NUREG/CR-4862
Vol. 2

Cognitive Environment Simulation: An Artificial Intelligence System for Human Performance Assessment

Modeling Human Intention Formation

Prepared by D. D. Woods, E. M. Roth, H. Pople, Jr.

Westinghouse Electric Corporation

**Prepared for
U.S. Nuclear Regulatory
Commission**

NUREG/CR-4862
Vol. 2
RX

Cognitive Environment Simulation: An Artificial Intelligence System for Human Performance Assessment

Modeling Human Intention Formation

Manuscript Completed: September 1987
Date Published: November 1987

Prepared by
D. D. Woods, E. M. Roth, H. Pople, Jr.*

Westinghouse Electric Corporation
Research and Development Center
1310 Beulah Road
Pittsburgh, PA 15235

*University of Pittsburgh and Seer Systems
Pittsburgh, PA 15235

Prepared for
Division of Reactor and Plant Systems
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555
NRC FIN D1167

Abstract

This report documents the results of Phase II of a three phase research program to develop and validate improved methods to model the cognitive behavior of nuclear power plant (NPP) personnel. In Phase II a dynamic simulation capability for modeling how people form intentions to act in NPP emergency situations was developed based on techniques from artificial intelligence. This modeling tool, Cognitive Environment Simulation or CES, simulates the cognitive processes that determine situation assessment and intention formation. It can be used to investigate analytically what situations and factors lead to intention failures, what actions follow from intention failures (e.g., errors of omission, errors of commission, common mode errors), the ability to recover from errors or additional machine failures, and the effects of changes in the NPP person-machine system.

The Cognitive Reliability Assessment Technique (or CREATE) was also developed in Phase II to specify how CES can be used to enhance the measurement of the human contribution to risk in probabilistic risk assessment (PRA) studies.

The results are reported in three self-contained volumes that describe the research from different perspectives. Volume 1 provides an overview of both CES and CREATE. Volume 2 gives a detailed description of the structure and content of the CES modeling environment and is intended for those who want to know how CES models successful and erroneous intention formation. Volume 3 describes the CREATE methodology for using CES to provide enhanced human reliability estimates. Volume 3 is intended for those who are interested in how the modeling capabilities of CES can be utilized in human reliability assessment and PRA.

Table of Contents

List of Figures

List of Tables

Acknowledgments

The authors would like to express their appreciation to Thomas G. Ryan, the NRC project officer, for the guidance and support he provided throughout the model development program.

We would also like to thank Michael Coombs, Allen Newell, Richard Pew, Jon Young, and John Wreathall, who participated in an NRC sponsored workshop to review the model development work, for their helpful comments.

We are grateful to the many colleagues who discussed with us how a model of intention formation could aid HRA and PRA, especially V. De Keyser, O. Svenson, J. Rasmussen, B. Kirwan.

Special thanks are due to James Easter, Glenn Elias and Michael Mauldin for helping to keep the model development on-track to capture the nuclear power plant as a problem solving world.

1. Introduction

This report documents the results of Phase II of a three phase research program sponsored by the U. S. Nuclear Regulatory Commission to develop and validate improved methods to model the cognitive behavior of nuclear power plant (NPP) personnel during emergency operations. In Phase II a model of how people form intentions to act in NPP emergency situations (Cognitive Environment Simulation or CES) was developed using artificial intelligence (AI) techniques. A methodology for using the model to enhance measurement of the human contribution to risk in probabilistic risk assessment (PRA) studies (Cognitive Reliability Assessment Technique or CREATE) was also developed.

This volume (2) of the report describes the structure and content of the CES cognitive model. It is intended for those who want to know about how CES models successful and erroneous human intention formation. Volume 1 provides an overview of CES and CREATE. Volume 3 describes the CREATE methodology. It outlines the steps involved in using CES as part of HRA/PRA studies, and it describes how CES can be used to better estimate human reliability. Volume 3 is intended for those who are interested in how the modeling capabilities of CES can be utilized in HRA and PRA.

1.1 The Importance of Modeling Operator Cognitive Activity for Human Reliability Assessment

The quality of human performance has been shown to be a substantial contributor to nuclear power plant safety. Some PRA studies have found that approximately one half of the public risk from reactor accidents can be related to human error (Levine and Rasmussen, 1984; Joksimovich, 1984). Studies of NPP operation and maintenance indicate from 30% to 80% of actual incidents in nuclear power plants involve significant human contribution (Trager, 1985). The analytical and empirical records clearly show that the human contribution to total safety system performance is at least as large as that of hardware reliability (Joksimovich, 1984).

A significant factor in determining human action under emergency conditions is *intention formation* -- deciding on what actions to perform.[1] Errors of intention are an important element of overall human contribution to risk, and the PRA community has recognized the need for more effective ways to capture this component of human error (Levine and Rasmussen, 1984).

---

[1] This is contrasted with *execution of intentions* -- carrying out the sequence of actions decided upon.

The U.S. Nuclear Regulatory Commission has embarked upon a program of research to build a computer model of human intention formation (how people decide on what actions are appropriate in a particular situation) in order to better predict and reduce the human contribution to risk in NPPs. The model simulates likely human responses and failure modes under different accident conditions, comparable to the analytic tools available for modeling physical processes in the plant.

This research program consists of three phases. Phase I (completed in April of 1986) was a feasibility study which determined that it is practical to build such a cognitive model based on techniques from artificial intelligence (AI) to provide useful input to human reliability analysis and probabilistic risk assessment (the results of the assessment are reported in NUREG/CR-4532). The feasibility study identified a specific AI software system which could serve as a vehicle for model development.

Phase II of the research project focused on model development and application to HRA based on the approach identified in Phase I. Specifically:

1. A model of how people form intentions to act in emergency operations in NPPs was developed using AI techniques. The model, called Cognitive Environment Simulation or CES, is the first analytic computer simulation tool which can be used to explore human intention formation in the same way that reactor codes are used to model thermodynamic processes in the plant.

2. A methodology, called Cognitive Reliability Assessment Technique or CREATE, was developed which specifies how this capability can be used to enhance measurement of the human contribution to risk in PRA studies.

An additional phase of the research project is planned whose objective is to conduct field evaluation and validation of the CES cognitive model and the CREATE methodology.

1.2 The Role of Modeling of Human Intention Failures in Risk Analysis

Model development addressed one part of human behavior: human intention formation (deciding what to do) and erroneous intentions to act. This scope was chosen, first, because models and techniques are already available to assess the form and likelihood of execution errors in human reliability studies (e.g., Reason & Mycielska, 1982; Swain & Guttman, 1983). A second reason

for selecting this scope is because erroneous intentions are a potent source of human related common mode failures which can have a profound impact on risk – as actual accidents such as Three Mile Island and Chernobyl have amply demonstrated. Intentions to act are formed based on reasoning processes. The scientific disciplines that study these processes are called cognitive sciences or mind sciences and include a variety of fields such as cognitive psychology and artificial intelligence. Models of these processes are called "cognitive models."

In Phase II a computer simulation of intention formation in emergency operations was developed. This system, Cognitive Environment Simulation or CES, is the first analytic computer simulation tool that can be used to model human intention formation in the same way that reactor codes are used to model thermodynamic processes in the plant.

CES is a simulation of cognitive processes that allows exploration of plausible human responses in different emergency situations. It can be used to identify what are difficult problem-solving situations, given the available problem-solving resources (e.g., specific procedural guidance, operator knowledge, person-machine interfaces). By simulating the cognitive processes that determine situation assessment and intention formation, it provides the capability to establish analytically how people are likely to respond, given that these situations arise. This means one can investigate

- what situations and factors lead to intention failures,
- the form of the intention failure,
- the consequences of an intention failure including,
    - what actions will not be attempted -- errors of omission,
    - what actions the intention failure will lead to – *commission errors* and *common mode failures*, that is, those leading to the failure of otherwise redundant and diverse systems due to misperception of plant state or another cognitive processing breakdown,
    - *error recovery* – whether the human intention failures or execution errors or failures of plant equipment to respond as demanded will be caught and recovery action taken (and information on the time until recovery),

- "improvised" action sequences that operators may take in different circumstances (responses other than the nominal response sequence in the procedure which hindsight suggests was most appropriate).

The ability of CES to predict errors of commission is particularly important since misapprehension of plant state by the operator can result in multiple actions which can have broad systemic effects. Intention failures are a major source of *human related common mode failures* -- multiple failures that are attributable to a common element (namely, the erroneous intention). Examples of this are cases where the situation is misperceived, and the operator deliberately decides it is appropriate to turn off multiple, otherwise redundant and diverse systems as occurred at Three Mile Island and Chernobyl. The PRA community generally recognizes the importance of identifying common mode failure points because they can have large and widespread effects on risk.

Because CES models the processes by which intentions to act are formed, it can be used, not only to find intention error prone points, but also to identify the sources of cognitive processing breakdowns and intention failures. This means that it can help to develop or evaluate error reduction strategies.

CES also provides an analytic tool for investigating the effects of changes in NPP person-machine systems including new instrumentation, computer-based displays, operator decision aids, procedure changes, training, multi-person or multi-facility (e.g., technical support center) problem solving styles. This means that proposed changes/enhancements to NPP person-machine systems can be analytically evaluated before they have been implemented.

CES, as a modeling environment, is a specific instance of an artificial intelligence problem solving system, EAGOL.[2] The EAGOL problem solving architecture embodies unique capabilities for reasoning in dynamic situations that include the possibility of multiple faults. CES uses these capabilities to capture the kinds of cognitive processes that contribute to intention formation.

Cognitive Reliability Assessment Technique (CREATE) is the method for using the capabilities of CES to better evaluate the potential for significant human errors in PRA analysis. In CREATE, CES is run on multiple

---

[2]EAGOL is a software system and proprietary product of Seer Systems. EAGOL builds on the conceptual framework of the CADUCEUS AI problem-solving system developed for medical problem-solving applications (Pople, 1985).

variants of accident sequences of interest. The variants are selected to represent parametric combinations of a plausible range of values along the dimensions that contribute to cognitive task complexity. The goal is to identify sets of minimum necessary and sufficient conditions (characteristics of the situation and/or the operator) that combine to produce intention failures with significant risk consequences. Once the range of plausible intention errors and the conditions under which they will arise are identified, a quantification procedure is used to assess the likelihood of these intention errors.

The CREATE methodology involves two main stages: a modeling stage where CES is used to find situations that can lead to intention failures and therefore to erroneous actions; and a systems analysis input stage where the results of the cognitive modeling are integrated into the overall systems analysis.

The main steps in the modeling stage are:

- Decide what NPP situations to investigate with CES and how these situations map into the CES simulation world,
- Set up CES to be able to run NPP situations,
- Run CES over a plausible range of demand and resource settings, given the analysis of this plant,
- Analyze CES behavior to identify the minimum conditions which produce intention failures and the actions that follow from an intention failure.

Because CES is a simulation code, it requires detailed and complete input to run and outputs specific predictions about human intentions. This means that using CES in the modeling stage ensures explicit consideration and detailed analysis of the factors that contribute to human intention errors.

The main steps in the systems analysis input stage are:

- Modify the systems analysis event/fault trees to reflect the effects of intention errors identified in the modeling stage.
- Employ a quantification procedure to assess the likelihood of these intention errors,

- Combine intention error estimates with execution error estimates.

Note that CES plays the same role in the CREATE methodology that simulation codes for physical plant processes play in reliability analyses of physical systems. In both cases we are dealing with complex, dynamic processes whose behavior is affected by too large a set of interacting factors to be tractable without a simulation. The modeling stage provides the backbone of the analysis in that it defines the critical elements to be aggregated and how they are to be aggregated. Frequency estimation techniques are then used to establish the probabilities to be aggregated.

1.3 Background for Model Development

This section briefly describes the background for the model development work carried out in Phase II including the goals to be satisfied, the behavioral science and NPP scopes to be addressed, and what activities are to be modeled. NUREG/CR-4532 contains a thorough discussion of these topics.

Objectives of Model Development.
The goal of the Phase II model development was to enhance the ability to predict human performance in NPPs, in particular, to enhance the ability:

- to predict the human contribution to risk in human reliability analysis (HRA) and probabilistic risk assessment (PRA);
- to identify situations prone to human error, particularly human related common mode errors and errors of commission;
- to understand the mechanisms that produce human error;
- based on increased knowledge about error mechanisms, to help develop error and risk reduction strategies;
- to predict the effects of changes in the NPP person-machine system (procedures, training, sensors, displays, operator aids) on human performance.

Intention Errors and Cognitive Processing.
Model development focused on one part of human behavior: human intention formation (deciding what to do) and erroneous intentions to act. Intentions to act are formed based on reasoning processes that determine how plant data are monitored, what situation assessments are formed, what explanations are built and what responses are judged appropriate to carry out under these

perceived circumstances. The scientific disciplines that study these processes are called cognitive sciences. Models of these processes are often called "cognitive models."

What is cognition? The word cognitive describes one approach to understanding human behavior which assumes that description, explanation, and prediction of observable human actions depends on understanding the chain of information processing or mental events that mediate between observable events in the world and human responses.

> Definition of Cognition: The cognitive approach asserts that human performance varies because of differences in the knowledge that a person or team of people possess (both the form and the content), in the activation of that knowledge, and in the expression or use of knowledge.

How knowledge is activated and used is based fundamentally on an iterative cycle of data-driven activation of knowledge and knowledge-driven observation and action. An item in the world is noticed (e.g., an alarm) which triggers some knowledge (e.g., what the message means about changes in system state); this knowledge, in turn, leads to new observations or actions which trigger other knowledge, etc. Cognitive models differ in the particulars of how data activate knowledge and how activated knowledge leads to particular observations and actions in different contexts.

Scope.

The model development addresses the cognitive processes that affect successful and erroneous human intention formation in NPP emergencies. This area of human behavior includes what is sometimes called "rule-based" behavior and "knowledge-based" behavior up to the point of creative problem solving (Rasmussen, 1986).

Model development focused on one part of the NPP: operations during abnormal and emergency conditions, i.e., activities carried out by the emergency response system including the control room and branching out to the technical support center.

What Cognitive Activities Need to be Modeled?

An effective cognitive model must be able to capture the kinds of cognitive activities that occur in emergency operations in order to produce valid predictions that are relevant to NPPs. Let's call this target the basic competencies of the desired cognitive model. These competencies are kinds of

behavior or information processing the model must exhibit that reflect aspects of the processing that people carry out to meet the demands of problem solving during control room emergencies.

To build a model to do this we must know -- What kinds of problem solving situations occur in NPP emergency operations? What must people know and how must people use that knowledge to solve these problems? How do people actually respond in these types of problem solving situations? The answers to these questions come from current empirical and analytical results on the cognitive demands and activities that arise in emergency operations (these are described in Chapter 4 of NUREG/CR-4532).

There are four primary characteristics of the NPP world that determine the kinds of problem solving situations that can arise in emergency operations.

1. NPPs are composed of a large number of highly interactive parts and processes (systems, functions, goals).

2. Emergency operations occur in a dynamic, event-driven world where incidents unfold in time, and events can happen at indeterminate times during an incident.

3. There is uncertainty -- a demanded position indication may not reflect actual position or sensors can fail -- and there is risk -- possible outcomes can have large costs.

4. There is a high degree of automation which means that multiple agents (machine controllers, machine decision makers, and multiple people) are involved in the response to emergency incidents.

The result is that actual NPP emergency incidents are difficult because multiple, interacting events (machine and human failures) can and do occur in the face of uncertain evidence and risky choices.

To solve problems in a world with these characteristics, operators must know about the many parts and processes and their interrelationships. They must be able to use this knowledge in a changing situation to determine the state of the plant (sustained monitoring) and how to respond (e.g., take into account side effects). This is complicated by uncertainties in the available evidence and the likelihood of multiple faults. Because of the high workload, operators must make decisions about timesharing/scheduling of activities. Because the world can be constantly changing, the ability to revise one's assessment of the situation, current goal, and current response strategy in response to new information is basic to problem-solving in this domain. This

means the cognitive activities of the operator are best modeled as being "opportunistic" or interruptable by perceived changes in the state of the world. Emergency response can crystallize into a situation where the operators must make a choice among response strategies based on an uncertain situation assessment and risky possible outcomes (e.g., as occurred during the Ginna steam generator tube rupture incident).

In summary, the cognitive demands of NPP emergency operations produce the following processing requirements:

- process evidence to build a situation assessment given the possibility of multiple failures,
- sustained monitoring of evidence because it is a dynamic changing world,
- only a portion of the available evidence or possible explanations can be examined or pursued at any point -- attentional focus -- because of high workload and limited mental resources,
- there must be control and revision of attentional focus because it is a dynamic changing world,
- attentional focus is controlled through an interactive cycle of opportunistic, interruptable processing of new signals or events and knowledge driven choices about where to focus next,
- choice under uncertainty and risk.

A Cognitive Model.

The principal aim of a model is to efficiently capture relations among significant variables in order to describe, explain, and predict the behaviors of interest. To do this, models contain *concepts* and relations among concepts which specify what is really important in producing and controlling behavior in the situation of interest.[3] The concepts suggest what to look at and how to describe the situations that arise. CES is based on concepts about how intentions are formed and how they go astray that are derived from specific studies of human performance in NPP emergencies and general results in cognitive psychology.

---

[3]As Eddington (1939, p. 55) remarked, "in physics everything depends on the insight with which the ideas are handled before they reach the mathematical stage."

Second, models are *representations* of some aspects of the situation of interest. They do not duplicate the modeled world; there is a relation between the modeling system and the modeled system. CES is a modeling environment designed as a parallel world to actual emergency operations. CES translates from a description of the evolution of an incident and recovery responses in terms of NPP engineering language to a description in terms of a cognitive problem-solving language in order to identify difficult or error prone problem-solving situations.

Third, models have some *machinery* to formalize the concepts and to generate specific and reproducible outputs given some inputs. Concepts about the processes involved in intention formation require formalization as symbolic processing or AI mechanisms. CES was developed based on the knowledge representation and processing mechanisms of the EAGOL AI software system developed by H. Pople and Seer Systems.

This volume describes the concepts and scope of applicability for the cognitive model, and the AI techniques used to formalize it. It is intended for those who want to know about how CES models successful and erroneous human intention formation.

Finally, models have multiple uses. This model was developed in order to better capture the human contribution to risk in probabilistic risk assessment studies. The methodology for using the cognitive model in PRA is summarized in Chapter 3 of the executive summary and described in more detail in Volume 3. Volume 3 is intended for those who are interested in how the modeling capabilities of CES can be utilized in HRA and PRA.

