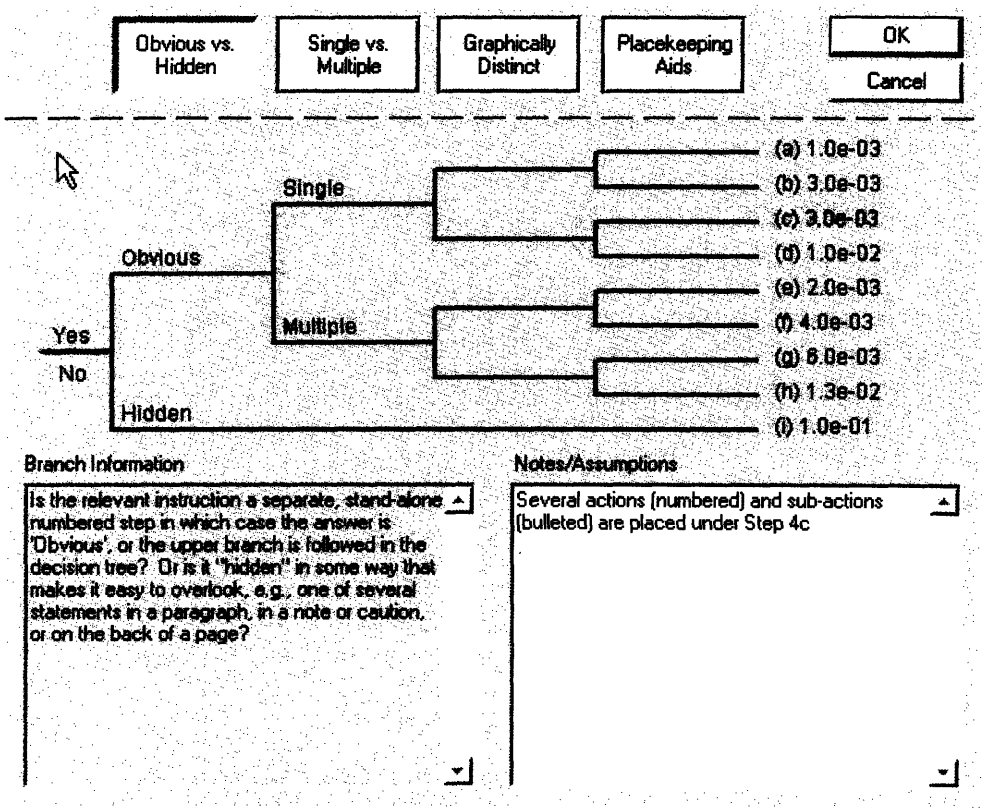


Figure 19: Decision Tree for Failure Mode e, Skip a Step in Procedure[25]

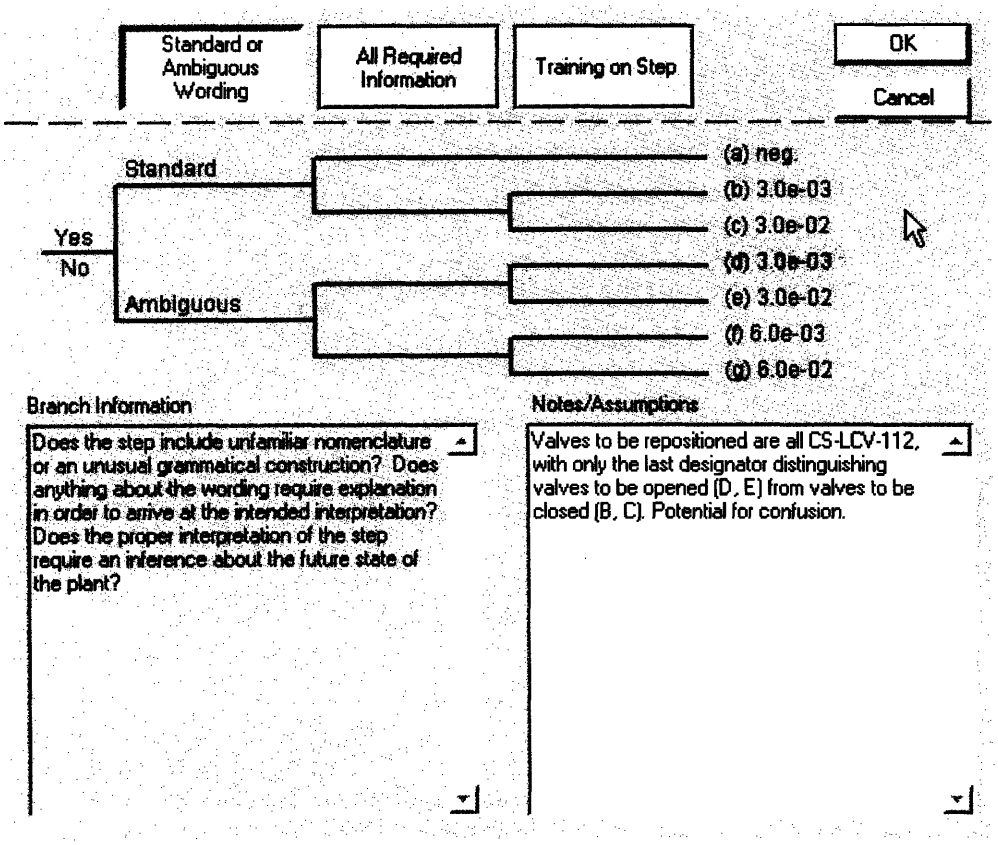


Figure 20: Decision Tree for Failure Mode f, Misinterpret Instruction[25]

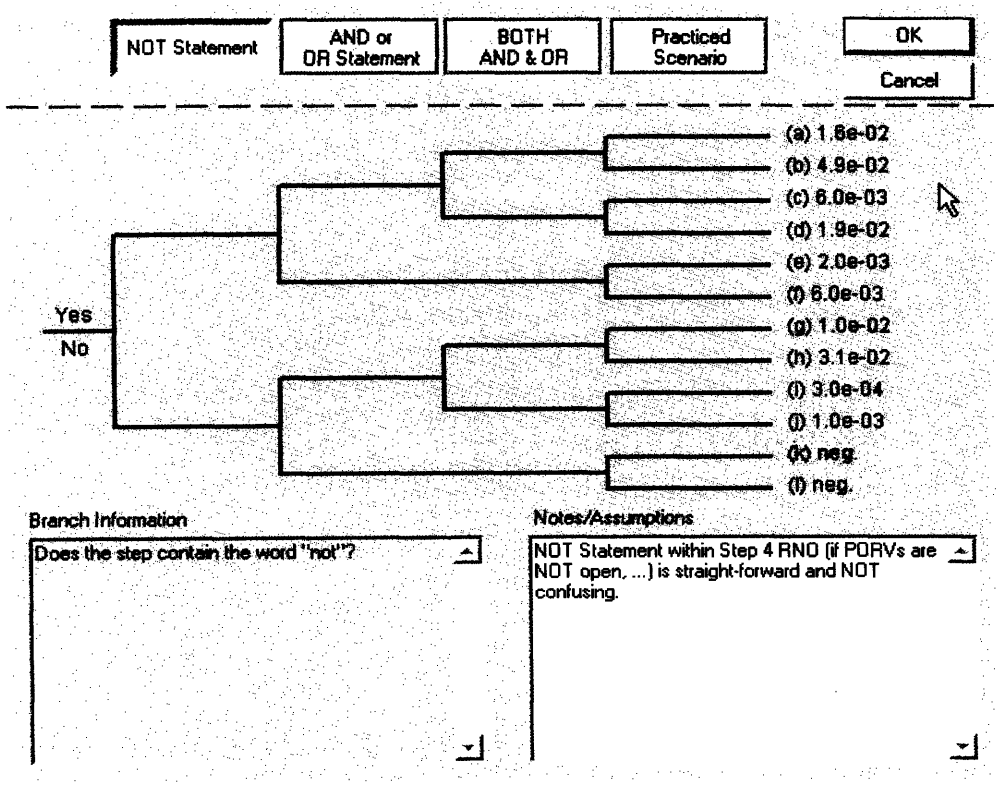


Figure 21: Decision Tree for Failure Mode g, Misinterpret Decision Logic[25]