2. The Approach to Modeling Human Performance

2.1 Introduction

The feasibility study done in Phase I found that all attempts to provide causal models of human performance in worlds where a broad range of cognitive activities occur result in *framework* models (e.g., Pew & Baron, 1983; Baron, 1984; Pew et al., 1986; Mancini et al., 1986). Framework models use one kind of modeling concept or technique to build a structure for the different kinds of cognitive activities that occur in the domain of interest and to capture how they interact. Narrower scope modeling concepts derived from heterogeneous sources provide depth at different points in the structure. This modeling strategy is used in many domains because there is a tradeoff between the desire for a formal model and the need to cover a broad scope of human behavior when modeling complex technological worlds (see sections 2.5 and 3.2 of NUREG/CR-4532).

The framework for the modeling system developed in this research program is based on a *model of the problem-solving environment* that is emergency operations. The emphasis is first on modeling the cognitive demands imposed by the problem-solving environment (the nature of the emergency incident, how it manifests itself through observable data to the operational staff, how it evolves over time). Then, concepts from narrower scope psychological models (monitoring dynamic systems, e.g., Moray, 1986; choice under uncertainty and risk, etc.) can be brought to bear to represent the factors that affect human behavior in meeting these demands and to constrain the model of the problem-solving environment. The most fundamental psychological constraint relevant to the NPP world is that people have limited cognitive processing resources, and this cognitive model was designed to simulate a *limited resource problem solver in a dynamic, uncertain and complex situation.*

Because this modeling approach was chosen, the resulting modeling capability has been named a *Cognitive Environment Simulation or CES.*

CES is a *causal model* in the sense that it generates predictions about operator action by simulating the processes by which intentions are formed. This contrasts with correlational approaches that base predictions on descriptive regularities between situational variables (e.g., time available to respond) and performance (e.g., likelihood of making an error) without simulating the processes that produce the error. The ability to simulate the processes that lead to a particular intention makes it possible to predict

likely behavior in complex and dynamic situations where operator intentions depend on a large number of interacting factors (e.g., what plant data he has available, number of issues competing for his attention, what he knows about the meaning of observed plant behaviors, the order that different kinds of explanations come to mind that could account for patterns of data) that would otherwise be intractable. Furthermore, it enables identification of the form of the error (e.g., a fixation error) and the sources of the error (what aspects of the situation confronting the operator and/or his knowledge or cognitive processing limitations contributed to the error.)

CES is formally expressed as an AI based computer problem solving system that carries out cognitive processes that are critical to intention formation in complex dynamic worlds -- it monitors plant behavior, forms a situation assessment, generates one or more explanations for the plant state, forms expectations as to the future course of plant behavior (e.g., that automatic systems will come on or off), and generates intentions to act. In particular, CES is a specific instance of the EAGOL artificial intelligence problem solving system that is capable of reasoning in complex dynamic worlds(see Footnote 2). Among EAGOL's unique strengths are the ability to reason in multiple fault situations and to reason in situations that evolve over time (i.e., where evidence accrues over time, where evidence may disappear or become occluded by new events, where beliefs about the state of the world must be revised, etc.).

Degrading these capabilities or what we call the basic cognitive competencies of CES, leads to error vulnerable problem solving behavior. Poor performance -- errors -- emerges from a mismatch between demands (the incident) and the knowledge and processing resources. Varying CES knowledge and processing resources increases or decreases the program's vulnerability to getting offtrack or, once offtrack, staying offtrack. In this view, errors are the outcome of a processing sequence, and a model of error mechanisms depends on a model of processing mechanisms. Thus, the cognitive activities that underlie the formation of an intention to act are encompassed in CES and errors arise due to limitations of these cognitive processes. This is the imperfect rationality approach to modeling human performance and error (e.g., Rasmussen, Duncan & Leplat, 1987).

Modeling consists of matching CES resources to those present in some actual or hypothetical NPP situation. The specific processing mechanisms in CES are not intended to be "micro" models of human cognitive processing. It is the outcome of the computer's processing activities that are assumed to be the same -- what data are monitored, what knowledge is called to mind, what situation assessment is formed, what explanations are adopted, and what intentions to act are formed, *given* the incident (the demands of the

problem-solving situation), the representation of the world (i.e., as reflected in the displays by which the operator interacts with the world), and the set of knowledge and processing limitations set up in CES.

The CES modeling environment provides powerful facilities for exploring how what a person knows, what data about the world are available to him, and his monitoring and problem-solving strategies can lead to successful or unsuccessful performance in different dynamic situations. Users of the model can express different particular NPP situations by selecting the demands (the incident or variant on the incident) and by adjusting the resources within the simulation to analyze and predict "what would happen if."

2.2 Overview of the Cognitive Environment Simulation

CES is a dynamic simulation capability for human intention formation. As shown in Figure 2-1, CES takes as input a time series of those values that describe plant state which are available or are hypothesized to be available to be looked at by operational personnel. Any valid source of data about how the plant would behave in the incident of interest can be used to create the inputs to CES. This includes data on plant behavior in actual incidents or simulation results derived from training simulation models, engineering simulation models, thermohydraulic codes.

The dynamic stream of input data constitutes a *virtual display board* which the CES simulation monitors to track the behavior of the plant over time, to recognize undesirable situations, and to generate responses which it thinks will correct or cope with these situations (intentions to act). Its output is a series of these intentions to act which are then executed and therefore modify the course of the incident.

CES is a modeling environment for the *supervisory role* during emergency operations. This is because CES does not actually execute its intentions. Another mechanism is needed to actually carry out CES's instructions on the power plant. For example, a person who has access to controls to a dynamic plant simulation can execute CES instructions. Whether this person executes CES's instructions correctly or not depends on the nature of the incident which the CES user wishes to investigate.

CES watches the virtual display board of potentially observable plant behaviors and generates actions that it thinks will correct or cope with the perceived situation. To do this, inside of CES there are different kinds of processing which are carried out "in parallel" so that intermediate results established by one processing activity can be utilized by another and visa versa. This allows a solution to be approached iteratively from different

levels of analysis.[4] There are three basic kinds of activities that go on inside of CES (Figure 2-2):

- Monitoring activities -- what parts of the plant are tracked when; are observed plant behaviors interpreted as normal-abnormal or expected-unexpected?

- Explanation building activities -- what explanations are considered, in what order, and adopted to account for unexpected findings?

- Response management activities -- selecting responses, either expected automatic system or manual operator actions, to correct or cope with observed abnormalities, monitoring to determine if the plans are carried out correctly, and adapting pre-planned responses to unusual circumstances.

An analyst can look inside CES to observe these activities as the incident it was stimulated with unfolds in time. The analyst can see what data the computer simulation gathered, what situation assessments were formed, what hypotheses were considered, pursued or abandoned, what plant behaviors were expected or unexpected. This can be done interactively, assuming that CES is being stimulated by dynamic plant simulation and assuming that CES intentions are being executed on the simulated plant. Or an analyst can examine a record or description of the knowledge activated and processed by CES after it has been stimulated by an incident. In both cases CES's processing activities and resulting intentions to act are available to be analyzed (1) to identify erroneous intentions, (2) to look for the sources of erroneous intentions, (3) to discover what other actions follow from erroneous intentions (Figure 2-3).

The CES user can vary the demands placed on CES -- how difficult are the problems posed by the input incident. The CES user also varies the resources within CES for solving the problems by modifying what knowledge is available and how it is activated and utilized. The dimensions along which CES performance can vary are called *CES Performance Adjustment Factors* (or PAFs). There are a variety of these adjustment factors designed into CES that provide tools for a human analyst to set up or model the particular NPP situations which he or she wishes to investigate within the cognitive environment simulated in CES. For example, CES should be

---

[4] In some psychological models there are linear stages of information processing where an input signal is processed through a fixed sequence of stages. In CES, different processing occurs at the same time and intermediate results are shared. This leads to formalization as an AI program.

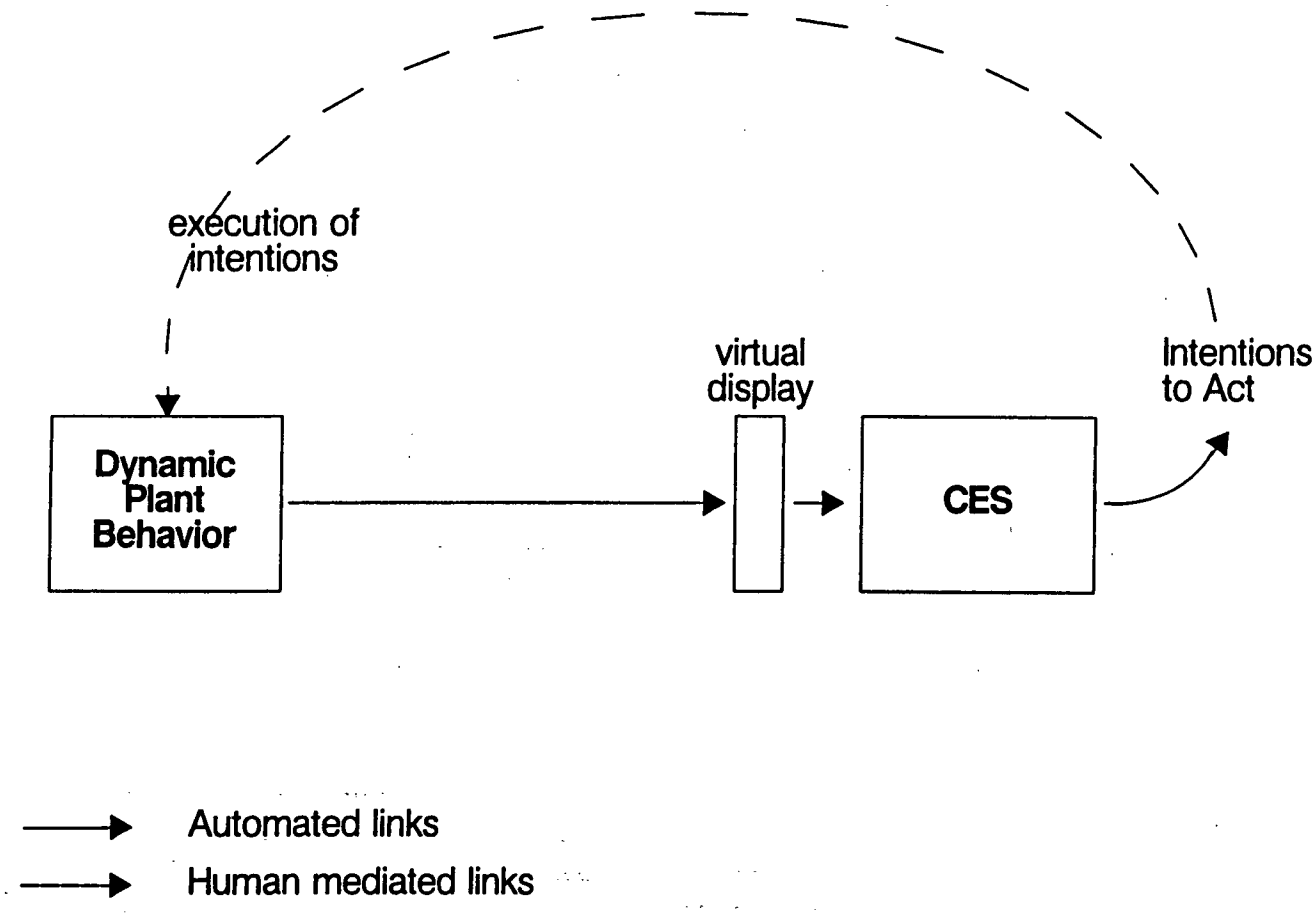


Figure 2-1: CES is a dynamic simulation capability for human intention formation. It takes as input a time series of those values that describe plant state which are available or are hypothesized to be available to be looked at by operational personnel. The CES simulation watches this virtual display board of potentially observable plant behaviors to track the behavior of the plant over time, to recognize undesirable situations, and to generate responses which it thinks will correct or cope with these situations (intentions to act). Its output is a series of these intentions to act which are then executed and therefore modify the course of the incident.

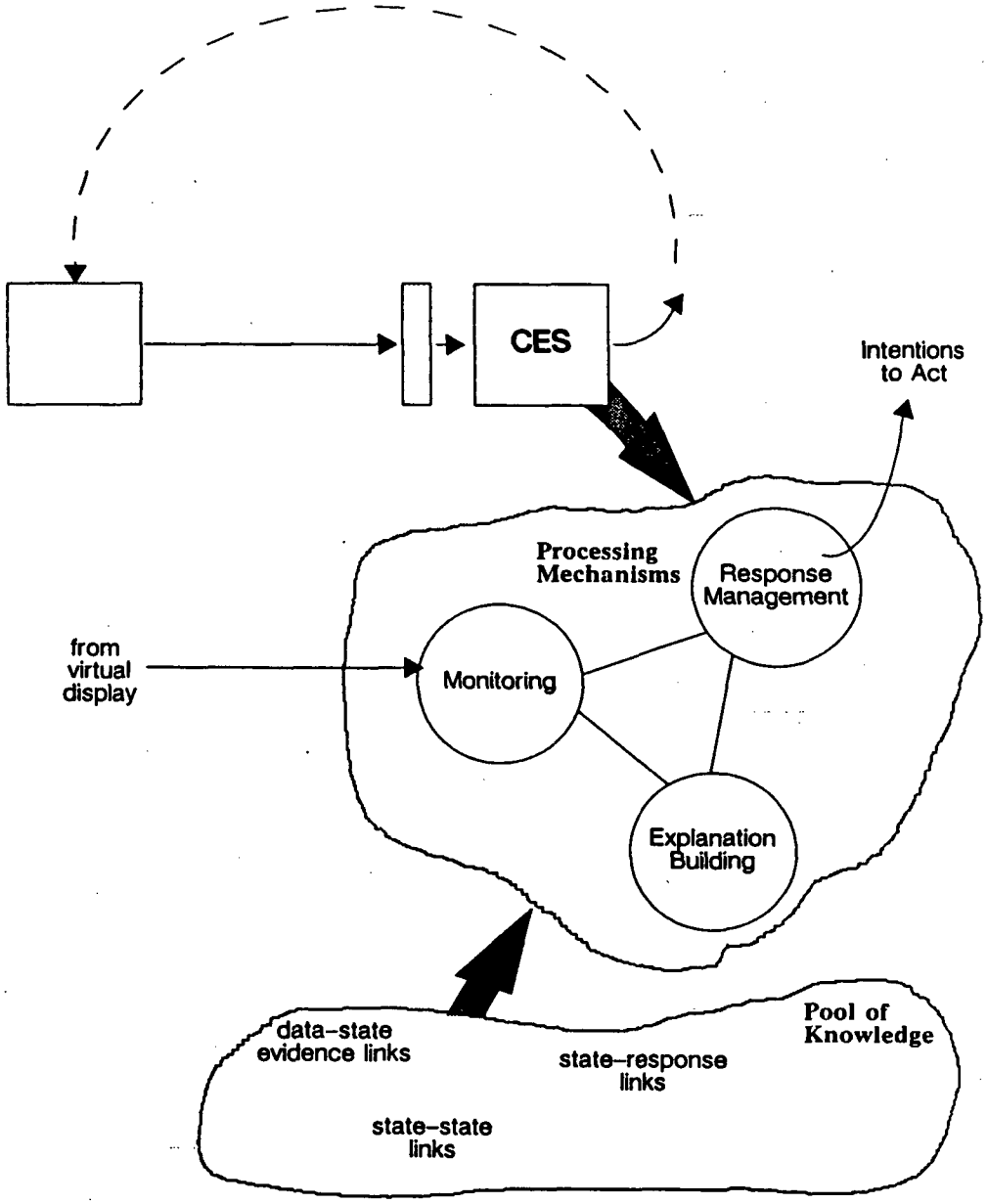


Figure 2-2: Inside of CES there are different kinds of processing which are carried out "in parallel" so that intermediate results established by one processing activity can be utilized by another and visa versa. This allows a solution to be approached iteratively from different levels of analysis. There are three basic kinds of activities that go on inside of CES: (a) monitoring activities – what parts of the plant are tracked when and are observed plant behaviors interpreted as normal-abnormal or expected-unexpected? (b) explanation building activities – what explanations are considered, in what order, and adopted to account for unexpected findings? (c) response management activities – selecting responses, either expected automatic system or manual operator actions, to correct or cope with observed abnormalities, monitoring to determine if the plans are carried out correctly, and adapting pre-planned responses to unusual circumstances.

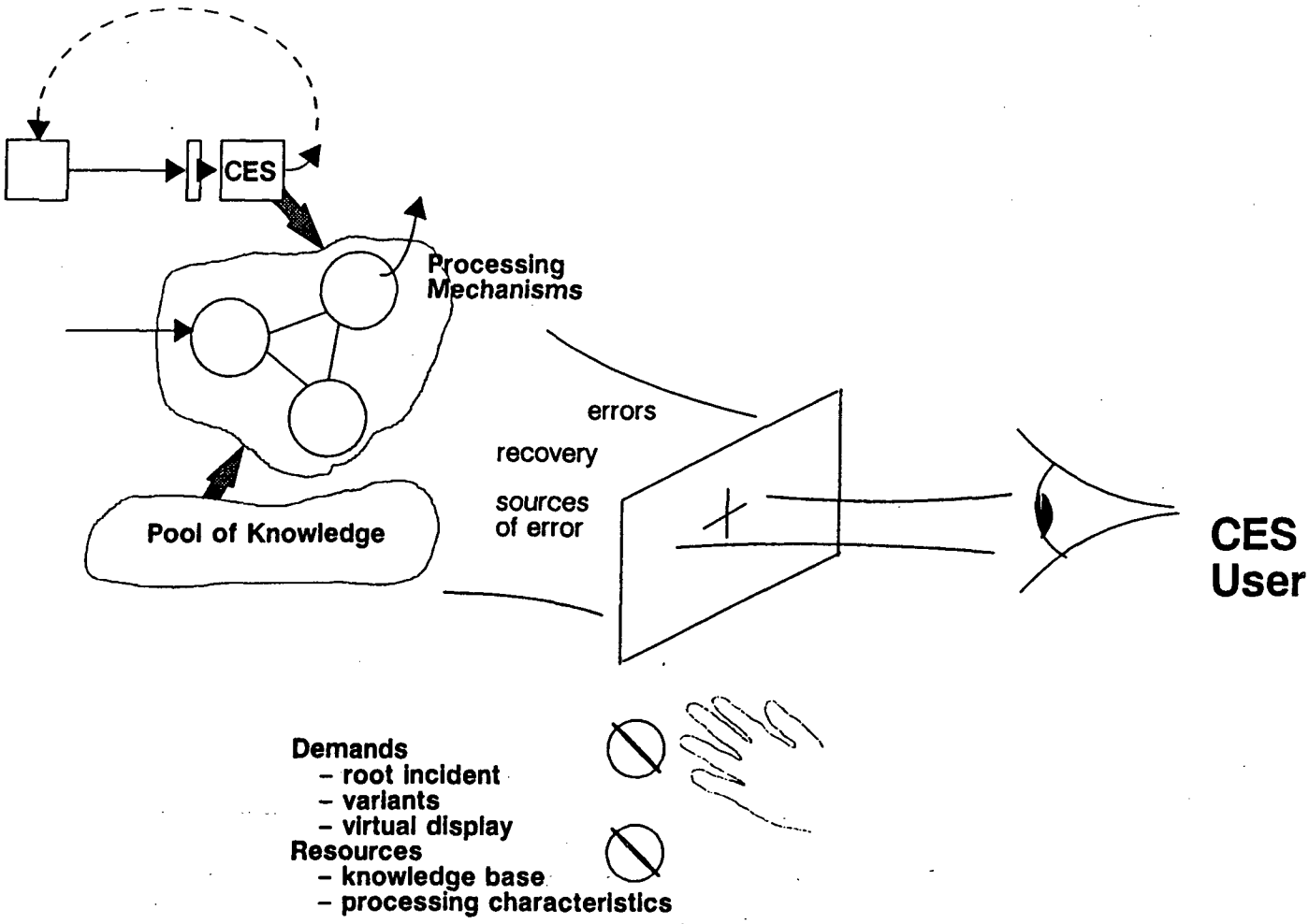


Figure 2-3: An analyst can look inside CES to observe these activities as the incident it was stimulated with unfolds in time. The analyst can see what data the computer simulation gathered, what situation assessments were formed, what hypotheses were considered, pursued or abandoned, what plant behaviors were expected or unexpected. CES's processing activities and resulting intentions to act are available to be analyzed (1) to identify erroneous intentions, (2) to look for the sources of erroneous intentions, (3) to discover what other actions follow from erroneous intentions.

The CES user varies the demands placed on CES – how difficult are the problems posed by the input incident. The CES user also varies the resources within CES for solving the problems by modifying what knowledge is available and how it is activated and utilized. The dimensions along which CES performance can vary are called CES Performance Adjustment Factors (or PAFs).

capable of responding in a "function-based" and/or in an "event-based" fashion to faults, and CES should be capable of being fixation prone or not being fixation prone in explanation building. Modeling NPP situations within the CES simulation environment is, in effect, a translation from the engineering languages of NPP incidents to a problem solving language as represented by the knowledge and processing mechanisms set up in CES. CES is then run to find the conditions that lead to erroneous intentions and the action consequences of these erroneous intentions.

This type of system can, in principle, function in the supervisory role of operational personnel during an unfolding NPP incident. The relationship between processing activities that occur in CES and processing activities of a person or team of people in some NPP emergency situation varies depending on the knowledge, processing resources, and processing mechanisms set up in CES:

- CES may carry out NPP tasks competently, in the sense of "ideal" performance;
- it may carry out these tasks like "good" human operational personnel or teams;
- it may err in these tasks like human operational personnel or teams err.

A human performance model must be built based on knowledge of what people actually do in the situations of interest. If one knew this completely, then the benefit of formal modeling is to eliminate subjectivity in the application of this knowledge to specific cases. But our knowledge of human performance in complex dynamic worlds such as NPP operations is incomplete (e.g., Hollnagel, Mancini & Woods, 1986). Given this state of affairs, formal models are needed (a) to objectively express the current state of knowledge, (b) to extrapolate from this to new situations, (c) to test whether the current state of knowledge is adequate through comparisons to new empirical cases, and (d) to revise and update the state of knowledge as appropriate (the model as a repository of current knowledge/best guesses/approximate models on operator behavior).

The Cognitive Environment Simulation allows one to formally represent the state of knowledge about what people do in emergency operations (or alternative views about what they do) and then to see the implications of that knowledge (or point of view) for human intention formation in new situations where there is no or sparse empirical data. Thus, a cognitive environment simulation allows one to generate analytical data on human

performance that complement, but do not replace, empirical data on human performance.

This state of affairs is analogous to the situation with analytical computer codes which model reactor behavior. In both cases, an ongoing cycle of model evolution and change is needed as our state of knowledge changes. The Cognitive Environment Simulation, as repository of the best current knowledge, then, becomes the best source for interpolating or extrapolating what human behaviors are likely in cases where there is no or limited experience – including evaluating changes to the human-machine system and hypothetical situations that arise in postulated incidents for which there is no or insufficient empirical data (rare incidents). Reactor thermodynamic models are essential tools for design and risk assessment of the physical NPP. The Cognitive Environment Simulation provides, for the first time, an analytical model of human intention formation in NPP emergency operations which will be an essential tool to assess human performance for the evaluation of human-machine systems in the NPP and for assessment of the human contribution to risk.

2.3 CES Development Process

The process by which CES was created is illustrated in Figure 2-4.

Concepts and relations about how intentions are formed and how they go astray derived from empirical results and knowledge about the structure and function of NPPs were used to formulate a set of basic *cognitive competencies* that CES should exhibit. As mentioned earlier, the basic competencies are imposed by the need to simulate a limited resource problem solver in a dynamic, uncertain and complex situation.

If CES was to function as a modeling environment, the cognitive competencies also needed to include the dimensions along which CES performance should be variable – CES Performance Adjustment Factors (PAFs). CES should be capable of competent performance given some set of Performance Adjustment Factor settings, and should be capable of incompetent performance given other Performance Adjustment Factor settings. Furthermore, the performance breakdowns which CES exhibited under different PAFs must be related to what is known about how human problem-solving can break down in dynamic situations.

The concepts about intention formation were derived from general results in cognitive psychology and from empirical studies of human performance in NPP emergencies (cf., Chapter 4 of NUREG/CR-4532). Empirical results used included both studies of operators solving simulated faults (Woods et

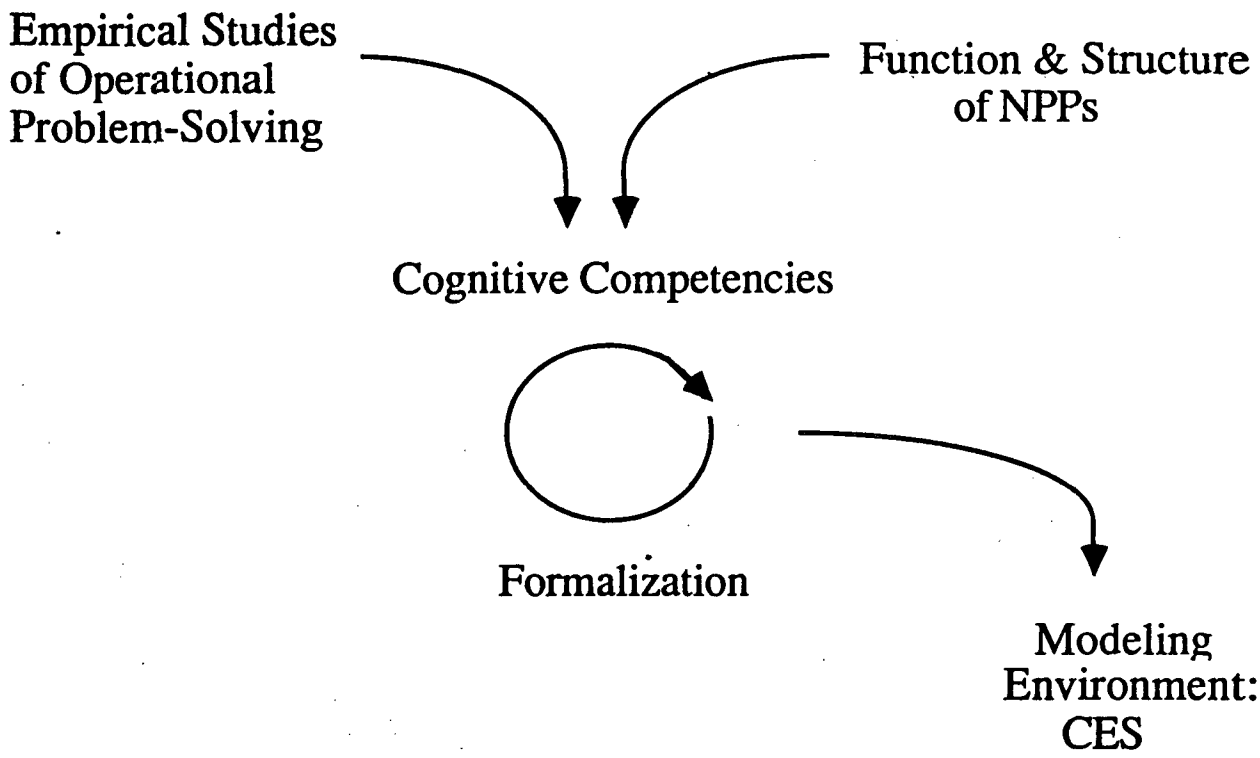


Figure 2-4: The CES development process. Concepts and relations about how intentions are formed and how they go astray were used to formulate a set of basic cognitive competencies that CES should exhibit. A formalization process followed where AI mechanisms were set up that could exhibit those competencies. Several iterations of formalization, leading to more refined statement of the basic competencies and then further formalization were carried out to develop CES to its current state.

al., 1982; Woods & Roth, 1982; and unpublished cases) and retrospective investigations of operator decision making in actual incidents (e.g., the four incidents analyzed in Pew et al., 1981; the Ginna and Oconne incidents analyzed in Woods, 1982 and Brown & Wyrick, 1982; the Davis-Besse incident reported in NUREG-1154; the San Onofre incident reported in NUREG-1190; the Rancho Seco incident reported in NUREG-1195).

In the CES development process, the different types of knowledge that a person might possess about the NPP were also taken into account. CES had to be capable of representing these different kinds of knowledge and different ways of organizing knowledge about the NPP. The formalism for organizing knowledge about NPPs that informed CES development is based on Gallagher et al. (1982), Woods & Hollnagel (1987), and Woods (in press).

Chapter 3 describes in detail the cognitive competencies which CES behavior should exhibit.

A formalization process followed where AI mechanisms embodied in the EAGOL artificial intelligence problem solving system were set up that could exhibit those competencies. The basic software mechanisms had to be capable of competent performance and capable of being degraded to exhibit the kinds of performance breakdowns that humans exhibit in high cognitive demand situations. Several iterations of formalization, leading to more refined statement of the basic competencies and then further formalization were carried out to develop CES to its current state.

Chapter 4 describes the mechanisms by which CES exhibits the competencies and the current state of CES development. Chapter 4 also contains samples of CES processing in different NPP situations.

2.4 Overview of the CES Architecture

As an instance of the EAGOL AI computer system, CES contains two major types of information. First, it contains a *knowledge base* that represents the operator's (or the team of operators') knowledge about the power plant, including the inter-relationships between physical structures, how processes work or function, goals for safe plant operation, what evidence signals abnormalities, and actions to correct abnormalities.

Second, it contains *processing mechanisms* (or inference engine) that represents how operators process external information (displays, procedures) and how knowledge is called to mind under the conditions present in NPP emergencies (e.g., time pressure). This part of the model determines what

knowledge is accessed when and what cognitive activities (monitoring, explanation building, response management) are scheduled when during an evolving incident.

The knowledge representation formalism from EAGOL (i.e., how knowledge about the NPP is expressed) provides a powerful and flexible mechanism for representing virtually any relation among NPP concepts. Concepts at any level of abstraction, whether observable or not, can be represented (e.g, a plant parameter reading; an intermediate disturbance category such as a "mass imbalance"; a fault category such as primary system break to containment; or a response such as "turn off the emergency cooling system"). Within the knowledge representation formalism, the full variety of relations among concepts that NPP operators would be expected to know such as plant data-state evidence links, state-state links, and state-response links can be expressed. This includes encoding of symptom-response "shortcuts" that form the basis for what has sometimes been termed operator "rule-based" behavior, as well as encoding of more abstract and functional relations that form the basis for more elaborated and thorough reasoning or what has sometimes been termed "knowledge-based" behavior (Rasmussen, 1986).

Included in the knowledge representation is a description of what data about plant state are directly available to the model to "see," reflecting what plant information would be directly available to the operator to observe. This description constitutes a *virtual display board*, that the model monitors to acquire data about plant state. The CES knowledge base includes a list of plant parameters or states that it can directly access (e.g., from a data file or as output from a simulation program). Depending on the plant being modeled these plant parameters can be direct sensor readings, or more integrated information about plant state such as the output of computerized displays or decision aids). Associated with each element on the "virtual display board" are parameters that reflect characteristics of how that information is presented in the plant being modeled (i.e., characteristics of the *representation* provided to the operator of that NPP).

The basic psychological concept behind CES is that people have limited resources in a potentially high workload environment. This means that CES, as a model of operational personnel, cannot access and utilize all possibly relevant pieces of knowledge (i.e., not all potentially relevant knowledge in the knowledge base can be *activated*) on any one model *processing cycle*, i.e., time step). Similarly, CES cannot examine all of the plant data available at any one processing cycle. Therefore, CES must be able to control what data are examined when and what knowledge (and how much knowledge) is activated in a given cycle. This is one of the basic cognitive competencies specified for CES.

Controlling what knowledge and how much knowledge is activated at a given point in an unfolding incident depends on:

- A cycle or interaction between *knowledge-driven processing* (such as looking for information to find an explanation for an unexpected finding) and *data-driven processing* (where salient data interrupts ongoing processing and shifts the focus).

- *Resource/workload interactions* where carrying out one type of processing precludes the possibility of doing other processing if there is competition for limited resources. Thus, there can be a need to choose which processing activity should be carried out next, e.g., acquire more data? or pursue possible explanations? or generate/track responses to detected abnormalities?

- A limited problem solver should focus first on *"interesting" findings.* There are several layers of criteria that define which findings are "interesting" or "important" that affect control of CES processing. For example, if an observation indicates an abnormality, then there is a need to pursue how to correct or cope with it; if an observation is unexpected, then there is a need to pursue what could account for it?

The formalization task then was to use the symbolic processing or AI mechanisms in EAGOL to control a limited focus of attention in these ways, e.g., what data are examined when, what possible explanation is pursued first.

The basic processing mechanism from the EAGOL system used in CES to achieve this behavior is to spawn an "analyst" when some criterion is met, who then performs some information processing work, accessing knowledge available in the knowledge base as it needs it. There are three basic kinds of "analysts" each with their own area of responsibility and with different criteria that trigger their processing activities. These are:

- *Behavior analysts* responsible for monitoring and analyzing plant behavior to decide if observed plant behaviors are expected or unexpected.

- *Situation analysts* responsible for analyzing the perceived situations and for postulating and pursuing possible explanations for unexpected findings.

- *Response plan analysts* responsible for selecting and adapting plans to correct or cope with perceived abnormal conditions.

These analysts are active processes that draw conclusions and "post" their results for other analysts to use as needed. Multiple instances of each basic type of "analyst" are generated or "spawned" as needed. A fundamental characteristic of this problem-solving architecture is that each analyst has a very narrow field of view and responsibility, and that complete problem solving involves communication and coordination among the multiple analysts.

Each analyst does not represent a different person, rather the cooperative set of analysts are intended to model a single problem-solving system -- be it an individual operator or a team of operators. The multiple analysts are intended to model the multiple types of processing (e.g., monitoring, explanation building, response planning) and lines of reasoning (e.g., multiple alternative explanations pursued) that occur in parallel and are interwoven during problem-solving

2.5 Modeling Human Intention Formation with CES

CES, as a modeling environment, is designed as a parallel world to actual emergency operations. The parallel is established by capturing in the simulation world the problem solving resources available in some actual or hypothetical NPP situation (operators, training, procedures, control board, etc.). If the parallel is well established, then the behavior of the simulation (the monitoring, explanation building and response management behavior of CES) in some incident corresponds to expected human behavior in the actual world, under the same circumstances.

Note that the specific mechanisms in CES are not intended as "micro" models of human mental operations. Rather CES was designed to exhibit the monitoring, explanation building, and response management behaviors of a limited resource problem solver in a dynamic and complex situation. For example, a limited problem solver must "decide" what plant data to look at a particular moment; therefore CES must be capable of tracking some data to the exclusion of others. This specifies a cognitive competence which CES should exhibit. If the mechanism that directs CES monitoring focuses its limited resources on data that are not relevant or are even misleading with respect to that actual state of the NPP, then CES can form erroneous situation assessments and therefore select erroneous actions. This means that CES is a dynamic simulation that can behave incorrectly if demands exceed resources.

The basic assumption in this modeling approach is that degrading a competency imposed by the problem solving environment can lead to more error vulnerable problem solving behavior. The power of this approach is

that, depending on how the competencies are formalized and implemented, it is straightforward to provide different levels of resources and to create different processes by which resources are allocated when competition occurs. The constraint on defining these resource limitations is that they can lead to known *forms of human errors* such as fixation errors.

2.5.1 Changing CES Resources: Performance Adjustment Factors

CES is a deterministic model. Given the same dynamic incident scenario, the same virtual display board characteristics, the same knowledge about the NPP, and the same processing resources, CES will generate the same series of intentions to act. There are large degrees of variability in human behavior; even when performance is good, people take different trajectories to reach the same outcome. CES is capable of large degrees of variation in its behavior as well, and it is capable of taking different problem solving trajectories to the same outcome.

Variability in CES behavior arises from several sources. First, variability in CES behavior arises due to variability in details in how the incident in question unfolds. This is one reason why dynamic plant behavior is needed as input to CES. Second, CES behavior varies as a function of variations in its knowledge and processing resources. The assumption is that human variability arises from differences in relatively enduring knowledge (e.g., knowledge of how x works) and processing characteristics (e.g., a fixation prone personality), longer term changes in knowledge and skill (e.g., skill acquisition from training or experience), or from more moment-to-moment variations in processing resources (e.g., a narrow field of attention due to stress or fatigue).

There are a set of factors designed into CES which allow one to vary CES knowledge and processing resources. These Performance Adjustment Factors (PAFs) provide the tools for a human analyst to establish parallels between the cognitive environment simulated in CES and NPP situations which he or she wishes to investigate. The analyst uses PAFs to represent the resources available (or thought to be available) in a particular NPP situation within the CES modeling environment. The CES user then stimulates CES with data on plant behavior in different incidents, checks how CES solved those problems (intention failures, omission and commission errors that follow from intention failures, error recovery), re-adjusts PAFs to explore variants, and re-runs CES to identify the conditions under which intention errors occur, the consequences of intention errors, and the sources of intention errors.

Traditional performance shaping factors (e.g., experience level, stress, organizational climate) are examples of variables that are thought to affect

human behavior. CES Performance Adjustment Factors (PAFs) are variables that affect CES behavior. To simulate a NPP situation in CES, the factors operative in that situation which are thought to affect human behavior are mapped into CES PAFs in a *two step inference process.* First, one must specify what is the impact of the factor of interest (or a change in that factor) on cognitive activities. This can be derived from theoretical concepts (e.g., the effect of team structure on problem-solving processes), empirical data, or analysis. In any case, it is the effects on the *processes* involved in activating and utilizing knowledge which must be specified. Second, the specified effects on cognitive processing are translated into adjustments in PAFs.