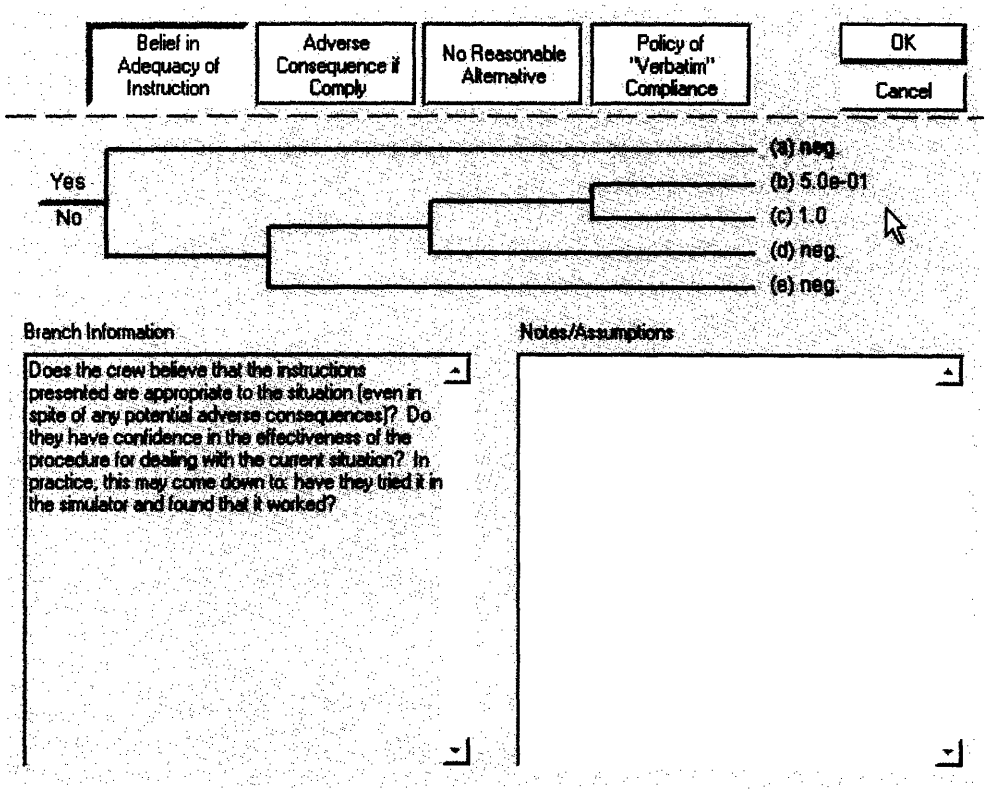


Figure 22: Decision Tree for Failure Mode h, Deliberate Violation[25]

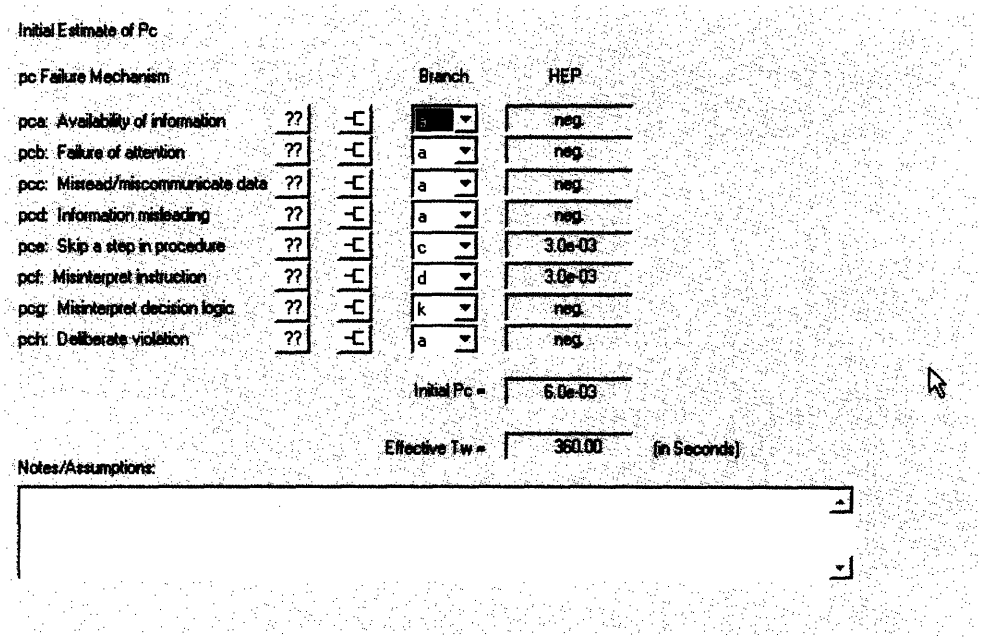


Figure 23: Summary Decision Tree Branches and Resulting Probabilities[25]

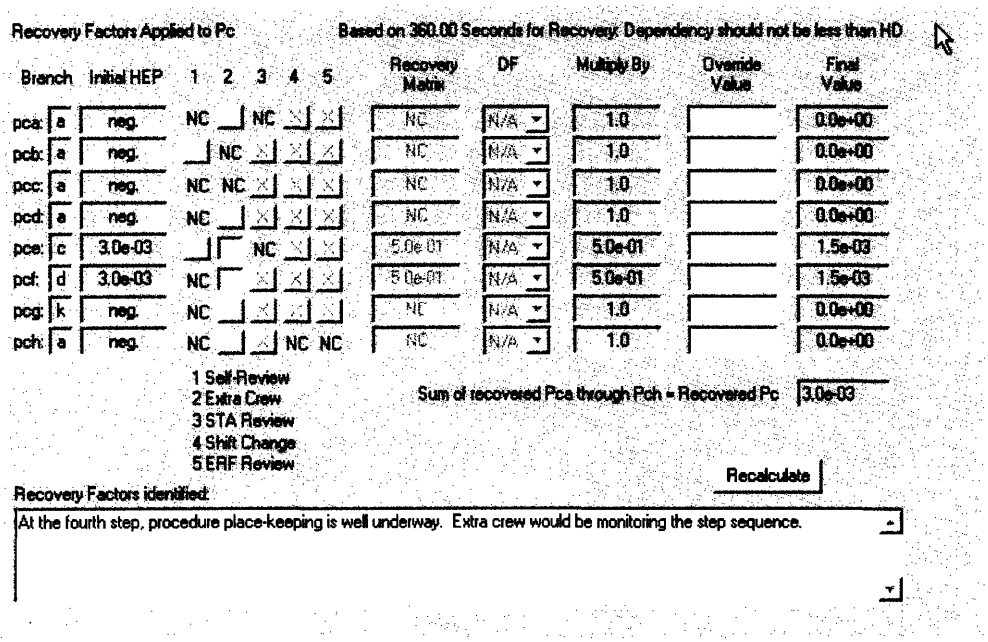


Figure 24: Recovery Assessment for Boron Injection[25]