For some kinds of performance shaping factors this two stage inference process is a straightforward, tractable analytical task. For example, with respect to the effect of procedures on performance, the specific guidance on corrective responses encoded into the procedures (e.g., specification of corrective responses to take) would be extracted and entered into the CES knowledge base. With respect to issues of display quality, the relative salience of different plant data on a control board or in a computer display system would be determined by the CES user analytically and used in the set up of the virtual display board.

Other kinds of factors can be specified based upon straightforward empirical investigation. For example with respect to effects of training or experience, one can use simple "quick and dirty" techniques or more sophisticated techniques to find out what particular operators actually do know about how some plant process works (e.g., natural circulation), about the basis for some response strategy, or about what possible hypotheses are brought to mind by some plant behavior(s).

Finally, some factors require a specification of *how they are assumed to affect problem-solving processes* in order to be mapped into CES PAFs. For example, how does stress affect problem-solving (e.g., high stress might narrow the field of attention) or how do different organizational structures affect problem-solving? The answer to this question specifies what PAF settings should be used to investigate the consequences of this factor on intention formation errors in different incidents and over various other PAF settings.

Note that the answer to this question requires taking a theoretical position on how the factor in question impacts on the processes involved in problem solving. The theoretical relation asserted can be derived from behavioral science research (e.g., the impact of team structure on problem solving) or analyst judgment. This is an example of how CES is a framework model that utilizes more specific models in some areas.

The mapping between the NPP and CES PAFs is many-to-many. This means that there are many circumstances which might affect, for example, the degree of fixation proneness (various external factors such as displays, decision aids, team structure and various internal factors). Similarly, there are multiple Performance Adjustment Factors which could be manipulated individually or jointly to affect, for example, the degree to which CES is prone to fixation. Different versions of CES can be set up via the PAFs to represent different kinds of human problem-solving behavior, for example, a fixation prone version of CES can represent a "garden path" problem solver. Table 2-1 contains other example pre-settings of CES processing PAFs that might capture known emergent patterns of human cognitive processing.

Adjusting CES PAFs to represent different NPP situations requires knowledge about what factors affect human cognitive activities, and it requires knowledge about how CES represents knowledge and about how CES processing mechanisms function.

2.5.2 Cognitive Processing and Erroneous Intentions

An analyst uses CES PAFs to change the knowledge and processing characteristics within CES or the virtual display of data to CES. This allows the CES user to explore under what conditions intention failures occur and to see the consequences of intention failures for further actions on the plant in different incidents or variations on a root incident. Errors -- failures to form the appropriate intentions for the actual situation -- depend on how CES activates and uses knowledge, given the demands of the incident under investigation.

Finding intention failures with CES is based on the concept that the difficulty of any given problem-solving situation is a function of

1. *The problem-solving demands.*

    - Processing requirements imposed by the characteristics of the incident (e.g., a multiple fault incident where one masks the evidence for another is inherently more difficult to isolate that a single fault incident with a clear signature).

    - The *representation* or window on the world by which the problem-solver views and interacts with the incident (e.g, the displays available on the control board; integrated information available on computer-based displays).

Table 2-1: Examples of modeling different types of problem solvers via Performance Adjustment Factors.

This table contains examples of some emergent patterns of human cognitive processing which are relevant to NPP emergency operations and the settings of CES PAFs which might produce these patterns in CES behavior.

- **1. Vagabond:**
  A vagabond problem solver (after Dorner, 1983) abandons the current issue for each new one that arises. The tendency is to jump from issue to issue without satisfactory resolution of any. It is characterized by an incoherent view of the situation and incoherent responses to incidents. This pattern could emerge due to the following, especially when there is some time pressure:

  - failure to synthesize or converge multiple views of the situation,
  - many potential views of the set of significant findings are activated but remain independent,
  - a response orientation emphasized over explanation building so that more coherent explanations never emerge,
  - too interrupt-driven so that every new finding seizes priority.

- **2. Hamlet:**
  This type of problem solver looks at each situation from multiple viewpoints and considers many possible explanations of observed findings. However, the result is a tendency to examine possibilities too long before acting because

  - its criterion for judging what is an acceptable explanation is missing or is too general (too many possibilities satisfy it)
  - explanation building is greatly emphasized over response management activities.

Table 2-1, continued

- **3. Garden Path:**
  A garden path or fixation prone problem solver shows excessive persistence on a single issue or activity – easily fixated, fails to consider revision in face of disprepant evidence. PAFs relevant to produce this type of behavior include:

  - pursues (or is biased to pursue) only a single point of view to explain findings;
  - not interrupt driven enough; in the extreme case, no new issue interrupts ongoing activity until scheduled activity is completed;
  - insensitive to violations of expectation, after an initial explanation is accepted, because too narrow a field of view or because response management overrides explanation building.

- **4. Inspector Plodder:**
  This type of problem solver slowly and deliberately builds up and then narrows in on possibilities. It exhibits very thorough consideration of evidence and possible explanations via explicit chains of reasoning (minimal reasoning shortcuts). The result is good, thorough, but slow problem-solving. Performance adjustment factors related to this pattern are a narrow field of attention; low interruptability; sequential, deep exploration of possible explanations, good criteria for scheduling competing activities, and good criteria for what is a good explanation.

- **5. Expert Focuser:**
  This problem solver is adept at seeing and focusing in on the critical data for the current context so that it is always working on the most relevant part of the situation. The whole situation tends to fall quickly into place, and revisions are easily made when appropriate. Performance adjustment factors related to this pattern are a wide field of attention; high level of interruptability; good criteria for scheduling competing activities; good criteria for what is a good explanation.

2. The available *problem-solving resources.*

   - The base of knowledge about the NPP that is available to use in problem-solving. This includes knowledge about the structure and function of the NPP and knowledge about NPP disturbances/faults, and how to correct these.

   - The processing mechanisms and their characteristics (e.g., size of the field of attention, how fixation prone, degree of communication among different processing mechanisms).

Errors emerge when there is a mismatch between demands and resources. For example, a narrow field of attention (low resources) cannot lead to intention failures if the incident in question produces no situations where a wide field of view is needed for timely detection of important plant behaviors (low demands).

Intention formation errors are the end result of a processing sequence which starts and develops due to failures to call to mind relevant knowledge -- a plant behavior is missed (which could happen due to several factors, such as, because of low observability of the data or because the focus of attention is elsewhere), the knowledge it would have evoked is not brought to mind and does not lead to an accurate situation assessment (e.g., plant behavior x is interpreted as expected instead of unexpected), the erroneous situation assessment affects what explanations are pursued or not pursued and what responses are evoked or not evoked. Varying CES processing resources through PAFs increases or decreases the program's vulnerability to getting offtrack or, once offtrack, staying offtrack.

Failures to call to mind relevant knowledge are the seeds from which intention errors emerge. But a failure to call to mind knowledge at one point in an evolving incident does not mean that erroneous intentions and actions are inevitable. Often, they represent different paths by which someone sees, thinks through, and acts in the same situation. Furthermore, as an incident evolves there are opportunities to update and revise one's situation assessment and get back on track.

An analyst observes CES attempt to recover from an incident just as he might watch people attempt to recover from an incident on a training simulator or on a real plant. CES itself does not analyze its own behavior; a user of CES can watch how it behaves in some situation and can judge when it forms erroneous intentions and therefore would take erroneous actions (although note that most of the difficulties that arise in analyzing human behavior during simulated or real incidents, such as ambiguities in

defining what are errors, will arise in analyzing how CES behaved in some incident as well). The CES user can adjust problem-solving demands and the available problem-solving resources and rerun CES to identify points where mismatches occur – i.e., where errors emerge.

2.6 Expressing NPP Situations in CES

This section contains illustrative examples of how PAFs can be adjusted to express different NPP operational situations within the CES modeling framework.

In some plants the prescribed operational philosophy is that operators should only monitor and verify proper automatic system responses for the first five minutes following an automatic reactor shutdown (i.e., no operator diagnosis and action). One expressed intent for this rule is help avoid erroneous initial diagnoses. The evaluation question is: what is the impact of this operational philosophy?

This philosophy can be translated into the CES simulation world in terms of the relative priority between monitoring, explanation building, and response management activities. In particular, monitoring activities dominate any explanation building, and response management activities go on only with respect to verifying expected automatic system responses. Note however that observed findings cannot be prevented from calling to mind knowledge about the situation or possible explanations. In CES terms this means data-driven activation of explanations occurs although knowledge-driven pursuit of these possibilities is temporarily postponed. This captures a finding from cognitive psychology that people have difficulty in trying to suppress the knowledge evoked by some signal.

The consequences of these processing modifications then can be investigated by actually running CES in different incidents. For example, does this philosophy result in a reduction in erroneous diagnoses or not?

2.6.1 Multiple People, Multiple Facility Problem-Solving

One question that frequently arises is how can a cognitive model address multi-person problem-solving situations? Is there a separate version of the program to represent each person in the situation? The answer to the latter question is no. Multiple people are represented in CES in another fashion.

In CES, as a symbolic processing based model, different kinds of processing are carried out in "in parallel" so that intermediate results established by

one processing activity can be utilized by others. This can be taken advantage of to partition CES processing activities to represent the interaction of multiple problem solvers in different roles or with different NPP areas of responsibility.

Different roles played by different people can be represented by mapping each role into the kinds of processing activities within CES that are associated with it, i.e., at a high level, the monitoring, explanation building, and response management activities. When people play different roles, the capacity of the corresponding processing activities is higher and bottlenecks are reduced. However, just because there are multiple people present does not mean they will amplify processing resources. For example, in the technical support center during the Ginna incident (Woods, 1982) the entire staff of this facility was engaged, for the most part, in deciding one choice under uncertainty and risk (for about 90 minutes). The organizational processes that went on in this incident resulted in most problem solving resources being focused on one response management choice to the exclusion of other problem solving activities (failures to observe other abnormal situations that would have benefited from corrective action; cf., Woods, 1982).

One way that multiple people are organized is by responsibility for different areas of the NPP. For example, the usual prescribed organization of the control room is one supervisor and two operators at the control board, one responsible for the reactor systems (the reactor or RO operator) and the other responsible for the balance of plant systems (the balance of plant or BOP operator). Different areas of responsibility can be represented in CES by partitioning plant data into groups corresponding to the different areas of responsibility. Interesting plant behaviors within the area of responsibility would trigger further CES processing (data-driven processing), while interesting plant behaviors outside of the area of responsibility could be noticed and processed only if some processing activity specifically requested to know the state of that piece of plant data (knowledge-driven processing). This manipulation is one PAF. This type of partitioning means that explicit communication between parts of CES is needed that corresponds to verbal requests for information between people with different areas of monitoring responsibility. Similarly, explanation building activities needs to be partitioned into groups so that separate explanations are pursued, one starting from findings about secondary side systems and the other from findings about primary side systems. The partitioned explanation building activities then need to explicitly communicate their own results if they are to cooperate in accounting for unexpected findings. The communication between these kinds of partitions within CES processing mechanisms can be varied to represent crews with poor inter-communication or methods to increase

communication, such as operator training (drill on calling out what they see to the other operators) or display systems designed to provide a common context among multiple operators.

For example, in a conventional control board configuration the reactor operator position might not have direct visual access to data on the secondary system (e.g., steam generator level). If this were the case, then this operator would only know about the status of steam generator level if he/she explicitly requested this information. For example, in one simulated accident (Woods et al., 1982) a complicating factor was added to the incident that removed the leading indicator of the disturbances (a loss of offsite power preceded a steam generator tube rupture so that secondary radiation indication was disabled). The reactor operator was trying to find an explanation for a persistent decrease in pressurizer level and pressure. After puzzling over this for several minutes, he asked the BOP operator, "do you see anything funny about the steam generators?" It was only after this prompt that the BOP operator realized and reported that there was clear evidence for an abnormal steam generator -- a finding that resulted in the entire picture becoming clear for the team.

2.6.2 Procedures

Another major characteristic of emergency operations is that operators have guidance available to them in the form of procedures. The role of the guidance available in procedures in intention formation is handled by decomposing procedures into the kinds of knowledge about the plant, accidents, and corrective responses that they provide to operational personnel. Procedures in the nuclear industry provide guidance about what situations should be monitored for (e.g., symptoms, disturbances to safety functions, accident classes), what evidence signals when a system or function is disturbed, cues about when to monitor different parts of the plant or different issues, and, of course, how to respond to different situations should they arise.

For the most part, the different types of knowledge carried in procedures is represented in CES in the links in the knowledge base. The grain of response varies both within a single set of procedures and across different styles of procedures (event based versus symptom or function based). This is captured in the structure of sets of situation-response links that are expressed in the knowledge base. Note that there is no global performance shaping factor for procedures requiring subjective judgments of what are "good" or "bad" procedures. Instead, the CES user decomposes procedures into what guidance they provide (a clear cut analytical task) and then runs CES for objective results on how this guidance supports operational personnel during accident recovery.

3. CES Cognitive Competencies

3.1 Introduction

This chapter describes the cognitive competencies which must be exhibited for a cognitive environment simulation to capture how intentions to act are formed, both correctly and incorrectly, in NPP control rooms during emergencies. It is a description on paper of the processing mechanisms needed for CES to function as a modeling environment.

One can think of the competencies as the fundamental properties of a problem-solving system within a particular kind of world, analogous to the fundamental properties of thermodynamic systems. And one can think of the processing mechanisms as a particular design for a particular purpose (finding intention failures) analogous to a kind of reactor as a particular design for particular purpose or tradeoffs. The commodities dealt with differ – energy and water versus knowledge and information. The properties of these are different; therefore, the principles and techniques applied are different. But just as reactor behavior is modeled by computer simulations of thermodynamic processes embodied in plant equipment, so human behavior can be modeled by computer simulations of problem-solving mechanisms. The competencies needed to create such a simulation for NPP emergency operations include control of attention in a limited resource and high mental workload environment, monitoring a changing world, situation assessment including forming expectations, qualitative reasoning, building explanations to account for unexpected findings, and plan selection/monitoring/adaptation to correct detected abnormalities.

This chapter describes CES as a conceptual model, but a conceptual model designed to be formalized through implementation as a symbolic processing (or AI) computer program. The conceptual model (CES) describes the target behaviors which should be exhibited by the computational machinery. The Eagol software system provides the base of computational machinery for formalizing the target capabilities of CES.

This chapter does not address the computational mechanisms themselves which will implement CES concepts. This is addressed in Chapter 4. This chapter also does not contain descriptions of how CES could be used in the context of HRA or PRA. This is the subject of Volume 3.

3.2 Problem-Solving with Limited Resources and High Workload

Given that people have finite resources and that the NPP is dynamic and consists of many highly coupled parts, a basic feature of the cognitive environment is the need for *control of attention.* This means that a problem solver cannot simultaneously attend to all data or evidence about the state of the world and he/she/it cannot simultaneously think about all issues, possible explanations or possible responses. Because of this limited capacity, attentional processes about how to select, spread and change attentional focus in pace with the changing state of the world are a critical part of CES. The basic question is what factors affect, given some observed event in the world, *what is called to mind?* An important source of performance breakdowns is failures to call to mind relevant knowledge (although failures to call to mind relevant knowledge do not always lead to observable erroneous actions).

What CES "calls to mind" depends on –

- does it look and does it see the event in the world – *observation failures*?
- does it have the relevant knowledge available to call to mind – *missing or inaccurate knowledge*?
- does available relevant knowledge actually get called to mind given other competing activities – *inert knowledge* (e.g., Bransford et al., 1986; Perkins & Martin, 1986)?

CES contains processing mechanisms which monitor the changing stream of plant data, observe events or occurrences, and "call to mind" knowledge relevant to what is observed. The mechanisms that control what knowledge and how much knowledge is activated at a given point in an unfolding incident include: (a) an interactive cycle between knowledge-driven and data-driven processing, (b) resource/workload interactions, (c) layers of criteria which define what are "interesting" or "important" findings.

Attentional focus is partly determined by the incoming data stream where salient data interrupt ongoing processing and shifts the attentional focus to the issues called to mind by the observed data (interrupt- or data-driven processing). Attentional focus is partly determined by ongoing processing such as looking for the information needed to explain an unexpected finding or the information needed to generate responses to correct perceived abnormalities (knowledge-driven processing). An example of knowledge-driven processing is when one decides to look at data on the letdown system, pressurizer relief valves, and containment status because one is trying to find

an explanation for an unexpected decrease in pressurizer level. An example of data-driven processing is when an automatic safety injection signal occurs, an operator interrupts a search to determine what caused an automatic reactor shutdown and goes to verify proper automatic systems responses to the safety injection signal.

Another critical factor in behavior is resource/workload interactions. Processing activities have associated costs. Carrying out one type of processing precludes the possibility of doing other processing when there is competition for limited resources. Thus, a need to choose which processing activity should be carried out next can arise -- acquire more data, pursue possible explanations, generate/track responses to detected abnormalities?

A major constraint on whether an individual processing act is carried out and how well it is done is the degree to which mental resources are already occupied. This is a function of how much processing resources are available and how many processing activities are competing for those resources. A processing activity may need to be suspended or dropped when new findings demand unavailable resources. Alternatively, the implications of a new finding may not pursued. Resource competition implies there is some agenda of processing items that could be chosen to carry out next and a method for selecting among these competing items. The degree to which resource competition (new findings exceed the available resources) occurs depends on a variety of factors -- the incident partly determines the rate of possibly interrupting stimuli, the amount of resources available can vary, the time it takes to carry out selected processing is important because new events may occur before the ongoing processing has been completed (relative pacing between CES processing durations and the time evolution of the incident itself).

A limited problem solver should focus first on *"interesting" findings.* The question, then, is what is an interesting finding. There are several definitions of what is interesting that control processing.

First, of all the things that could be observed about plant state, what occurrences are worth noting? For continuous variables, departures from target regions, limit crossings, direction of movement (up, down, stable), rate of change (e.g., slow, fast, stable), or more complex behavior patterns (such as decrease to a low equilibrium, increase to a maximum value and then decrease) are all interesting plant behavior. For processes, systems or components, interesting state changes include active-inactive, should/should not be active, available-unavailable. If an "interesting" occurrence is observed, then there is a need to monitor and track the item in question.

This level of criteria leaves a potentially large pool of "interesting" findings that could be pursued. One filter is a competition for a limited attentional resource – is a potential findings salient? has its status been requested? If a potential finding is observed – is it abnormal and is it expected? If it is abnormal, then there is a need to pursue how to correct it or how to cope with it? If it is unexpected, then there is a need to pursue what could account for or explain it?

Another layer of filtering may be needed because there can be several unexpected findings to explain and several abnormalities to correct. Strategic choices may need to be made to decide whether to respond to current perceived abnormalities or to build deeper explanations for findings and therefore to select more effective responses, i.e., the relative priority between monitoring, explanation building, and response management activities. Goal or consequence information may be needed to chose which abnormality to deal with first.