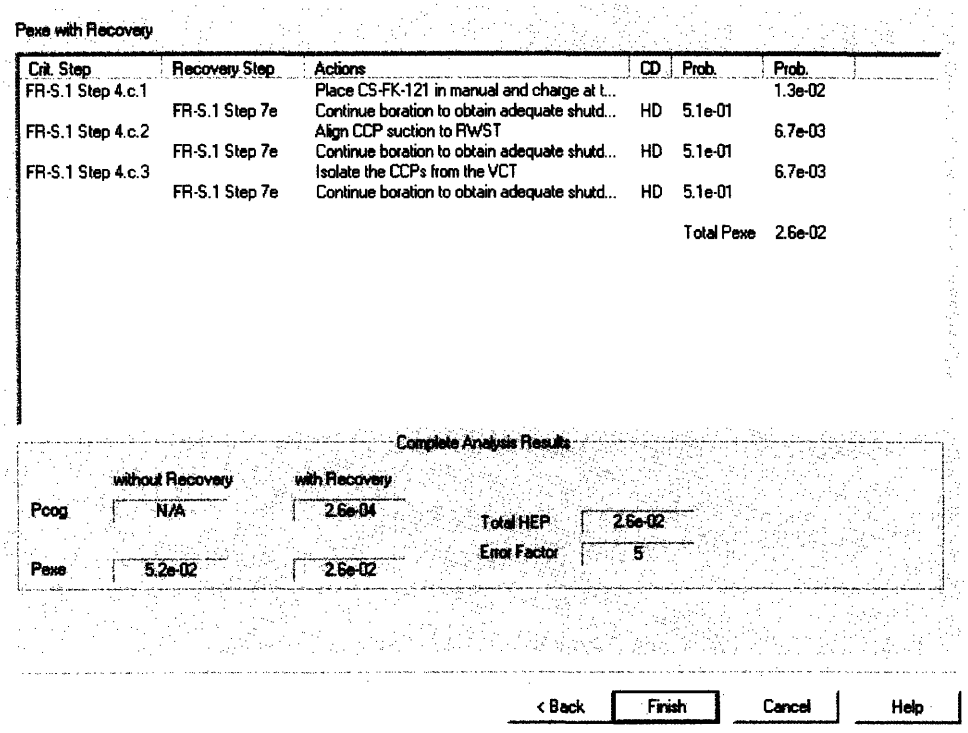


Figure 25: Probability of Execution and Summary of CBDT Results for Boron Initiation Following ATWS, Nominal Case[25]

In trying to use CBDT to produce a better estimate based on the Halden data, we can use override values to attempt to account for some of the results of time pressure and information load, i.e., mis-prioritization. In Halden, some crews had difficulty with prioritization, choosing to address failures of the AFW and the pressure relief valve before initiating boration. In a few of these cases, operators had difficulty with these tasks. Furthermore, these trials (along with other time pressure/information load scenarios not presented here) exhibited poor division of labor. To account for these effects of time pressure and information load, an analyst might choose to take away credit for recovery, as all control room personnel were otherwise occupied. It is not clear from the data that all personnel were in fact occupied the entire time, and in a few cases review (of varying degrees of success) was exercised, but to be conservative and to account for time pressure, lack of procedural guidance in prioritization and poor division of labor, we will assume that no extra crew is available. (If this assumption is not made, the failure probability with and without time pressure/information would be exactly the same, which we see from the Halden data is not appropriate). Without recovery credit, the total HEP becomes 3.2E-02.

5.1.3 Time Pressure/Information Load and ATHEANA

As quantification using ATHEANA requires expert opinion, it is beyond the scope of this work to attempt to quantify this event using ATHEANA; however, there is nothing in the ATHEANA method that precludes this analysis from being done.

5.2 Masking

A plant configuration which is misleading for operators in understanding the scenario and taking appropriate action is considered a masked scenario. In this scenario, the operators have two tasks to perform: "isolate a leakage from the cooling system for shutdown reactor" (task 1) and "diagnose the site and isolate a leakage of the reactor's pressure relief system."[19] There are four variations in this trial:

- base case: task 1 only, no masking
- variation 1: task 1 and 2, low masking
- variation 2: task 1 and 2, medium masking for task 2
- variation 3: task 1 and 2, high masking for task 2

where masking levels are defined in the scenario description below. The purpose of this experiment was not only to observe the effects of masking on task 2, but also to see the effect of a masked task 2 on task 1. Here we are only concerned with the effects of masking on task 2. A full scenario description is as follows, taken from the Halden report HWR-758:[19]

Scenario Description:

The base case scenario consists of a leakage in the shutdown cooling system outside the containment. The in- and outboard containment valves, which are in series on the same pipe, failed to close automatically upon the automatic isolation signal. To isolate the leakage, one of two containment valves must be closed or a large LOCA would ensue. One containment valve could be closed from the control room, while the

other had to be closed in the field due to a "stuck open" failure. This was determined to be an "easy" task.

The second task arises from a steam leakage through the reactor's relief system to the condensation pool. For the low masking case, there was a missing open indication of an open main relief valve (leakage through main relief valve). In the medium masking scenario, the leakage was through a closed valve in one pipe of the relief system, which resulted in an "open" indication for a main relief valve in a connected pipe. The high masking case was much like the medium case, but the temperature indication on the leaky valve falsely indicates a normal temperature.

5.2.1 Masking Results

Figure 26 contains the crew response time for stopping the leakage in the reactor relief system (task 2) in all variations of the simulated scenario. For this scenario, the success criterion for task 2 was a response time of less than 1080 seconds.[19] The data shows that masking affects the performance of task 2.

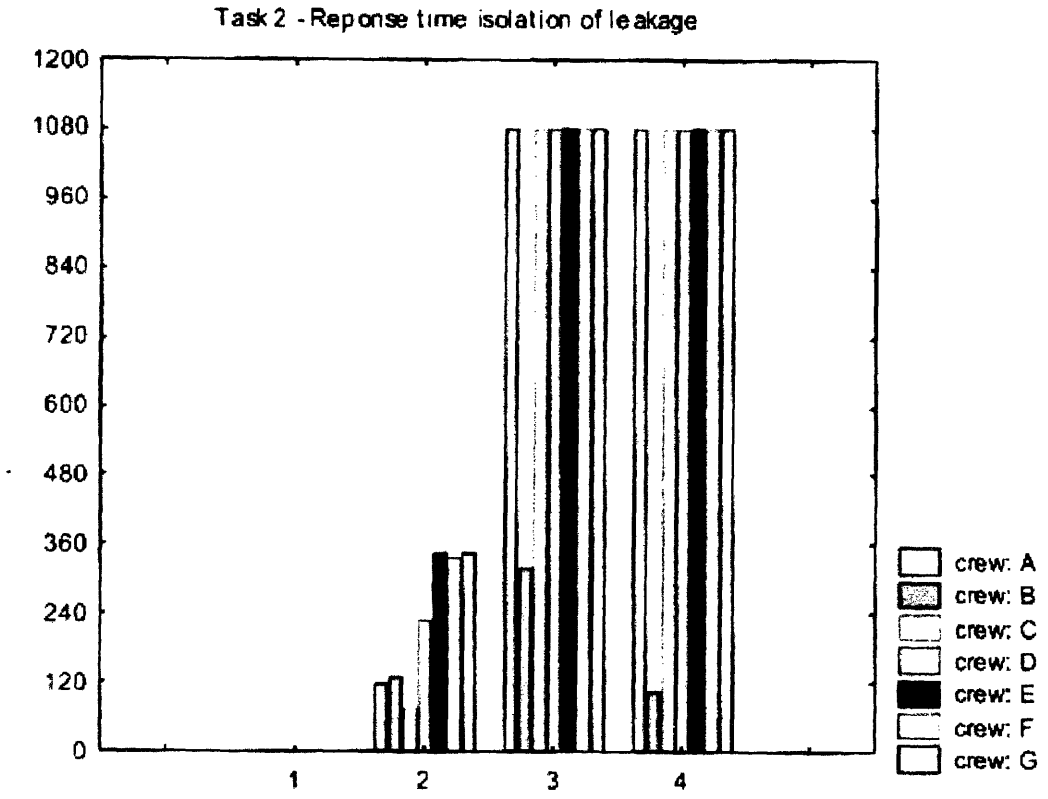


Figure 26: Crew Response Times for Task 2[19]