Depth of processing is controlled in part by the layers of answers to the what is interesting question. If nothing interesting is perceived in some area of the plant, then no processing effort is devoted to that area. As interesting things begin to happen or to be observable, then more processing effort should be allocated to that area, if it is available. When there are numerous processing activities which could be pursued, then items need to be scheduled to be carried out. For example, building an explanation for unexpected findings may have to wait until responses to cope with perceived abnormalities have been selected and correct execution verified.

We will now examine each of the basic competencies. It is important to note that in specifying the desired cognitive competencies we are also specifying the target performance adjustment factors for CES. These factors will determine when CES exhibits competent performance and when it is prone to cognitive failures. The cognitive competencies specified for CES are:

- Control of Attention,
    - data-driven control of attention
    - competition for limited resources
    - evidence processing,

- Situation Assessment
    - expectations
    - qualitative reasoning
- Explanation Building
- Response Management
- Knowledge Representation

3.3 Monitoring and the Control of Attention

A basic operator activity is monitoring plant state, that is, checking available data in order to detect interesting changes that may have taken place. The key here is the mechanisms that determine *what* data or plant issue are checked *when*. The relevant questions are what governs where CES looks next? Given it looks, what does it see? Given what is observed, what knowledge is evoked?

The limited problem solver assumption means that CES can only process a limited amount of the incoming plant data "at one time." The amount of input data and, more importantly, which input signals it processes can be thought of as a field of attention. The size of the field of attention to changes in plant data will likely be smaller than the elements competing for attention. Which evidence will be processed is determined by the cycle or interaction between knowledge-driven and data-driven processing, how much processing resources are available and how much demand there is for those resources, criteria which define what are "interesting" or "important" findings. (a composite function of factors associated with a piece of data such as observability, interruptability, time value of information/time since last observation).

3.3.1 Data-driven Control of Attention

Data can be monitored when knowledge-driven processing is interested in the state of some part of the plant. One major example of this is explanation building activities where plant data are checked in order to account for unexpected findings (see Sections 3.5 and 4.2.2 on explanation building). The other major example of knowledge-driven search occurs in monitoring to check whether desired responses (either automatic or manual) have occurred.

However, salient data can interrupt ongoing processing and capture the focus of attention. *Interruptability* is a function of two factors: how strongly a datum shouts for attention or *perceptual salience* and how receptive one is to incoming data in general or to incoming data on a particular issue or *receptivity.*

Perceptual salience is a function of how much the physical signal commands attention (signal strength) given the level of background activity (noise). An example of signal strength is that associating an auditory signal or a blinking visual signal with a state change or limit crossing increases the salience of that change. An example of the level of background activity is that a particular auditory signal is more salient against a quiet background than against a background of many other similar auditory signals.

Another reason to check the state of some signal is periodic sampling of available data in order to detect interesting changes that may have taken place. This can be modeled through time dependent functions on salience on plant signals that operators periodically sample.

Receptivity is a function of how "important" is the piece of data in the current problem-solving context. If one is trying to account for an unexpected level decrease, then one is more sensitive to incoming data that help refine what accounts for the unexpected behavior. Thus, issues that come to mind affect the *level of interest* in a piece of data. It also depends on the *current workload.* Interruptability depends on what other processing activities are competing for limited resources.

One can think of the interaction of these factors as a threshold mechanism. If the salience of a signal exceeds the threshold, then the signal is observed and placed on an agenda for processing. However, receptivity affects the threshold so that interest in an area means that lower salience signals will exceed threshold.

There are a variety of possible interruption scenarios: a signal may fail to interrupt ongoing processing, the new signal may be processed to check for "interestingness" and control returned to the previous line of processing, or the previous line of processing may be dropped completely and a new line, prompted by the new signal, taken up. Notice that if one is too sensitive to incoming signals (too interruptable), then processing of one issue may not be finished before jumping to another and another. This models one source of errors -- vagabonding (cf., Dorner, 1983). If it is too difficult to interrupt ongoing processing, then one may continue to work on the current issue regardless of important new signals which should shift attention. This setting models another source of errors -- fixation. There have been several cases in

NPPs and other complex worlds where fixation on one signal has retarded observation of other important aspects of system state and contributed to an accident (the most notable case of this is the Everglades L-1011 plane crash; National Transportation Safety Board, 1973).

It is important to be able to control whether a particular plant behavior can trigger or interrupt processing (i.e., is data-driven processing of a signal possible). If changes in a signal cannot or do not trigger processing, the plant behavior may still be observed, but only for a knowledge-driven reason (i.e., to check on desired corrective responses or to narrow possible explanations of an unexpected finding).

The inverse of the question, "when to look," is the question, "when to stop looking." There are a variety of criteria for deciding when monitoring of a data channel stops, for example, a time out or time decay function unless or until there is a stimulus to continue monitoring the datum in question.

3.3.2 Competition for Limited Resources

A competition for limited resources implies an agenda of processing items which could be chosen to carry out next and a method for selecting which to do next. The need to choose which activity to suspend, drop, or not pick up occurs when new findings exceed the available resources. For example, when monitoring demands exceed capacity, which items will be dropped from a monitoring agenda -- the oldest item, the least important, at random?

The degree to which competition occurs depends on a variety of factors -- the incident partly determines the rate of possibly interrupting stimuli, the amount of resources available can vary, the time it takes to carry out selected processing is important because new events may occur before the ongoing processing has been completed (the relative pacing between CES processing duration and the time evolution of the incident itself).

As a result, the concept of effort can affect processing. For example, one psychologically plausible rule is try to minimize resource expenditures unless the situation demands it. This rule would lead one to look at easy to acquire data first, or only to process a minimal set of evidence when checking the state of some part of the plant -- i.e., to use an evidence processing shortcut. Only if there is some reason to suspect that the shortcut is insufficient, e.g., an unexpected result occurs, is more thorough processing invoked (assuming the knowledge needed to support more thorough processing is in the knowledge base). The concept of effort is one reason to be able to vary the depth of processing in CES.

The level of effort may also vary during incidents because resources are not fixed but are elastic -- they stretch as a function of the demands placed on the system. This means that the level of resources would not be fixed; instead, there would be an elastic maximum capacity to monitor data which varied with workload. When the processing demands are low because little is occurring in the world to be processed, thorough processing (depth of processing) may not occur because effort may be low (the resources available are low). As more interesting findings are noted (abnormalities, unexpected plant behavior) more thorough processing should occur (more resources should be brought to bear on those particular issues and general level of resources available increases). However, the thorough processing still may not occur because demands may still outstrip resources. When a resource limit is exceeded, there are two ways to cope: one could attempt each activity but do each less well or one could carry out fewer activities (cf., research on resource limited cognitive processing, e.g., Lane, 1982).

Time pressure is another potent factor that determines the level of resource competition. In part time pressure is a demand characteristic of the incident in question, i.e., the time available to perform tasks. Time pressure is also a of function processing resources, i.e., how efficiently or quickly processing is carried out. This can be modeled by varying the pacing (the relative timing) between CES processing cycles and incident time progression (how much information processing goes on for each step of time in the incident evolution, e.g., number of processing cycles). Increasing the rate at which the incident evolves, relative to the processing cycles required by CES to complete analysis, introduces time pressure and constrains depth of processing.

The behavior of CES changes as the level of resource competition changes. This depends on the nature of the incident itself, the level of resources available, and how resources are allocated to different types of processing activities. Competition can also arise between the three basic processing types: monitoring, explanation building, and response management. Specific events during an incident could trigger an emphasis on one processing type over the others. The most notable example of this is a reactor trip where operators are drilled to go through a monitoring phase with little effort spent on explanation building. Different operators (e.g., at different skill levels) could emphasize one processing type over others. For example, a skilled control room supervisor may place a high emphasis on obtaining feedback about whether desired corrective responses were carried out correctly. Relative emphasis across these three activities may also capture different kinds of emergency response strategies A general bias towards response management to the exclusion of explanation building represents weight on a disturbance management approach to incident recovery (e.g., function or

symptom based procedures) -- prevent propagation of disturbances; the reverse bias puts weight on spending resources to diagnose and to repair broken equipment or systems (e.g., the latter emphasis was observed in the Davis-Besse incident; NUREG-1154; Woods & Roth, 1986). The relative emphasis on these three types of activity could also shift dynamically during an incident. For example, early in an incident the emphasis might be on monitoring the state of the plant. Later in the incident the emphasis would then shift to explanation building. Following acceptance of a good explanation the emphasis would then shift to response management (but, for an ideal dynamic problem solver, note how the occurrence of a new fault should re-shift the emphasis).

3.3.3 Evidence Processing

Knowledge about data-state mappings stored in the knowledge base, i.e., how the available data function as *evidence* to answer questions about plant state is another factor in CES monitoring activities.

There is a many-to-many mapping between available data and the states of parts of the NPP. There are many pieces of evidence which testify to the state of some part of the plant (many-to-one). One piece of data may testify to the state of multiple parts of the plant (one-to-many). For example, what is the state of the pressurizer spray system for controlling high primary system pressure? The evidence available to determine an answer to this question includes demanded valve position, actual valve position, automatic system setpoints on high primary system pressure, primary system pressure, the status of reactor coolant pumps, whether there is a steam space in the pressurizer. In different situations created by various process faults, sensor failures and automatic system failures, different parts of this evidence set may be critical in determining the actual situation.

An issue that arises can trigger the need to evaluate evidence on the state of that part of the plant -- knowledge-driven processing. For example, if one suspects that there might be a break in a pipe, then he, she or it would check to see if there is any evidence which points to the presence of abnormal amounts of water or energy in the container around the pipe (e.g.,if this is the containment, the relevant set of evidence includes containment sump level, high humidity, high temperature, high radiation).

When a particular piece of data about plant conditions (which can be seen by an operator in the hypothesized control room of interest) is checked, what could be called to mind depends on what that datum means about the state of some functional or physical part of the plant. Take for example, an auditory alarm sounds which indicates that a valve has closed. Most simply,

this message signals a component status. If this piece of data is combined with other knowledge about the valve or the current context, the change in status may be expected or it may be abnormal. If it is abnormal and depending on other knowledge, it could signal that a dormant process is unavailable if needed (flow is prevented) or that an active process is disturbed (there is no flow). This is an example of data-driven processing where an observed signal triggers the activation of other knowledge.

There is always uncertainty in the conclusions that could be drawn from observed data -- the sensor system could have failed, the display can be misread, only portions of the relevant set of evidence may be checked, a single fault or multiple faults may have produced the pattern of evidence. Uncertainty of a state assessment is related to attributes of recency, completeness, consistency, strength.

Because the NPP is dynamic, the value of evidence obtained from a plant parameter reading as an accurate indicator of plant state decays as a function of time. *Recency* refers to how up-to-date is the data (or the state computed from it). The rate at which old information decays or loses its value (the data half life or the time value of information) can be expressed as a function of the time since the data were last gathered and can vary for different data. There are a number of mechanisms for computing the relationship between information worth and the time since last observation which basically reflect the extent to which the last observation is a good predictor of current state (cf., Moray, 1986).

Completeness refers to whether the state (a) was assumed based on inferences or expectations, (b) was computed from a single datum, or (c) was computed from multiple data (Figure 3-1).

The question of *consistency* arises if multiple indications are used to determine a state. Multiple pieces of evidence can reinforce the same result (consistent evidence) or they can indicate different answers (inconsistent evidence). If the evidence is inconsistent, the difference may be resolved and a single answer passed on (with the appropriate basis indicated), or the difference may not be resolved in which case the implications of all possible answers may need to be examined.

Strength of relationship refers to how strongly a particular piece of evidence testifies to the state of an entity. A simple example is that indication of the demanded valve position is weaker evidence of valve position than an indication of actual valve position. Other examples are the reliability of the data channel (e.g., sensor failure as an explanation is more plausible if there is only a single as opposed to redundant sensor channels) and its precision

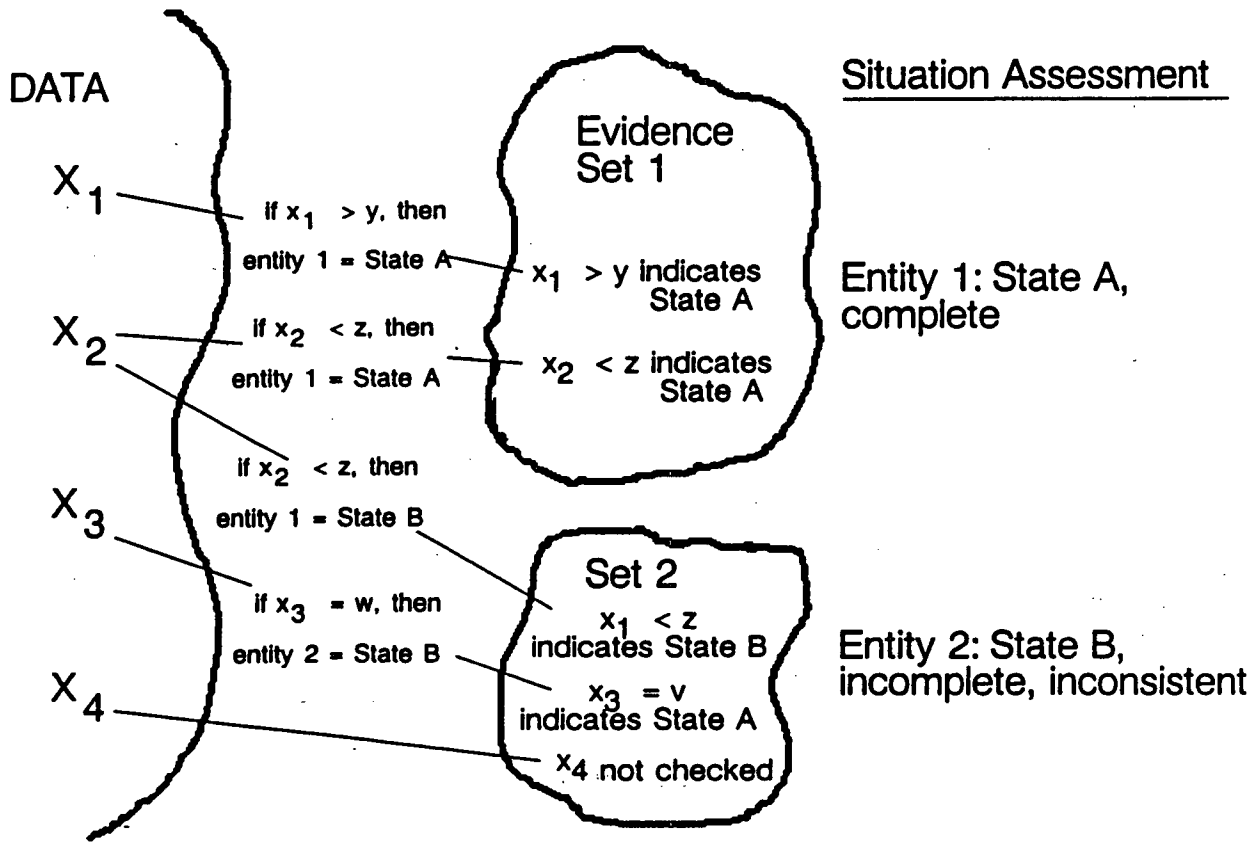


Figure 3-1: Evidence Processing. Given interest in an entity, CES must be able to examine the evidence set and corresponding data to compute the state of that entity and, given a new datum, must be able to establish what that value signifies in terms of the entities that it is evidence for.

(e.g., the error bounds on many measurements varies depending on containment conditions). In the Oyster Creek incident analyzed in Pew et al. (1981) operator difficulties were due, in part, to a lack of understanding of the different strength of evidence values of different indicators of the state of a critical plant function.

3.4 Situation Assessment

Situation assessment refers to the perceived state of the NPP at a point in the evolution of an incident. The situation assessment at some point in an incident in CES is the current results of ongoing processing activities including what plant behaviors have been observed, which systems or processes are perceived to be active, which expectations have been confirmed, what observations are considered unexpected, what observations are seen to be abnormal, what explanations have been established. It also includes knowledge about the current state of the problem-solving process, such as what items are currently being monitored and what explanations are under consideration. Note that these results are distributed over the individual processing activities. This is part of the organizational view of the mind.

Knowledge about intermediate results and ongoing activities can be posted and made available to different processing mechanisms. For example, when an explanation is established for an unexpected finding, knowledge about automatic or manual responses appropriate to that situation is evoked and can play a role in other processing activities, e.g, monitoring to check whether the response has been executed correctly. Thus, situation assessment is a virtual repository of information on the current, perceived state of the incident.

However, communication of processing results and ongoing activities among parts of CES may be restricted. This is especially important because it allows the simulation to capture situations where multiple people or facilities are involved in problem-solving (see Section 2.6).

3.4.1 Context Sensitive Reasoning, Abnormal Plant Behavior, and Unexpected Plant Behavior

An operator notices that steam generator level is a particular value. Is the state indicated by this value "abnormal?" One answer to this question is that the state is abnormal if the actual value has departed from some target value or region (e.g., "steam generator narrow range level should be at x% post reactor trip") or has crossed into an undesirable region (e.g., "steam generator narrow range level should not be less than y% otherwise heat

removal through the steam generators is impaired"). In principle, if level has departed from its target region or crossed into an undesirable region, operational personnel should drive level back towards target (e.g., increase feedwater flow) or supervise automatic system responses that do this if there are any (response management activities). If level has departed from its target region or crossed into an undesirable region and operational personnel know of no active process which would explain this observation, then the staff should pursue possible explanations that would account for this observation (explanation building activities).

However, the interpretation of the level value, and therefore what is an appropriate response to it, becomes more complicated depending on the particular context. To understand the kind of complications that arise and their consequences for cognitive processing in the NPP world, let us consider several cases where the steam generator level value is observed to be less than y%.

In Case 1, the reactor has just tripped off. Steam generator level is observed to be less than y%. However, the operational personnel know that steam generator level is always low for some period of time following a reactor trip because of the thermodynamic consequences of rapidly stopping the reactor. The operators do not pay further attention to this observation or to alarms that signal level is low (and given the assumption of limited problem-solving resources, they should not expend resources on such an observation). In other words, the operators "expect" level to be low immediately following a reactor trip, and the absence of this behavior would be "surprising", i.e., interesting to pursue further.

In Case 2, the operator thinks that the steam generator is broken (e.g., a break in the feedline to this steam generator). Given this situation assessment, the observation of level less than y% is a consequence that follows from the perceived fault. Again, the operators "expect" level to be low given a feedline break, and corrective responses evoked by level less than y% do not apply to this situation. What would be surprising (i.e., interesting) is the absence of this behavior, given the situation assessment of a feedline break. Ideally, the absence of the expected behavior implies that either the situation assessment is erroneous (there is not a feedline break) or there is an additional factor influencing steam generator level besides the feedline break.

These cases illustrate the requirement for a high degree of *context sensitive reasoning*, if one hopes to capture the problem-solving environment that is NPP emergency operations (actually any dynamic reasoning situation with limited resource problem-solving agents). These cases also point out the

critical need to distinguish *abnormalities or plant disturbances*, which are mismatches between the actual state and the desired state of parts of the plant, and *unexpected findings or discrepancies*, which are mismatches between perceived state or anticipated state and newly seen evidence (cf., Woods, Elm & Easter, 1986; Woods & Hollnagel, 1987). Failure to clearly distinguish these two concepts has greatly impeded study of fault diagnosis behavior and retarded the development of fault diagnosis systems and aids.

A critical cognitive competence for CES is to be able to judge whether observed plant behavior is expected or unexpected. Detection of an unexpected finding starts explanation building or explanation revision in order to account for the discrepancy. Thus, unexpected findings are a deeper definition of what is interesting plant behavior which should attract processing resources. Lack of sensitivity to discrepancies is one factor that contributes to fixation prone problem-solving.

Another error form that can arise from computing expectations occurs when an expected states is substituted for actual observation of the current state and used as the current state by other processing mechanisms. This can be efficient or necessary when resources are limited, i.e., strongly expected plant behaviors are of less interest relative to other items competing for limited attentional resources. However, it is also a source of observation failures that begin to lead processing astray.

In evidence processing activities, strength of expectation can interact with the strength and completeness dimensions of evidence. Less evidence or weaker evidence of some state may be sufficient to conclude that state if there is a strong expectation.

3.4.2 Influences on Plant Behavior, Observable Plant Behavior, and Qualitative Reasoning

How can expectations be generated? Various states of the NPP (pressure, energy, temperature, material inventories) are affected by various plant processes (material balances, energy balances, material flows; cf., e.g., Gallagher et al., 1982; Rasmussen, 1986; Woods & Hollnagel, 1987). Qualitative judgments about the behavior of these entities can be made based on knowing what processes or *influences* are currently active and dominance relations between these processes. There are a set of influences that could affect a given parameter. If an influence is active, then it will act to affect the parameter in question. How the parameter actually behaves depends on the combined effects of the *set of active influences* and not on the basis of the effect of a single influence. This is an important point: concluding that an influence is acting on a part of the NPP does not mean

that the observable behavior of that part of the plant will be consistent with the active influence. The distinction between influences acting on a part of the plant and the observable behavior of that part of the plant is critical in being able to generate expectations, build explanations, and handle multiple fault situations.

How are qualitative judgments about the behavior parts of the NPP made from knowledge about active influences? In the simplest case where there is only a single active influence, inferring expected behavior is straightforward -- if net charging inflow to the primary system is positive and there are no other influences acting on pressurizer level, then level should be increasing. If pressurizer level is observed to be decreasing, not increasing, then either net charging is not positive, the observed behavior is erroneous (failed sensor) or there is an *additional unknown influence* or influences acting on level that dominate the charging effect. Note that the influence relationship is deterministic; all of the uncertainty is displaced to questions about what is the set of active influences. When an unexpected plant behavior is noticed, deeper processing of this finding is called for -- explanation building (e.g., is there an unknown influence that would account for this behavior?)

Knowing about what influences are acting on a parameter allows only qualitative reasoning about its resulting behavior: determination of high information value landmarks in the behavior of continuous parameters (e.g., expected direction of change or decreasing to a new equilibrium point) rather than a complete statement of quantitative continuous functions. Consider the case where a new influence begins to act on a parameter that is currently increasing in value. Effective qualitative reasoning would allow one to know whether or not the new influence is sufficient to cause the parameter to reach a maximum and begin to decrease, whereas quantitative models would allow one to know the time and maximum value as well.

Qualitative reasoning is straightforward when there is only a single active influence or when there are two or more active influences that have the same kind of impact. Qualitative reasoning becomes much more difficult when there are multiple active influences which have different impacts. If one influence is to increase level and another is to decrease level, what will be the resulting level behavior (up, down, stable)? Or if two influences act in one direction but another one's influence is in the other direction, what will be the resulting behavior -- do two influences always dominate one influence? The difficulty is due in part to the need for information about "degree of influence."

This qualitative reasoning bottleneck can be dealt with in several ways. First, dominance relations can be directly entered into the knowledge base. Second, dominance relations can be learned. If the pattern of influences is ambiguous with respect to expected behavior, then the resulting behavior can be observed and remembered as the result of this pattern of influences for recall when this pattern is observed in the future.

However, knowledge of dominance relations will not be sufficient to characterize behavior in all possible patterns of influences. But operators clearly do not run mathematical models in their heads to calculate degree of influence. How do skilled operators avoid this bottleneck? The answer is that they do not try to directly estimate degree of influence; they let the evidence on continuing process behavior tell them dominance.

3.5 Explanation Building, Revision and Fixation

Evidence processing and other processing activities can generate a set of significant findings (violations of expectations) to be explained or accounted for. What is a good explanation that accounts for the current pattern of significant findings? For example, what state accounts for primary system level decreasing, primary system pressure slowly decreasing, and a very high charging flow? A loss of coolant break to containment is one possibility that may come to mind given this pattern.

Because the NPP is dynamic, the need to revise explanations also occurs. Breakdowns in this process (failures to detect discrepancies, failures to revise an explanation) are called fixation errors and are a major human error form in studies of NPP problem-solving (e.g., Woods, 1984). What evidence (unexpected findings) is sufficient to abandon or reconsider an accepted explanation? For example, assume the above signs and explanation. Then new evidence is observed that the primary system level and pressure decrease stops at a level of 11%, letdown is isolated, and charging flow remains extremely high. Does this evidence produce a revision of explanation (to a charging line break downstream of the flow sensor)?

Another critical capability for CES is reasoning in multiple failure situations. Given the possibility of multiple failures (actually the probability of multiple failures in serious real incidents), there can be multiple disturbances/discrepancies that have a single source or can be due to multiple sources. Thus, there are different paths to build towards a satisfactory explanation:

1. assume there is only one explanation for all significant findings -- severely fixation prone,

2. start with assumption of only one explanation; if it doesn't work, then consider multiple explanations -- liable to fixate,

3. start with possibility of separate explanations for each significant finding (assume multiple explanations first); then try to converge on fewer explanations -- fixation resistant but extra processing is required (convergence of multiple views) when the explanation is straightforward.

An operational definition of what constitutes a good explanation (e.g., the simplest explanation or parsimony) affects the process of explanation building by determining when a single explanation assumption should be abandoned or by determining how to orchestrate and converge if possible when considering multiple views.

Operator actions depend not only on building explanations to account for patterns of findings, but also on *commitment* to an explanation. Commitment refers to when an explanation is communicated to begin selection of appropriate corrective responses. A problem solver can err by committing too quickly (examine too little evidence) or too late (examine too much evidence or wait too long for more evidence to accrue). Varying the criteria governing commitment can affect how much evidence search will go on during explanation building.

3.5.1 Multiple Explanations; Alternative Explanations

There is a need for some mechanism to deal with the fact that there can be multiple explanations for a set of findings and alternative explanations for a set of findings. One concept is to think of an *explanation set* and a dimension of explanation where the number of explanations in the set ranges from a maximum which is equal to the number of findings to be explained to a minimum of one, or

$$1 \leq \text{number of explanations} \leq \text{number of findings-to-be-explained.}$$

The concept of an explanation set means that more than one explanation may be needed to account for all of the current set of findings and that there can be more than one way to map findings into explanations (Figure 3-2). The kind of explanation in the set can vary from a simple restatement of the perceived abnormality (a disturbance such as high pressure in a volume) to a fault category such as a loss of coolant accident (LOCA) or a response strategy category such as post-LOCA cooldown. More "integrated" explanations account for larger numbers of the findings-to-be-explained. The contents of the explanation set can vary in both the number and kind of

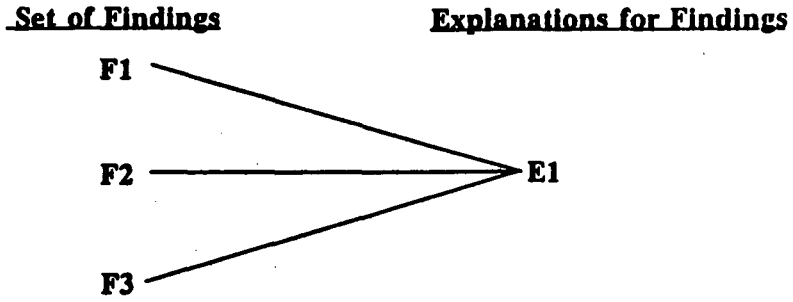
explanations as an incident evolves. The criterion for what is a good set of explanations can be modified to bias explanation building in different directions, for example, to prefer fewer but more integrated explanations when considering alternative ways to account for a set of findings.

This concept of an explanation set that is made up of a sliding number and kind of explanations provides a framework that integrates both function based and event based approaches to incident response. When the number of explanations used to account for findings is towards the maximum, the interpretation of plant state is function based. When the level of explanations is towards the minimum, the interpretation of plant state is fault oriented. The former produces more flexible responses; it can generate useful partial solutions and responses without waiting for a complete accounting for the situation; and it is better suited when the level of available data and knowledge about the plant supports only weak assessments of causes and response strategies. It is how function or symptomatic procedures tell operators to think. The latter supports stronger expectations about future plant states and responses; it is better suited for maximum efficiency or efficacy of response; and larger response strategies are selected rather than built up from the selection of elemental responses. However, stronger explanations lead to more brittle performance if the explanation would be wrong or incomplete. It is how event oriented procedures tell an operator to think.

CES must try to build the "best" set of explanations to account for the perceived set of findings. For successful explanation building it must be able to consider alternative interpretations for a set of findings at one point in time and over time as new evidence is observed. For erroneous explanation building to occur, under some conditions it must fail to consider possible alternative explanations, again, at one point in time or over time as new evidence is observed.

Building explanations to account for observed findings depends on access to knowledge about how plant behavior reflects different underlying plant states (plant behavior-plant state links in the knowledge base). The possibility of having explanations in terms of the perceived abnormality or disturbance itself (e.g., high pressure in a volume) or in terms of a fault or response strategy category (e.g., a loss of coolant accident) means that symptoms (i.e., observable plant behaviors) can be linked directly to fault or response strategy categories as in most diagnostic models and systems, or indirectly through the intermediate categories of different types of disturbances. This structure can support symptomatic diagnostic search behavior or topographic diagnostic search behavior or a mixture of both (Rasmussen, 1986) depending on the structure of the plant behavior to plant state links in the knowledge base.

Case 1:



Case 2:

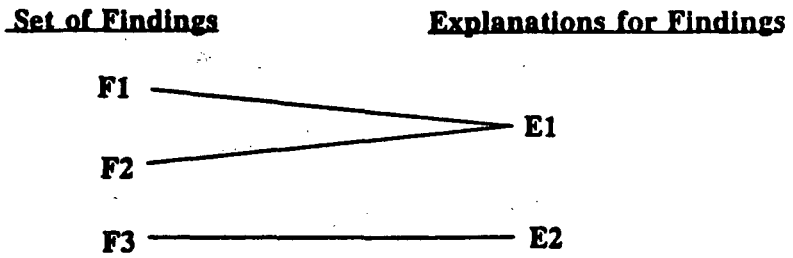


Figure 3-2: More than one explanation may be needed to account for all of the current set of findings and that there can be more than one way to map findings into explanations. In this example the same three findings can be explained either by a single explanation for all three (E1 in Case 1) or by using one explanation (E1) to account for two of the findings and a second explanation (E2) to account for the third finding. The kind of explanation included in the explanation set can vary from a simple restatement of the perceived abnormality (a disturbance such as high pressure in a volume) to a fault category such as a loss of coolant accident (LOCA) or a response strategy category such as post-LOCA cooldown. More "integrated" explanations account for larger numbers of the findings-to-be-explained.

The explanation set structure supports the need for changes in explanations that can occur during the evolution of an incident. For example, when there are ambiguities because the evidence for the actual underlying fault(s) emerges slowly over time, diagnostic behavior can follow a progressive refinement process. The nature of the trouble is initially characterized at a high level (e.g., there is enough evidence to recognize some form of loss of water inventory from the primary system), and the situation assessment becomes more refined as more evidence is gathered or becomes available (the loss is due to a steam generator tube rupture but the secondary radiation alarms were missing for some reason). Woods, Wise & Hanes (1982) contains decision protocols of operator performance in simulations of these kinds of incidents that exhibit this progressive refinement diagnostic behavior.

One consequence of attaching explanations to subsets of findings is that the explanation, as part of the situation assessment, can be linked to knowledge about plant behavior (state-plant behavior links in the knowledge base). This kind of knowledge can be used to generate *expectations about current or future plant behavior.* Observing these behaviors (through scheduling of the requisite monitoring activities and subject to the control of attention mechanisms) tends to confirm the explanation. Failures to observe these behaviors, or *negative findings*, weaken belief in the explanation and indicate the need to revise the set of explanations -- an explanation may be erroneous or additional explanations may be required which in total account for the complete pattern plant behavior.

Explanations can also be linked to sets of corrective responses. Different kinds of explanations may be linked to different grains of pre-planned corrective responses (see Section 3.6 on response management). For example, more integrated explanations such as a fault category will generally be linked to larger, more integrated sets of responses. This introduces the question of when has sufficient explanation building gone on to accept an explanation and to invoke the corrective responses linked to this explanation. This involves deciding whether sufficient evidence has been gathered (e.g., Cohen, 1987). Both moving too quickly or taking too long to accept provisional explanations can lead to intention failures. Notice that an explanation can be accepted independent of accepting an entire explanation set. This provides a constraint on further explanation building because a subset of the findings have now been accounted for.

3.5.2 Temporal Factors in Explanation Building

One important challenge in building explanations is the role of time. Time affects explanation building in several ways. First, there may be only weak evidence for an explanation because more time must elapse for additional

evidence to accumulate or because it is the wrong explanation. For example, a sensor value that is noticed and processed (say pressurizer relief tank -- PRT -- temperature off target and increasing) may be evidence for a disturbance in a plant entity (PRT as a sink for water and energy has abnormal increasing contents). This generates expectations that there is other evidence for PRT contents increasing (e.g., PRT pressure off target and increasing). If the simulation looks to see if this is the case but sees PRT pressure normal, then a data inconsistency and discrepancy (violated expectation) exist. A variety of possibilities exist to explain this: the PRT might be normal and the temperature evidence due to a failed sensor; the PRT might be abnormal but either too little time has elapsed for pressure to show this or the pressure sensor is failed. Which of these outcomes results or whether the simulation even tries to resolve the discrepancy depends on the perceived state of the processes in which the PRT plays a role and the agenda of items competing for the simulation's resources (i.e., workload), e.g., is there an expectation that the PRT is active (a process in which it participates is active) or should the PRT be active in the present context (primary pressure has been high).

Because the NPP is a dynamic world, negative findings, the failure to observe an expected plant behavior, can be detected only if there is a time function that defines the window of opportunity for observing the expected behavior.

Another complication introduced by time is that disturbance propagation, automatic systems, or operator action may remove part or all of the evidence that led to earlier explanations. Thus, the account for current findings must include elements of past explanations that may not be repeatable given current evidence. For example, at one time step decreasing primary system pressure and level and indications of abnormal containment conditions are used to conclude the existence of a primary system break to containment. At a subsequent time step, after emergency core cooling system (ECCS) actuation has occurred, primary system pressure and level are increasing and there are abnormal containment conditions. The simulation needs to use the past explanation that there is a break to conclude that ECCS inflow exceeds break outflow and that reducing ECCS inflow may be appropriate. Note that multiple explanations can be needed to account for the change in a finding over time. In the above example, primary pressure and level behavior are explained by the conjunction of a primary system break and ECCS flow greater than break flow.

Another aspect of time is time pressure. Time pressure is one kind of resource limitation as was discussed earlier. With respect to explanation building, when there is time pressure, it may not be possible to completely

assess all potential explanations that could account for a finding, to converge on a single explanation when considering multiple views, or to revise a previously accepted explanation.

The problem-solving agent can be aware of the time available to carry out tasks before negative consequences propagate (to a greater or lesser degree of specificity). Operators can consciously use this kind of knowledge in deciding how to respond in some emergency situations (e.g., the Davis-Besse incident; NUREG-1154). This means the knowledge base will need to include information about the time course of plant processes, the time course of disturbance propagation, and time requirements for recovery tasks.

3.6 Response Management -- Plan Selection, Monitoring, and Adaptation

Response management deals with

- how pre-planned responses that are relevant to the perceived situation are called to mind,
- how these response sequences are monitored to detect execution errors or failures of equipment to respond as demanded,
- how pre-planned responses are adapted or gaps filled in when unusual situations occur.

3.6.1 Plan Assembly

Situation assessment judgments (observed findings, explanations) trigger or call to mind knowledge about how to respond to this perceived situation (i.e., one kind of information in procedures). These can be responses both to disturbances (e.g., decreasing pressurizer level) and to faults (e.g., steam generator tube rupture). Response sequences include both automatic system responses and manual human actions. There can be sequences or stages of response over time, for example, as a disturbance grows, stronger and stronger counter measures. For example, the script for a decreasing progression in pressurizer level includes "increase/maximize net charging inflow" {"increase charging flow (automatic)," "start additional charging pump (optional manual)," "stop letdown outflow (automatic)"} and, if this response is insufficient, "switch to emergency injection (automatic)."

The basic process for forming intentions to act is plan assembly from pre-stored plan chunks guided by the current assessment of the situation and

accepted explanations. This is because emergency operations is a heavily proceduralized world. Situation assessment and explanation building control selection among chunks of pre-planned responses which are stored in the knowledge base. The type of response (function restoration or response to fault categories) depends on the grain of the response chunk that is linked to a particular perceived situation in the knowledge base. Judgments about what and when responses should be made can be output from CES to another mechanism which executes the action on a simulated plant or which determines how this action would affect the behavior of the NPP.

In the plan assembly process, the same corrective response may be in the knowledge base in several places, i.e., the same response is linked to the different situations for which it is an appropriate corrective response. Because responses are triggered from situation assessments, the different states linked to a response or to a response chunk that contains this response represent the different paths by which this knowledge could come to mind. This mechanism provides for the possibility of inert knowledge about corrective responses -- the case where there is no link from the perceived situation to what, in hindsight, is the appropriate response (cf., Bransford et al., 1986 and Perkins & Martin, 1986 for discussions of the cognitive processing strategies and the problem of inert knowledge). In other words, even though this response is located somewhere in the knowledge base (the operational staff "know" the response in some context) it is not called to mind in this situation. This is critical for CES to be able to model operational staff behavior in incidents like the Oconee steam generator tube rupture (Brown & Wyrick, 1982) and in the LOFT L2-5 test reactor experiment (Bray, 1982; Bayless & Divine, 1982) where potentially relevant corrective responses were not thought of given the staff's perception of the situation.