5.2.2 Quantifying Masking Effects with the EPRI HRA Calculator

Can the EPRI HRA Calculator adequately capture masking effects? The HCR/ORE method cannot, for the same reasons it was unequipped to handle time pressure and information load. What about CBDT? To answer this, we must go back to the failure mechanism decision trees to see if there are any relevant PSF that would change between the nominal scenario and the masking scenario. In this case, only failure mechanism d, information misleading, will be affected (see Figure 9d for the decision tree). In this case, assuming the crew had no specific or general training on this failure mechanism, the failure probability would be unity. The reasonableness of this number

cannot be assessed here, however, from Figure 26, we can see roughly 2/14 trials were successful with masking (yielding a failure probability of .86).

5.3 Implications of Halden Results for ATHEANA and the EPRI HRA Calculator

5.3.1 Updating the HCR/ORE Curves Using Halden Data

The ORE data were gathered in 1989. Since then, many advances have been made in the field of HRA, including improvements in communication methods (i.e., 3-way communication) and EOPs.[25,34] Furthermore, the distribution of sigmas for each cue-response type is relatively large (Table 6), and could use more resolution to confirm the reasonableness of the cue-response categorization. In order to update the HCR/ORE curves, crew time response data are needed. The question then becomes: can data from a Norwegian simulator using Swedish crews help us model American reactors? The answer is: sometimes.

Table 6: Sigma Values for HCR/ORE Curves by Cue-Response Structure[15]

Plant Type	Cue-Response Structure	Values for σ		
		Average	Upper Bound	Lower Bound
BWRs	CP1	0.70	1.00	0.40
	CP2	0.58	0.96	0.20
	CP3	0.75	0.91	0.59
PWRs	CP1	0.57	0.88	0.26
	CP2	0.38	0.69	0.07
	CP3	0.77	*	*

The Halden crew used event-based AOPs (EBAOP). The ORE curves were obtained using symptom-based EOPs (SBEOP). The difference between event-based AOPs and symptom based EOPs can be significant enough to preclude Halden data from being reasonably used to update the HCR/ORE curves. There are several reasons for this. The main reason is EBAOPs require diagnosis and prioritization, both of which require time which is saved by following a symptom-based procedure. As we saw in the ATWS scenario, the time it takes a crew to perform an action is influenced by prioritization. The $T_{1/2}$ for the PWR was very conservatively put at 2 minutes, whereas the median time for Halden ranged from 3 minutes to just over 6 minutes depending on the scenario variant

(overall median was about 3.5 minutes). The time variation for EBAOPs is overstated compared to what is reasonable for SBEOPs because the former includes cases where the actions were mis-prioritized (AFW pumps and/or pressure relief valve addressed before Boron injection). This mis-prioritization in a SBEOP would require a deliberate deviation from procedures, which is a rare event.[25] A similar time variation due to diagnosis time can be seen in the LOOP/Fast_transfer_of_bus_bars time pressure/information load scenario from Halden (not presented here, see Reference 19).

Another reason that this data may not be valid for the ORE is that SBEOPs are perhaps less sensitive to time pressure and information load.[25] For information load, many SBEOPs start with a master silence of alarms (i.e., all Westinghouse plants), where operators then only check the alarm list and indicators, as the EOP dictates. While information load will probably still affect the operators, the effect is minimized because the procedures allow the operators to focus on the task at hand without, again, worrying about prioritization. Also, awareness of time pressure might be slightly alleviated because the procedures take away some of that burden: the procedures are written and practiced in such a way that if the operators move at reasonable pace through the procedures then plant damage will be avoided. In EBAOPs, this is also generally true, but only if the diagnosis is correct and done in a timely manner.

While the Halden data are a very new and thorough source for crew time variation, the fact that EBAOPs are used makes it an unlikely candidate for updating the HCR/ORE

curves. Another data source, such as the OPERAS[35] or HERA[9] database, might be better suited for this task.

5.3.2 The EPRI HRA Calculator and Special Circumstances

As demonstrated by the time pressure/information load examples, both methods presented in the EPRI HRA Calculator are constructed to address the average conditions only, and do not lend themselves to quantification of specific evolutions of a given scenario. The Halden data addressed two types of special circumstances: masked scenarios and scenarios with high time pressure/information load. Masked scenarios are generally specific enough (i.e., they require an exact low-frequency event to be coupled with a specific low-frequency hardware failure) that they are not considered to be important contributors to human error during normal accident scenarios using SBEOPs,[25] as seen by the PWR ATHEANA trial in Section 2.4. However, they become very important when diagnosis is necessary or when key hardware failures become more probable. During shutdown the plant is in an abnormal state: often random systems are unavailable due to maintenance. Also, during shutdown, significant diagnosis is necessary because EBAOPs are used instead of SBEOPs. Therefore, masked scenarios are potentially significant for a shutdown HRA. Another general area where masked scenarios may be significant is for external events (i.e., fire, earthquake, flood, etc.) because the probability of hardware failures increases significantly.

The HCR/ORE data used to construct the curves only incorporate some PSFs (i.e., differences between crew), but it does not explicitly incorporate masking effects. Furthermore, in masking situations, diagnosis becomes important, which would also suggest that the HCR/ORE method is inappropriate for quantification where masking is important.

Likewise, the CBDT method cannot adequately capture masking effects.[25,34] Quantification of the Halden masking event led to a failure probability of unity. While this may be reasonable given the frequentist failure probability from the data was .86, a probability of unity is not reasonable for every masked scenario. For example, both of the EFCs in the ATHEANA PWR trial (Section 2.4) were masking events – only one of them had an effect on operator success. Therefore, the CBDT can only account for masking effects in a very conservative manner.

The Halden experiments show that time pressure and information load have an impact on human performance. These PSFs are incorporated into HCR/ORE calculations because the ORE data were taken from different scenarios including varying levels of time pressure and information load. The CBDT can deal with variations in information load (work load) and time pressure (work load and possibly recovery), but this is done in a very round-about, generic way requiring override values for high workload or other such adjustments.

5.3.3 ATHEANA and Special Circumstances

From the results of the time pressure/information load trial, it is clear that both PSFs had some effect on operator performance, but this effect was not uniform across crews. However, this set of data clearly represents the type of distribution to be expected from experts taking into account "aleatory" PSFs. For example, Crew G took longer than 5 minutes in 3 of the 4 trials (taking significantly over 5 minutes for 2 of the 4 trials). In two trials (8:55 and 11:43) the crew prioritized the pressure relief valve over boration. In these trials, the crew closed the valve only after several failed attempts. In one trial (8:55), the crew did not notice that boration needed to happen until 5 minutes and 32 seconds into the event. In another trial (11:43), the crew realized early that boration was necessary, but deliberately waited until the pressure relief valve was open because they feared otherwise the Boron will boil away to the condensing pool (a misdiagnosis). This crew is a clear example of a "bad" crew that experts quantifying this event would have to appropriately account for.

The masking scenarios demonstrate that masking can have a very strong effect on HEPs, thus lending credibility to ATHEANA's claim that significant error forcing contexts can and should be identified. Figure 26 shows the time variability between crews for the nominal (low masking), medium, and high masking cases. These data could bring insight into what a strong EFC is and how crew response might vary given a strong EFC. In this case, both the medium and high masking cases would be considered EFCs. For both EFCs, only one crew out of seven successfully completed the task. This

information could be used as the starting point for a statistical approach at quantifying $P(UA_j|EFC_i,S)$ in the ATHEANA formulation. This will be further described in the next Section.

6 Conclusions

At the beginning of this endeavor, we sought to "explore the capabilities and limitations of HRA methods, understand the methods, what they offer, how the differ and how they can be used."[18] We also asked: how can current data-gathering efforts become integrated with the current methods?[18]

Sections 2 and 3 contained a description of ATHEANA and the EPRI Calculator, Section 4 addressed how the two methods differ, and Section 5 addressed what each method offers and how it can be used. The discussion presented here will first explore and summarize the capabilities and limitations of both quantification methods, and then address the data integration question.

6.1 Capabilities and Limitations: The EPRI HRA Calculator

To address the capabilities and limitations of the EPRI Calculator, we shall examine the HCR/ORE and CDBT methods separately and then look at the interaction between the two methods.

6.1.1 HCR/ORE

The HCR/ORE method approaches quantification of p_C, the probability of failure to initiate, from a statistical angle. Realizing that available time is generally a dominating PSF, this method relies upon response time distributions taken from ORE data and time available to quantify p_C for a given HI. Rather than dissecting the HI into specific failure modes, the HCR/ORE method approaches p_c from a procedural-level examination of HIs (i.e., failure of Boron initiation given an ATWS is not examined to see what could cause a crew to fail to initiate the proper actions for successful boration). It is this high level approach that is at the heart of both the capabilities and limitations of the HCR/ORE method.

The ORE curves, which form the bases of the HCR/ORE method, combine aleatory and epistemic uncertainties. The only categorization, or epistemic uncertainty, that is explicitly captured by the ORE curves is cognition via the various cue-response structures. Otherwise, the two types of uncertainties are inseparable. To attempt to extract a probability for a specific context from the ORE curves would be impossible. Therefore, quantification using this method does not require – and does not support – further decomposition of a HI into specific sequence evolutions or failure modes.

In the HCR/ORE method, this kind of resolution – being able to distinguish between failure modes and specific contexts within a given HI – is traded off for low resource intensity. Given that most plants do not have a dedicated HRA expert on staff,

the Calculator does not require this level of expertise in order to be useful. Any PRA analyst with proper training on the Calculator can use this method to complete a basic HRA that is consistent, reproducible, and transparent. The results are transparent because every assumption and parameter is well documented, and a deterministic relationship exists between the assessed parameters and the HEP (see Equation 3). Furthermore, this method can be applied entirely "in house" to complete the plant's HRA. This is important to the transparency of the HRA, and will be elaborated upon in the next section on ATHEANA.

For all the benefits of low resource intensity, the tradeoff is, again, a limited scope. Lack of resolution also means that quantification of special circumstances, where specific failure modes must be addressed (i.e., determining the effect of time pressure/information load, or masking), is outside the scope of the HCR/ORE quantification method. Also, because the curves ignore failures due to misdiagnosis, the HCR/ORE is also not suitable for quantification of events which require extensive diagnosis, or have a large available time with respect to the median response time. This is why the HCR/ORE method is not the only method in the EPRI HRA Calculator – the cause-based decision tree method exists as well.

Other capabilities and limitations of the HCR/ORE method arise from its dependence on plant-specific simulator data. Some of these capabilities include strategy monitoring and a formalized link between plant-specific data ("as built, as operated") and quantification via $T_{1/2}$ estimation. For a given HI, $T_{1/2}$ ideally is taken from walkthroughs

and simulator data (usually from observing training scenarios). By virtue of obtaining plant-specific data points for a given HI, it becomes possible to use this data to observe the response time impact of different strategies used by different crews. Depending on the response time, this data can then flag good and bad practices to aid in training and procedure improvement.[15] Furthermore, as data becomes available over the lifetime of a plant (i.e., from more training runs), $T_{1/2}$ can be refined and the consistency of overall operator performance can be assessed (perhaps as a proxy for assessing quality of training). Thus, the HCR/ORE method can also be a useful tool for training assessment and improvement.

Finally, this method rests squarely upon acquired simulator data. Validation of this method, then, requires assessing how "good" simulations are (generally routine training scenarios) are in modeling reality. Here it may be possible to turn to other industries – particularly the military (air force) and the aeronautics industry – to help with this assessment.[36]

6.1.2 CBDT

The Cause Based Decision Tree method is rooted in THERP and is based on the identification of eight possible failure mechanisms for any given HI. Each failure mode comes with its own decision tree to aid the analyst in determining the p_C contribution due to that mechanism. These trees are based on two types of cognitive failures: failure to obtain required information to make the decision and failure to decide correctly. Each

decision tree is predetermined, with each node of the tree corresponding to a performance shaping factor that has been judged to be an important contributing factor for that given mechanism. The total p_C is then the summation of the p_C for each failure mechanism. The contribution of each failure mechanism is obtained by selecting the branch that most accurately describes the plant conditions for that scenario and using the default THERP values given for that branch. This probability is adjusted for recovery based on opportunity for review.

Like the HCR/ORE method, CBDT has all the benefits of being consistent, transparent and easy to use. However, as a method that attempts to deal with diagnosis errors under special circumstances, CBDT is very coarse. This method can decompose an HFE into different failure modes, however, this is done in a very generic way and does not necessarily address all the important PSFs. Furthermore, the probabilities associated with the EOC failure mechanisms (trees d and h) take their values from THERP, which maybe outdated and overly conservative (as seen through the example in section 5). The EPRI HRA Calculator recommends using override values when possible, however, there is no guidance on how to assess these probabilities. Using override probabilities would take away the transparency and consistency which are the strengths of the Calculator.

6.2 Capabilities and Limitations: ATHEANA

ATHEANA decomposes a HFE into specific error forcing contexts to identify and quantify human error types that arise from unusual circumstances. This EFC

identification involves an elaborate search process geared towards understanding how an operator might believe his erroneous actions are correct. This search process dives deeply into the specific contexts of a given accident scenario in order to find ways operators can end up in regimes that are characteristic of severe accidents (i.e., plant behavior is out of the expected range, plant behavior is not well understood by operators, plant procedures are not helpful/appropriate or the actual plant state is not recognized by operators). This method strives to be all-inclusive with respect to various PSFs, and is not limited to the examination of certain PSFs. The quantification process, however, has to reduce the identified EFC into a few PSFs that are considered to be most important in creating that EFC. ATHEANA does not provide specific guidance on how this reduction should be done.

While in theory ATHEANA can quantify all types of HEFs, it is not intended to quantifying ordinary slips and lapses because of the recovery-based screening process. Quantification in cases where the probability of recovery is determined to be "high," which is the usual case for slips and lapses, is foregone unless that particular slip/lapse can then create an EFC and force a mistake.

By addressing epistemic and aleatory PSFs separately, the consensus-based expert elicitation process allows ATHEANA to incorporate the effects of PSFs on a HEP without having to quantify the probability of these PSFs occurring.