3.6.2 Plan Monitoring

CES as a supervisor of other agents, both human and machine, needs to be able to expect when responses should be made and to see that they occur correctly. If a response is expected, then there is a need to monitor to see if the response has happened (e.g., did letdown isolate?) and if the new influence has affected the behavior of the datum in question (e.g., is level now stable or increasing). Notice that the need to explain why level was decreasing in the first place remains (what level influencer exists that is not in the current set). The level behavior following the addition of a new (or stronger) influence from the response script (assuming that the response was actually made) provides more information for explanation building (did the response counteract or dominate the unknown influence or does the unknown influence continue the level decrease). If the unexpected behavior continues unabated, one possibility is that the expected response was not properly

carried out. In this way CES is capable of detecting human execution errors, failures of automatic systems, and failures of equipment to respond as demanded.

3.6.3 Plan Adaptation

Complications can arise in situations which go beyond pre-planned responses sequences. Operational personnel sometimes face circumstances where they need to (or think they need to) improvise action sequences that are not the nominal response sequence in the procedure which hindsight suggests was most appropriate. In these kinds of cases, choosing a response may require more processing than the simple response assembly process.

Two or more items in the current explanation set may be linked to incompatible intentions (e.g., the preferred response to x conflicts with the preferred response to y). Incompatibilities also occur when an intention generated by one explanation in the explanation set, if acted on, would create a disturbance (an anticipated disturbance) given the current situation assessment. Note that the latter implies a mechanism to envision the consequences of the intended action before it is sent for execution. Usually, the incompatibility can be resolved by choosing a response that satisfies both (perhaps a less preferred but acceptable response) or by making multiple responses (a second response is needed to prevent a negative side effect of another response). A mechanism is needed to resolve incompatibilities by finding an intended response that will satisfy both.

Sometimes adapting responses goes beyond simple planning to include choice under uncertainty and risk. For example, if no response can be identified to resolve incompatible options, then one must determine how to sacrifice one (e.g., as occurred during the Ginna incident; Woods, 1982). These choice situations include cases where one must decide whether to act on the basis of the current explanation for observed findings or whether to gather more data (which often means wait for more information to accrue) and cases where one must decide whether to attempt to diagnose and repair the disturbed system or process versus whether to implement alternative processes in order to cope with the consequences of the disturbance (disturbance management to prevent disturbance propagation in the short term) as occurred in the Davis-Besse incident (NUREG-1154). These choices can be based on knowledge about requirements and obligations associated with actions on processes, side effects of an action in the current context, the exposure of different operations (e.g., operators can be reluctant to implement a higher exposure to risk alternative), the relative impact of alternative actions on different goals, and goal priorities. This type of reasoning to adapt plans also requires the capability to look ahead to envision the consequences of potential action choices.

3.7 Representing Knowledge About the Plant

Knowledge about the NPP must be represented and made available for the processing mechanisms in CES. The knowledge base expresses relationships between NPP entities. This must include different kinds of relationships and different kinds of NPP entities (e.g., data, system states, parameter behavior, goal satisfaction, actions). For example, knowledge is needed about the mapping between potentially observable data and states of plant functions, systems, components and visa versa. CES monitoring activities use knowledge about what are interesting changes in a particular piece of plant data. Knowledge is needed about what observable behaviors occur if a plant function, system, or component is in a particular state. Knowledge is needed about how plant systems function to meet their design goals. Knowledge about how one entity (process) *influences* another entity (a parameter or a process) is needed to support the qualitative reasoning used to formulate expectations. Knowledge is needed that links different plant states to pre-planned response sequences.

Varying the kind of knowledge expressed in the CES knowledge base and the structure of the knowledge will drastically affect CES problem solving. Changes in the contents and organization of the knowledge base can be used to capture many aspects of NPP person-machine systems including much of training and procedures.

3.8 Representing What CES Can "See": A Virtual Display Board

The knowledge representation must also include a description of what data about plant state are directly available to CES to "see," reflecting what plant information would be directly available to operational personnel to observe. This description constitutes a *virtual display board*, that CES monitors to acquire data about plant state.

Changes in control board instruments, the organization and layout of control boards, computer-based displays of information, the organization of computer-based display systems, human-computer interaction can be expressed in terms of characteristics of the virtual display board. Not all of the dimensions needed to capture variations between different concepts and media for information display have been set up in the current implementation of CES.

One parameter is the *salience* of plant data which reflects the degree to which the *form* or *content* of a data point commands attention or can capture control of processing resources. For example, in terms of form, changes in a datum displayed as a meter in a field of similar meters are less likely to command attention compared to a representation of the changes as

blinking lights or audible sounds. For example, in terms of content, changes in some plant data are more "important" than others (e.g., better indicators of system state; an operator's idiosyncratic or favorite data points to check), such as changes in pressurizer pressure are much more salient than changes in pressurizer relief tank pressure, independent of the form of presentation.

A second parameter is the *observability* of plant data which expresses how easy it is to see its value or state. This represents factors such as the cost of acquiring data (physical or virtual distance), and the amount of processing of the display needed to extract the desired information. The greater the "cost" to the operator of acquiring the information, the more likely it is that the display will not be examined unless knowledge is activated that indicates that the information from that display is relevant to disambiguate a state or contribute to problem solution (i.e., knowledge-driven monitoring).

A third measure is *ease of integration.* This reflects the extent to which the displays in the control room collect and integrate the evidence available with respect to issues (i.e., the number of data elements and computations required to process directly observable data to reach a desired state conclusions). For example, to determine subcooling margin, is there a single display that provides that information (i.e., a one-bit decision) or does the operator have to compute the margin by taking readings from multiple sensors and performing some calculations?

The characteristics of the virtual display board contribute to CES monitoring activities by specifying what data are available to be sampled on plant state (typically, this is the sensors on display in the control room; however, it could include new computer-based displays or advanced interfaces, telephone/intercom communication with remote personnel) and the representational properties of the data such as observability and the cost or effort to acquire (e.g., physical distance, number of commands to call up a computer display, potential for misreading displays, the amount of processing needed to decide on a state).

4. Architecture of the Cognitive Environment Simulation

This chapter describes the current architecture of the Cognitive Environment Simulation – what are the processing mechanisms that exhibit the target cognitive competencies? In the process, this chapter also describes the current state of implementation of this architecture.

As an AI-based computer system, CES has two basic parts. First, it contains a *knowledge base* that represents the operator's (or the team of operator's) knowledge about the power plant, including the inter-relationships between physical structures, how processes work or function, goals for safe plant operation, what evidence signals abnormalities, and actions to correct abnormalities. Second, it contains processing mechanisms (or inference engine) that represents how operators process external information (displays, procedures) and how knowledge is called to mind under the conditions present in NPP emergencies (e.g., time pressure). This part of the model determines what knowledge is accessed when and what cognitive activities (monitoring, explanation building, response management) are scheduled when during an evolving incident.

The following sections describe the knowledge representation, the current processing mechanisms, samples of CES processing in different NPP situations, and the ways the program's behavior can be modified (CES performance adjustment factors).

4.1 Knowledge Representation

The basic unit of knowledge representation consists of a link between two pieces of knowledge. These *knowledge couplers* express a relation about the NPP (Figure 4-1a). For example, "pressurizer level" is coupled to "steam generator tube rupture." When either terminus (or node) is activated, the item it is coupled with is suggested (thus reasoning can flow in both directions). If "decreasing pressurizer level" is activated (e.g., a level decrease is observed from some instrument or display), then it suggests the possibility of "steam generator tube rupture"; and if "steam generator tube rupture" is activated (i.e., suspected or deduced), then it suggests "decreasing pressurizer level."

The nodes can represent potentially observable plant behaviors (e.g., "decreasing pressurizer level"), plant states that are inferred from observable data (e.g., "inadequate heat sink" or "steam generator tube rupture"), corrective responses (e.g., "automatic isolation of the letdown system" or

"manually isolate feedwater to steam generator x"). This means that couplers in the knowledge base can encode knowledge operational personnel would be expected to have about NPP data-state evidence links, state-state links, and state-response links.

The relation expressed in the coupler can be qualified by the state of some contextual knowledge (Figure 4-1a). Contextual knowledge expressed in a *qualifier* modifies the coupler link in two ways.

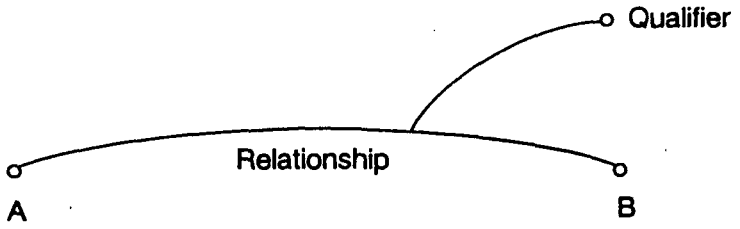
1. It specifies different conditions that vary the specific nature of the relationship between the two items. For example, the relationship between "pressurizer level" and "steam generator tube rupture" varies depending on "primary/secondary pressure differential."

2. It specifies the conditions under which the relation expressed by the coupler is relevant (one should not call to mind this relationship unless X; similar to the "unless clause" construct of Michalski & Winston, 1986). For example, the relationship between "primary/secondary pressure differential" and "pressurizer level" is not relevant unless "steam generator tube rupture" has been activated.

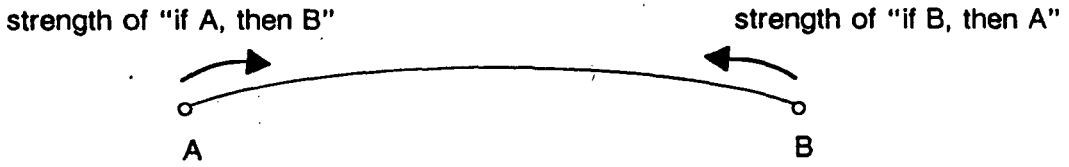
The qualifier is a powerful concept because it governs *under what conditions* one concept will activate another. Thus, it provides a measure of context sensitivity so that the phenomenon of "inert" knowledge can arise, that is, knowledge triggered only in very narrow contexts (e.g., failing to apply procedures learned for responding to situation A, to situation B which superficially appears different, but is fundamentally the same).

A coupler also encodes the *strength* of relation between the two items linked; that is how strongly the activation of one item calls to mind the other (Figure 4-1b). For example, associated with the coupler linking "pressurizer level" and "steam generator tube rupture" is a strength value that reflects how strongly knowing that pressurizer level is decreasing would bring to mind the possibility of a steam generator tube rupture. Since couplers allow bi-directional inference and since the strength of implication need not be identical in both directions, there are two strength values associated with each coupler, one for the strength of association in the each direction. For example, a steam generator tube rupture definitely is an influence to decrease pressurizer level, given a positive primary-secondary pressure difference, while decreasing pressurizer level by itself only suggests the possibility of a steam generator tube rupture because it could indicate other conditions as well. The strength parameters take on values between 1 and 5, with 5 indicating the strongest relation.

A.



B.



C.

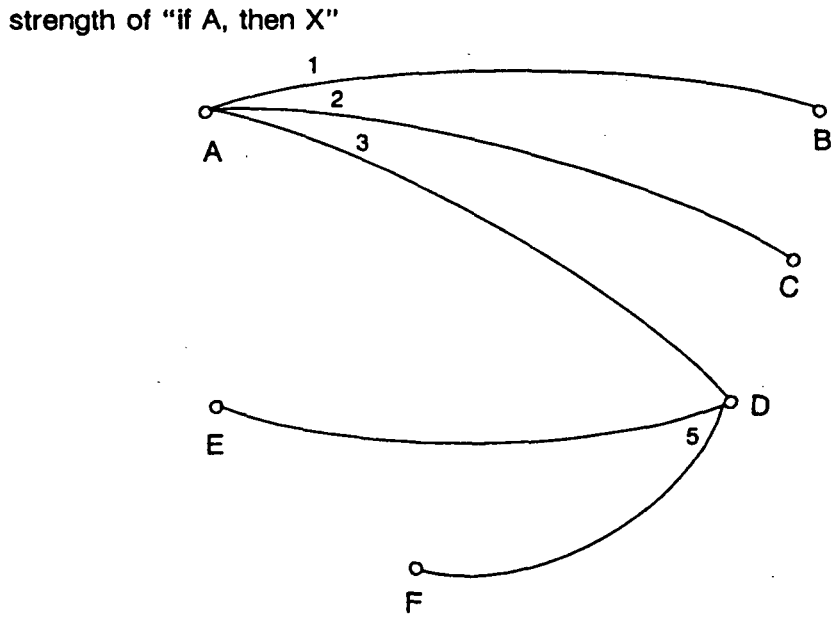


Figure 4-1: The basic unit of knowledge representation consists of a link between two pieces of knowledge called ***knowledge couplers*** that express a relation about the NPP (e.g. "pressurizer level" is coupled to "steam generator tube rupture.") When either terminus (or node) is activated, the item it is coupled with is suggested (bi-directional inference). The relation expressed in the coupler can be qualified by the state of some contextual knowledge (Figure 4-1a). A coupler also encodes the ***strength*** of relation between the two items linked, that is, how strongly the activation of one item calls to mind the other (Figure 4-1b). There are two strength values associated with each coupler, one for the strength of association in the each direction. Couplers are used to build up larger structures of interconnections between pieces of knowledge about the plant (Figure 4-1c).

In CES the strength parameters can indicate how strongly each item in a coupler would call to mind the other. The strength parameters bear some similarity to certainty factors often used in conventional expert systems in that these values can be used to indicate the degree of confidence in inferring one item in a coupler from the other. One important use of this information in CES is *to order* what knowledge is activated and processed when there are multiple links from one item in the knowledge base (Figure 4-1c). For example, if there were couplers in the knowledge base linking "decreasing pressurizer level" with "steam generator tube rupture" of strength 2 and with "break to containment" of strength 3, then CES would consider the latter as a stronger possibility (all other factors being equal) or as the first possibility that comes to mind, if resources are limited. The strength parameters play an important role in determining which hypothesis that could account for an observed finding will control further CES processing.

Couplers are used to build up larger structures of interconnections between pieces of knowledge about the plant. One can think of a network or topology of interconnections (Figure 4-1c). For example, a node is activated by some processing mechanism, e.g., a behavior analyst notices "decreasing pressurizer level". A relation in the knowledge base links this observation to the possibility of "steam generator tube rupture." This node, in turn, has connections to other nodes that represent other observable behaviors associated with this state ("increasing steam generator level"), other plant states ("loss of primary system water inventory"), and corrective responses ("reduce primary system pressure"). Other processing analysts use these relationships to draw conclusions within their areas of responsibility, if the analyst is active and if it is the next scheduled activity.

The network of couplers provide a powerful and flexible mechanism for representing virtually any knowledge about plant structure and function, disturbances and faults, goals and responses that NPP operators would be expected to know. Set up of the knowledge base requires data or hypotheses about what operational personnel do know (analysis of training programs). Knowledge provided to operators through external means such as procedures is also captured here. One can modify the information encoded in the knowledge base to represent differences that might exist among operators with respect to the depth and accuracy of the knowledge they possess about a NPP issue (e.g., simplistic vs. highly accurate mental models of an NPP process), and the accessibility or ease of calling to mind that knowledge. If knowledge is missing or distorted (buggy knowledge; e.g., Brown & VanLehn, 1980) and if it is relevant to the demands imposed by the nature of the incident, then CES performance will be degraded.

Couplers can be used to represent NPP knowledge at different levels of abstraction, whether directly observable or not, such as a plant parameter reading, an intermediate disturbance category such as a "mass imbalance," a fault category such as primary system break to containment, or a response such as "turn off the emergency cooling system." The level or levels of concepts depends on analytical or empirical results about how operators at different experience levels think about the state of the plant or how different procedure or training systems teach them to think about the plant.

The accessibility of knowledge is governed in part by the strength values associated with couplers and by the existence and content of any qualifiers on a coupler expressing a relation. Larger patterns among sets of couplers also determine the contexts in which knowledge is called to mind so that missing links can lead to inert knowledge.

The network of couplers also can represent different kinds of knowledge about the plant. For example, there can be a multi-step thorough reasoning process in going from observations to conclusions about plant state and selecting corrective actions that includes the intermediate conclusions operators may draw while systematically working through a problem. There can be *reasoning shortcuts* where there is a direct link that shortcuts through the multi-step reasoning, for example, a direct link from an observable plant behavior to a response chunk, a direct link from an observable behavior to an explanation (such as a direct link between a radiation monitor reading in the secondary side and a steam generator tube rupture event), or a direct link from a plant state to a response chunk. The thorough reasoning path will be more error resistant (assuming the knowledge is not buggy); the shortcut will be more efficient. Either or both of these coupler structures can be set up. The shortcut/thorough reasoning patterns are related to one interpretation of Rasmussen's "skill, rule, and knowledge" levels of reasoning behavior (Rasmussen, 1986).

4.2 Processing Mechanisms

Not all possible couplers are accessed and utilized (i.e., *activated*) on any one model *processing cycle* (i.e., following the input of one time step of plant data). Similarly, not all the plant data available for processing at any one processing cycle are examined.[5] The basic architecture of CES is designed to control what knowledge (and how much knowledge) is activated in a given cycle and what data are examined (the focus of attention).

[5]The statement that all available plant data are not processed is an over-simplification. In fact there is a preprocessor that scans all available data at the start of each time step to check for parameter values that meet the initial criterion to start processing by the model -- e.g., for spawning a behavior analyst.

CES contains processing mechanisms that enable monitoring plant behavior, situation assessment and explanation building to account for plant behavior, and formulation of responses to the perceived plant situation. Processing in CES is accomplished through the joint activity of multiple independent processing agents or *analysts*. The basic processing mechanism is to spawn an "analyst" when some criterion is met, who then performs some information processing work accessing knowledge available in the knowledge base as it needs it. There are three basic kinds of "analysts" each with their own area of responsibility. These are:

- *Behavior analysts* responsible for monitoring and analyzing plant behavior to decide if observed plant behaviors are expected or unexpected.

- *Situation analysts* responsible for analyzing the perceived situations and for postulating and pursuing possible explanations for unexpected findings.

- *Response plan analysts* responsible for selecting and adapting plans to correct or cope with perceived abnormal conditions.

These analysts are active processes that draw conclusions and "post" their results for other analysts to use as needed. Multiple instances of each basic type of "analyst" are generated or "spawned" as needed. A fundamental characteristic of this problem-solving architecture is that each analyst has a very narrow field of view and responsibility, and that complete problem-solving involves communication and coordination among the multiple analysts.

The knowledge base is accessed and utilized by the analysts to provide criteria for creating, interrupting, stopping analysts and to provide plant knowledge the analysts draw on to generate expectations, draw conclusions, and determine responses.