Many of the limitations of ATHEANA arise from its reliance on an expert opinion-based quantification process. There are several problems associated with expert opinion elicitation for HEP – these probabilities are often biased, inconsistent, opaque, and non-reproducible. To some degree, these effects are mitigated by employing a highly structured, *consensus-based*, expert opinion elicitation process.[23] This type of expert opinion elicitation, however, is highly resource intensive.

Part of identifying EFCs includes running simulator trials to test the strength of these EFCs. These simulator trials can bring significant insight to the training personnel on how their operators can actually fail. Unlike the HCR/ORE simulations (for $T_{1/2}$), ATHEANA requires a special simulation to be run rather than just observing simulations from regular training sessions. The downside to these simulator tests is the resource intensity – in addition to the normal resource burden of simulator trials (i.e., PRA analysts need to take time to observe the scenario and process it), new contexts must be developed and simulated, and extra crew time (beyond normal training) is also required. Furthermore, because knowledge-based errors are so dependent on the particular operator's knowledge level and cognitive approach, EFC simulations which do not result in a failure do not necessary "prove" that the EFC is not strong. As we saw from the Halden masking trials, EFCs do not have a binary effect – all crews fail or none fail. Therefore, the usefulness of one simulation trial in cases that do not produce a failure is reduced.

One last note on ATHEANA: it has been criticized as being highly resource intensive. The effect of this characteristic extends beyond pure economics to impact the transparency of the HRA. As a participant of an ATHEANA pilot program pointed out, a plant's "tribal knowledge" is important to the quality of the HRA and how it is used. That is, ATHEANA requires many experts (including an HRA expert), which means that a plant will most likely turn to a contractor to do the work rather than doing it "in-house." This affects the transparency of the HRA because it becomes substantially more difficult to know what exactly went into the HEP and what problems arose.

6.3 Incorporating Data

Simulator experiments are a relatively low-cost method of obtaining valuable human performance data. Unfortunately, "methods and data activities sometime run independently."[18] To get maximum benefit, the two sectors must be integrate. How, then, can data and feedback be incorporated into these methods? What data is useful?

6.3.1 EPRI HRA Calculator

For EPRI, the HCR/ORE curves are a primary candidate for incorporation of data into the HRA model. Furthermore, because of the large variation in the standard deviations for a give ORE curve (Table 6), it would be quite useful to gather and incorporate more data into the HCR/ORE curves to gain better resolution on the sigmas. For this update, data from Halden cannot be directly used because Halden does not use

symptom-base operating proceduress. However, the design of the Halden experiments can be used because it produced results that show important differences in crew performance under certain conditions. A cautionary note for using the Halden experiment design: for TRCs to be meaningful, "tricky" (i.e., time pressure, information load, masking, etc.) situations must be included only at the frequency they would occur in plants; otherwise the curve will be skewed.

Another way to incorporate data into the EPRI Calculator would be to use qualitative insights from Halden-like experiments (using symptom-based operating procedures) to determine the reasonableness of the THERP values for the CBDT, especially with respect to misdiagnosis failures and other special circumstances. For example, in the Halden masking scenario, 2 of 14 masked trials were successful (Figure 26). This raises the question: is the THERP HEP value of 1 reasonable? In the time pressure/information load scenario, the CBDT found no difference between the high workload case and the nominal case, while the Halden data (Table 5) suggest there is a difference. By examining this type of simulator data, if the THERP values are found to be significantly different, then EPRI may consider updating the THERP values using Bayesian updating or expert opinion, or providing specific guidance to the plants themselves on how to assess/when to use override values for CBDT. For shutdown and external events, where event-based operating procedures are used, the Halden data itself can provide a check on the THERP values. Finally, from a qualitative perspective, the Halden scenarios can help analysts better apply CBDT by providing an example of what the failure modes and PSFs "look" like.

6.3.2 ATHEANA

ATHEANA uses expert opinion elicitation for quantifying HEPs, and so, in the current ATHEANA framework, simulator data can only be incorporated qualitatively. This can be done through the expert calibration process. In the ATHEANA quantification process, the EFCs are determined, but are limited to a few important PSFs and hardware failures. The Halden experiments are geared towards testing the failure contribution due to specific PSFs (or, more accurately, specific categories of EFCs, like masking or time pressure/information load). When quantifying the error probability for a given EFC, it might be useful to review the experimental results for a given category of EFCs (i.e., masking) as a way to calibrate experts to the effects of a general category of EFC and the variation in response.

This method of data incorporation is not ideal. The EFCs are very situation-specific, and so presenting experts with information from a generalization of that EFC may be misleading. For example, Figure 26 would be of the type of data that would serve to "calibrate" the experts for masking-type EFCs. Both EFCs examined in the ATHEANA pilot in Section 2.4 (MLOCA) are examples of masking. However, after the simulator trial, one masking EFC was found to be considerably stronger than the other. Therefore, the failure probabilities associated with each EFC cannot be calibrated to the same point without some sort of judgment being made with respect to the strength of the EFC in comparison to the experimental scenario. This might complicate the process and overwhelm the experts with too much information.

6.4 Final Remarks

The goal of comparing different HRA models is not necessarily convergence on one universal model, but, rather, it is to know what the state of the art is and what it is capable of.[18] To assess the state-of-the-art, the *context* of the HRA must be examined – what an HRA "can" and "can't" do is practically limited by a cost-benefit trade off between resources required and resolution needed. Therefore, the applicability or "goodness" of a specific approach at HEP quantification should be examined in an application-specific context, not just generally. Here we will summarize how ATHEANA and the EPRI HRA Calculator fulfill the requirements for an advanced HRA method, and then take a look at some of the application-specific contexts to illustrate the "goodness" of each approach.

Both the EPRI HRA Calculator and ATHEANA, at least nominally, fulfill the requirements for an advanced HRA in that they: yield transparent results, are not overly resource intensive, incorporate some aspect of the effects of organizational factors, give special attention to cognitive errors, provide a well-defined quantification procedure and incorporate data and feedback into the model. However, because of the resource/resolution tradeoff, neither method addresses all these points completely. Rather, their differing objectives dictate which of these requirements are most thoroughly addressed. For example, the EPRI HRA Calculator, created to enable a general PRA analyst to carry out a consistent HRA without significant outside resources, pays special attention to ensuring transparent results, having low resource intensity, providing a well-

defined quantification procedure and quantitatively incorporating data and feedback into the model. ATHEANA, on the other hand, strives to identify and quantify failures that involve specific "tricky" contexts and yield high HEPs. To carry out this objective, ATHEANA spends its resources on incorporating effects of organizational factors, giving special attention to cognitive errors, and qualitatively incorporating data and feedback into the process.

Because of their differing objectives, these two methods lend themselves best to different applications. The main value of a PRA is that it allows decision makers to allocate safety-improvement resources in the most effective manner by ranking risk contributions. In this case, the absolute accuracy of the HEP is not as important as the relative accuracy – being able to model a wide range of events easily and consistently is crucial. In this case, the EPRI HRA Calculator would be an appropriate choice for a first-pass PRA. ATHEANA, on the other hand, is generally too resource intensive to be suitable for modeling all HFEs in a full scope PRA, and is probably not the best choice for a first pass PRA. However, if this "first pass" reveals troublesome HFEs that involve cognitive stumbling blocks for the operators beyond what the CBDT method is capable of handling, one might be inclined to switch to ATHEANA for a more in-depth investigation. Shut down and external event HRAs especially might benefit from ATHEANA, for these are situations where troublesome contexts become much more common because the plant is in an unusual state.

Applications such as the significance determination process and risk-informed decision making generally have a very specific scope and context that require an in-depth analysis. These applications are also very number-oriented for regulatory purposes. These characteristics suggest that ATHEANA is a good candidate for these analyses because it is has a greater capability to dive deeply into a HFE, dissect it according to specific accident evolutions (contexts), and then reflect all that information in the resulting HEP. In these situations, the EPRI HRA Calculator is not without its uses. While for some troublesome situations the Calculator may be overly conservative (i.e., assigning a probability of unity for some masking scenarios as seen in Section 5), for risk-informed decision making, sometimes the Calculator is "good enough." For instance, in using screening methods in risk-informed decision making (see the Reinert-Apostolakis method[37] and NUREG-1764[38]), the coarse number the EPRI HRA Calculator yields may be sufficient for the analyst to determine whether or not a given HFE can have an effect on the overall decision. In this case, the EPRI Calculator could act as a screening method, allowing the analyst to scrutinize only the HFEs which can impact the decision at hand. And so, while each method has the same general goal (calculate a HEP) and considers the same inputs (PSFs, accident sequence, and cognitive factors), they have very different objectives and scopes. Consequently, neither the EPRI HRA Calculator, nor the ATHEANA method is "best," but rather, the better method is some combination of the two.

7 References

[1] Swain, A.D. and H.E. Guttmann, "Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications," NUREG/CR-1278, SAND80-0200, Sandia National Laboratories, August 1983.

[2] Swain, A.D., "Accident Sequence Evaluation Program Human Reliability Analysis Procedure," NUREG/CR-4772, SAND86-1996, Sandia National Laboratories, February 1987.

[3] Williams J.C., "A Databased Method for Assessing and Reducing Human Error to Improve Operational Performance," in Proceeding of the IEEE Fourth Conference on Human Factors and Power Plants, Monterey, CA, June 59, 1988, pp.436-450.

[4] Park, J. and W. Jung, "OPERA – a Human Performance Database Under Simulated Emergencies of Nuclear Power Plants." *Reliability Engineering and System Safety* Article In Press.

[5] Gertman, D.I., L.N. Haney and N.O. Siu, "Representing Context, Cognition, and Crew Performance in a Shutdown Risk Assessment." *Reliability Engineering and System Safety 52* (1996), pp. 261-278.

[6] Chen, S.H., A.A. Dykes, J.W. Stetkar and D.C. Bley, "Quantification of Human Error Rates Using a SLIM-Based Approach," IEEE Fourth Conference on Human Factors and Power Plants, Monterey, CA, June 59, 1988.

[7] Embrey, D.E., Humphreys, P., Rose, E.A., Kirwan, B., and Rea, K, "SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgment." (Vols. I & II), NUREG/CR-3518, Brookhaven National Laboratory, Upton, NY, 1984.

[8] Gertman, D.I., H.S. Blackman, J. Byers, L. Haney, C. Smith and J. Marble, "The SPAR-H Method." NUREG/CR-6883, US Nuclear Regulatory Commission, Washington, D.C., 2005

[9] Lois, E., D.I. Gertman and D. Dudenhoeffer. "Framework for the Human Event Repository and Analysis (HERA) System and its use to Quantify Human Actions." New Orleans, LA; Probabilistic Safety Assessment and Management Conference (PSAM 8), sponsored by the International Association for Probabilistic Safety Assessment and Management (http://www.iapsam.org/), May 14-18th, 2006.

[10] Baumont, G., F. Menage, J.R. Schneiter, A. Spurgin, A. Vogel. "Quantifying Human and Organizational Factors in Accident Management Using Decision Trees: The HORAAM Method," *Reliability Engineering and Safety System 70* (2000), pp. 113-124.

[11] B. Kirwan, "The development of a nuclear chemical plant human reliability management approach: HRMS and JHEDI." *Reliability Engineering and System Safety* 56 (1997), pp. 107-133.

[12] Hollnagel, E. "Cognitive Reliability and Error Analysis Method CREAM." New York: Elsevier; 1998.

[13] Bieder, C., et. al. "MERMOS: EDF's New Advanced HRA Method," (1998) in Proceedings of Probabilistic Safety Assessment and Management Conference

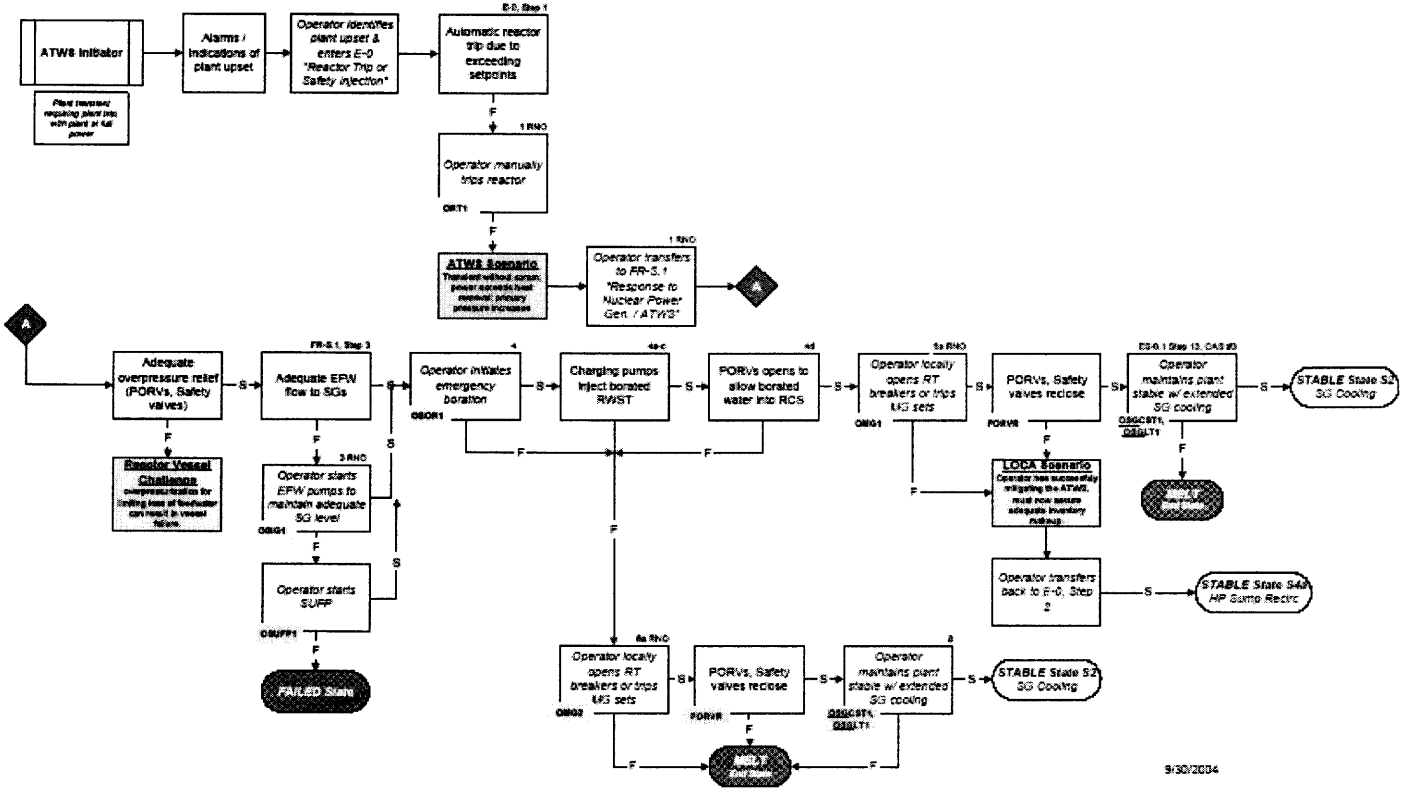
(PSAM 4), sponsored by the International Association for Probabilistic Safety Assessment and Management (http://www.iapsam.org/), Vol. I, pp. 129-134.
[14] "Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA)," NUREG-1624, Draft Report, US Nuclear Regulatory Commission, Washington, D.C., May 2000.
[15] "An Approach to the Analysis of Operator Actions in Probabilistic Risk Assessment;" EPRI TR100259, Final Report, June 1992
[16] "SHARP1 – A Revised Systematic Human Action Reliability Procedure;" EPRI TR-101711, Final Report, December 1992.
[17] "SHARP1 – A Revised Systematic Human Action Reliability Procedure;" EPRI NP-7183-SL, Interim Report, December 1990.
[18] Bye, A. "Minutes, Workshop Meeting on HRA Method Investigation." New Orleans, LA; Probabilistic Safety Assessment and Management Conference (PSAM 8), sponsored by the International Association for Probabilistic Safety Assessment and Management (http://www.iapsam.org/), May 13-14th, 2006.
[19] Laumann, K., et. al. "The Task Complexity Experiment 2003/2004." HWR-758, OECD Halden Reactor Project.
[20] W.D. Jung and J.W. Kim, "Analysis of limitations on human reliability analysis in nuclear power plants and development of requirements for an advanced method." *J Kor Inst Ind Safety* 14 (1999), pp. 178-199.
[21] "ATHEANA User's Guide," NUREG-1842, Draft Copy, US Nuclear Regulatory Commission, Washington, D.C., 2006.
[22] "An Empirical Investigation of Operator Performance in Cognitively Demanding Simulated Emergencies," NUREG/CR-6208, US Nuclear Regulatory Commission, Washington, D.C., July 1994.
[23] Forester, J., D. Bley, S. Cooper, E. Lois, N. Siu, A. Kolaczkowski, and J. Wreathall, "Expert elicitation approach for performing ATHEANA quantification." *Reliability Engineering and System Safety* 83 (2004), pp. 207-220.
[24] Kolaczkowski, A., et. al. "Example Application of ATHEANA: Pressurized Thermal Shock (PTS) Analyses," a Presentation to the Advisory Committee on Reactor Safeguard, PRA and Human Factors Subcommittees, Rockville, MD, June 28, 2006.
[25] Rau, L., Personal Communications (July – August 2006).
[26] Poloski, J., J. Knudsen. *Standardized Plant Analysis Risk Model (SPAR)* of a PWR. Idaho National Laboratory, 2004.
[27] Orvis, D.D., A.J. Spurgin, C.T. Laio, J.P. Spurgin, and P. Moieni. Operator Reliability Experiments at Maanshan: Final Report. San Diego, CA: Accident Prevention Group, APG Report #8910, 1989.
[28] Moieni, P., A.J. Spurgin, "Advances in Human Reliability Analysis Methodology." *Reliability Engineering and System Safety* 44 (2004), pp. 27-66.
[29] Hannaman, G.W., A.J. Spurgin, and Y.D. Lukic, Human Cognitive Reliability Model for PRA Analysis, NUS-4531, Electric Power Research Institute, 1984.
[30] Spurgin, A.J., et al., Operation Reliability Experiments Using Power Plant Simulations, EPRI NP-6937, 1990.
[31] Julius, J. and J. Grobbelaar. "Integrating Human Reliability Analysis Approaches in the EPRI HRA Calculator." New Orleans, LA; Probabilistic Safety Assessment

and Management Conference (PSAM 8), sponsored by the International Association for Probabilistic Safety Assessment and Management (http://www.iapsam.org/), May 14-18th, 2006.
[32] Jeff Julius presentation at ACRS meeting June 28, 2006.
[33] Apostolakis, G.E., "A Commentary on Model Uncertainty," in: *Proceedings of Workshop on Model Uncertainty: Its Characterization and Quantification*, A. Mosleh, N. Siu, C. Smidts, and C. Lui, Eds., Annapolis, MD, October 20-22, 1993, Center for Reliability Engineering, University of Maryland, College Park, MD, 1995 (also published as Report NUREG/CP-0138, US Nuclear Regulatory Commission, Washington, DC, 1994).
[34] Julius, J., Scientech, Personal communication, August 2006.
[35] Spurgin, A.J. and J.P. Spurgin, 1994, "A Data Collection and Analysis System for Used with a Power Plant Simulator," Institute of Mechanical Engineers Seminar, "Achieving Efficiency Through Personal Training – The Nuclear and Safety Regulated Industries," London, England.
[36] Schiflett, S., L. Elliott, E. Salas, M. Coovert. Scaled Worlds: Development, Validation and Applications, Ashgate Publishers, 2004.
[37] Reinert, J.M. and G.E. Apostolakis, Including Model Uncertainty in Risk-Informed Decision Making." *Annals of Nuclear Energy* 33 (March 2006), pp. 354-369.
[38] "Guidance for the Review of Changes to Human Actions: Final Report." NUREG-1764, U.S. Nuclear Regulatory Commission, Washington, DC, February 2004.

Appendix A: ATWS Event Sequence Diagram for a PWR

Event Sequence Diagram: Full Power Model

FIGURE 5.8-1 ATWS Initiator ESD



Appendix B: EPRI Calculator Auto Generated Report for PWR Boron Injection

OBOR - Operator Initiates Emergency Boration[25]

BASIC EVENT: HH.OBOR1.FA Initiate Emergency Boration (ATWS)

Summary

TASK: The operator must initiate emergency boration of the RCS.
CONTEXT: This action is included in top event OBOR in the ATWS tree. It is only asked for ATWS sequences where initial attempts to manually trip the reactor (ORT) failed.
SUCCESS: Success requires at least one charging pump supplying borated water to the RCS.

Diagnosis

DIAGNOSIS: At step 4 in FR-S.1 the operator is instructed to initiate emergency boration. This step is classified as an immediate action step.

Execution

EXECUTION: The 3 critical steps to initiating boration within FR-S.1 are: FR-S.1 Step 4c.1, Place CS-FK-121 in manual and charge at the maximum rate. FR-S.1 Step 4c.2, Align CCP suction to the RWST. FR-S.1 Step 4c.3, Isolate the CCPs from the VCT.
EXECUTION RECOVERY: FR-S.1 Step 7e, Continue boration to obtain adequate shutdown margin

Performance Shaping Factors

TIME REQUIRED (T1/2): 120 sec. From observations of the operators performing this action during simulator exercises, diagnosis errors for initiating emergency boration is negligible. During the three exercises conducted on the Seabrook simulator, the response times for beginning emergency boration were 33, 46, and 62 seconds. In an observation 2/4/2005 GTK. Step 4 FRS.1 was reached in approx. 2 min. Thus, 2 min is used as a best estimate.
TIME AVAILABLE (Tsw): 10 min. Long-term shutdown is assumed to be required within 10 minutes following the initiating event. The 10-minute time limit is conservatively based on the initial time required for the reactor to become critical again because of plant cooldown. Westinghouse Owner's Group, "Background Information for Emergency Response Guideline FR-S.1 "Response to ATWS," HP-Rev.1.
TIME DELAY (Tdelay): 0
TIME Manipulation (Tm): 120 sec, estimated time for board manipulations in Step 4c.
STRESS: HIGH based on the seriousness of an ATWS event without success of manual reactor trip.

Quantification

HEP = 2.9E-02 (5) CB
Cause Based HEP: Pcog = 3.00E-03 Pexe = 2.61E-0 Ptotal = 2.91E-02
Time Based HEP: Pcog = 2.64E-04 Pexe = 2.61E-0 Ptotal = 2.64E-02