Each analyst does not represent a different person, rather the cooperative set of analysts are intended to model a single problem-solving system — be it an individual operator or a team of operators. The multiple analysts are intended to model the multiple types of processing (e.g., monitoring, explanation building, response planning) and lines of reasoning (e.g., multiple alternative explanations pursued) that occur in parallel and are interwoven during problem-solving, i.e., the organizational model of the mind. Each analyst posts partial results from his unique point of view that other analysts can access and build upon. The performance of this type problem-solving system depends heavily on the degree of communication and coordination among the analysts.

4.2.1 Behavior Analysts

The first processing layer is focused on the behavior over time of potentially observable data on the state of the NPP. Associated with each plant parameter state is a behavior analyst who has the responsibility of tracking that parameter (Figure 4-2). However, behavior analysts are not always active (i.e., they are not automatically created at the start of a run). There are two ways to create or "spawn" a behavior analyst.

One way is "data-driven" and occurs when a parameter or state change meets some criterion. The criterion can be defined in terms of deviations from target, out of limits, rates of change, or other change in behavior that an operator would regard as worth tracking if it was noticed.

A second way a behavior analyst is created is "knowledge-driven" and occurs when a situation analyst or response plan analyst requires information about a plant parameter for which a behavior analyst does not yet exist, and thus creates one.

Once created a behavior analyst performs several functions. It monitors the behavior of the datum which generated it based on criteria which define what behaviors are "interesting" at this initial level of analysis (e.g., direction of change, rate of change, off target, limit violations) and produces a stream of observed events, e.g., level is in a decreasing progression, limit x has been crossed.

The behavior analyst uses perceived or expected active influences, the set of processes which are presumed to be currently acting on the datum in question – to determine *expected* behavior.

The behavior analyst then assesses whether observed behavior is consistent with the expected behavior. If the observed behavior is consistent with expectations, then the behavior analyst continues to monitor the datum. If the observed behavior is inconsistent with expectations, then an unknown influence is postulated which needs to be identified.

A behavior analyst can report several types of findings to other analysts: observed behavior, abnormalities, confirmation of expectations, violations of expectations.

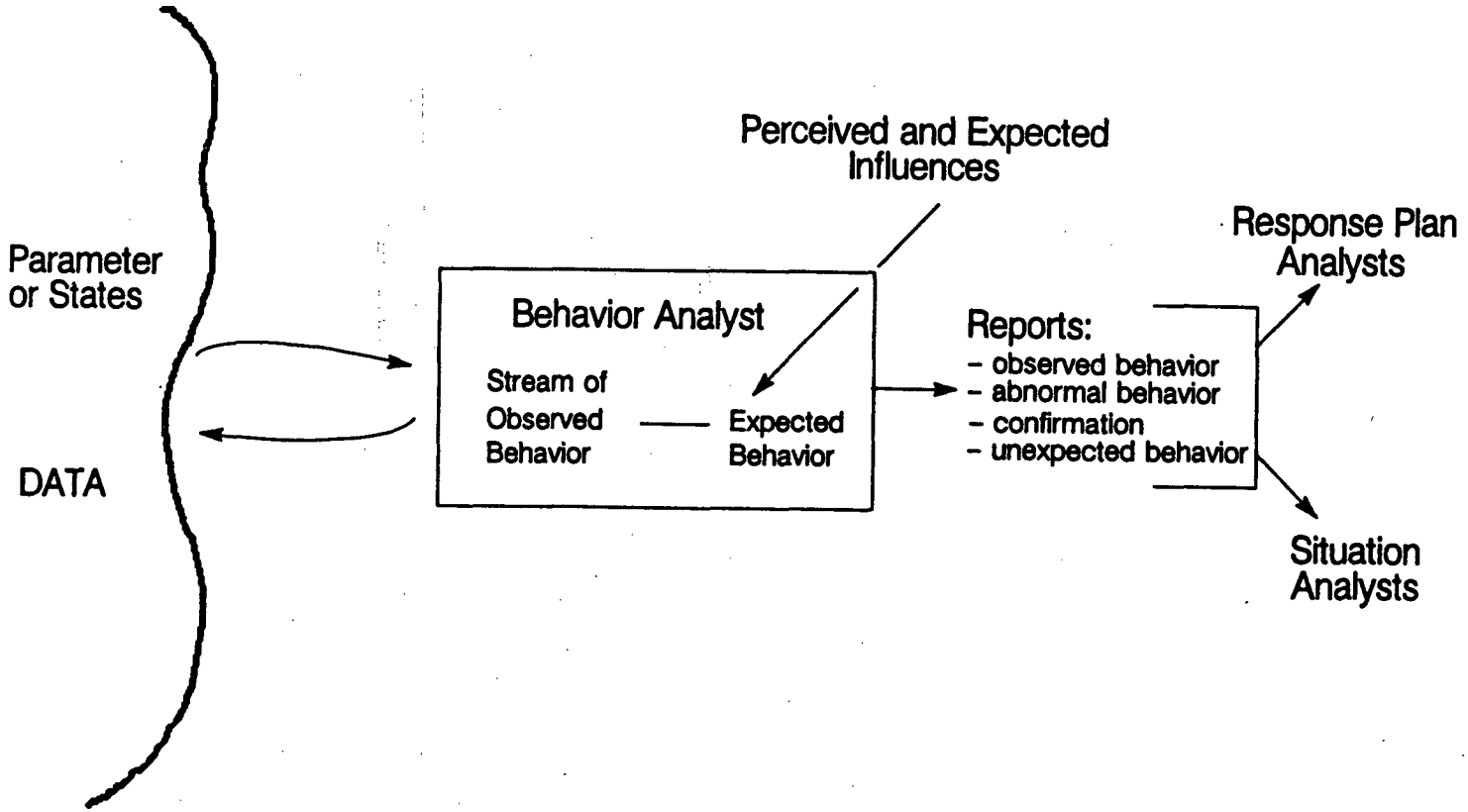


Figure 4-2: The first processing layer in CES is focused on monitoring of NPP data over time. Associated with each plant parameter state is a behavior analyst who has the responsibility of tracking that parameter Behavior analysts are not always active. They are created either when a plant parameter or state change meets some criterion (i.e., "data-driven") or when a situation analyst or response plan analyst requires information on the parameter that it is responsible for monitoring (i.e., "knowledge-driven").

4.2.2 Situation Analysts

Observed behavior that cannot be accounted for by the behavior analyst given the presumed list of active influences constitute another answer to the question of what is "interesting," and these findings are passed on to a second type of analyst -- situation analysts. Violation of an expectation leads to the creation of a situation analyst (assuming one does not already exist trying to explain this datum's behavior) or a new input to one that has already been created (Figure 4-3).

The situation analyst's responsibility is to try to *explain the unaccounted for behavior* of the datum in question in the context of other data and knowledge about the state of the NPP. The Situation Analyst considers *what changes in the active set of influences could account for the observed behavior.* For example, if the finding to be explained is a decreasing progression in primary system pressure, then the situation analyst would consider processes that can influence pressure in this way, such as pressurizer power operated relief valves (by considering elements linked to decreasing primary system pressure in the knowledge base). To do this, the situation analyst explores to see *if other evidence about the state of the NPP is consistent with this hypothesis* (knowledge-driven data search) by asking or creating behavior analysts to provide it with specific evidence (e.g., what is the state of the pressurizer relief tank). They can also signal an active or newly created behavior analyst that it is interested in future findings -- setting a lead. For example, notify me if any interesting changes occur in the pressurizer relief tank. This has the effect of increasing receptivity to incoming data points that are evidence on particular issues.

How does a situation analyst work?

To illustrate the basic processing involved in explanation building, consider a simple example where a single situation analyst is created to account for an unexpected finding. A change is observed in a parameter (P_1) which its behavior analyst determines is unexpected given the current active influence set for this parameter (e.g., I_1 suggests that P_1 should be increasing, but P_1 is observed to be decreasing). This implies the existence of another active influence, which is the signal to activate a situation analyst -- S_1. S_1 calls to mind potential influences on P_1, such as I_2, and evaluates the new influence set (I_1 plus the hypothesis of I_2) to see if it could explain the observed P_1 behavior (Figure 4-4). Other potential influences are called to mind and ordered as possible explanations, given the set of findings to be accounted for.

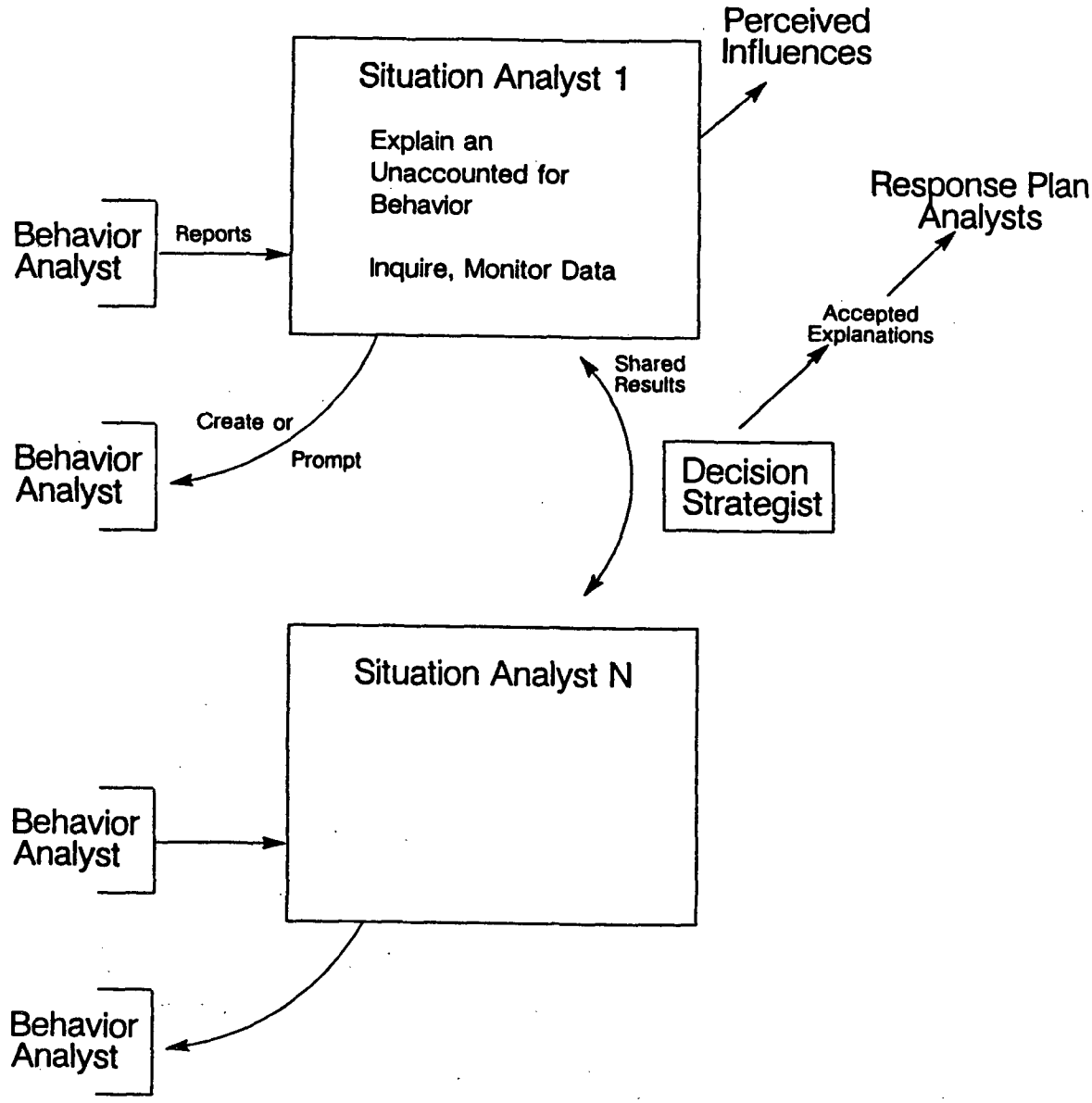


Figure 4-3: Observed behavior that cannot be accounted for by the behavior analyst are passed on to a second type of analyst -- situation analysts. The situation analyst's responsibility is to try to *explain the unaccounted for behavior* of the datum in question in the context of other data and knowledge about the state of the NPP. The Situation Analyst considers *what changes in the active set of influences could account for the observed behavior.* To do this, the situation analyst explores to see *if other evidence about the state of the NPP is consistent with this hypothesis* (knowledge-driven data search) by asking or creating behavior analysts to provide it with specific evidence.

The strength values associated with couplers in the knowledge base play an important role in this process of ordering possible explanations. For example, if there were couplers in the knowledge base linking "decreasing pressurizer level" with "steam generator tube rupture" of strength 2 and with "break to containment" of strength 3, then CES would consider the latter as a stronger possibility (given only that single finding) or as the first possibility that comes to mind, if resources are limited. The strongest possibility then controls further CES processing: the influence set of I_1 plus the hypothesis of I_2 is evaluated to see if other plant behaviors are consistent with this hypothesis.

Given unlimited resources, a situation analyst will explore all possible explanations it knows about completely. However, it can be interrupted before following up each possibility depending on the level of resources and competing activities. This means the order in which CES considers possible explanations can be an important determiner of what explanation is built to account for unexpected findings. The strength parameters play an important role in determining that ordering and can be used to implement whatever ordering is plausible given the operators or hypothetical operators in question.

When a postulated influence could explain observed P_1 behavior, the next step is to see if the postulated influence is consistent with other plant data (is there evidence this influence is active? is the behavior of other parameters consistent with this additional influence?). Thus, knowledge-driven data gathering is triggered. In considering whether a postulated influence is active, data gathering can consist of (a) simple requests to report currently observed behavior, (b) setting leads for specific behaviors of interest, and (c) triggering general searches of some region of the plant for "interesting" behaviors. *Setting a lead* increases for some period of time the receptivity of a behavior analyst to specific plant behaviors that are relevant to the situation analyst's explanation building activities. If behavior analysts return with observations that are consistent with the explanation currently being pursued, then the belief in that explanation increases. If behavior analysts return with observations that are inconsistent with the explanation currently being pursued, then the belief in that explanation decreases.

The order in which CES considers possible explanations affects the order in which knowledge-driven monitoring requests go out to behavior analysts. For example, when operators detect a moderately fast unexplained decrease in pressurizer level and pressure, they often check the pressurizer relief valves and letdown system before they pursue the possibility of loss of coolant break, before they pursue the possibility of charging system trouble (e.g., Woods et al., 1982). The ordering (i.e., the relative strengths on couplers in

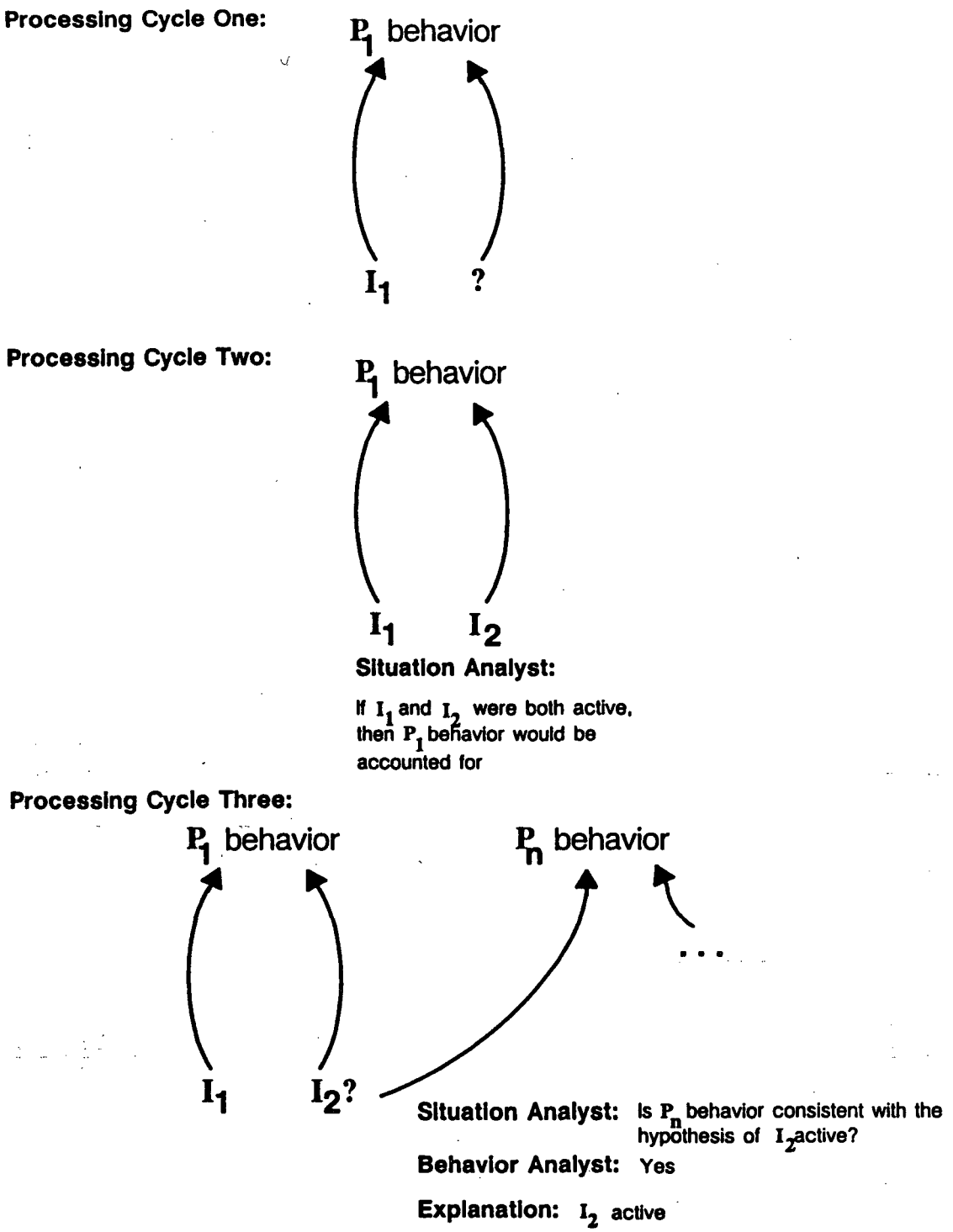


Figure 4-4: Case 1. A change is observed in a parameter (P_1) which its behavior analyst determines is unexpected given the current active influence set for this parameter (e.g., I_1 suggests that P_1 should be increasing, but P_1 is observed to be decreasing). This implies the existence of another active influence so situation analyst S_1 is triggered. S_1 calls to mind potential influences on P_1, such as I_2, and evaluates the new influence set (I_1 plus the hypothesis of I_2) to see if it could explain the observed P_1 behavior. Other plant behaviors are then examined to see if they are also consistent with this hypothesis.

the knowledge base) could be based on any of several criteria for establishing strength of association: strength of implication (e.g., decreasing primary system pressure and level alone are stronger evidence for a loss of coolant break than for a steam generator tube rupture), severity of consequences (e.g., a loss of coolant break has more severe safety consequences than a charging line break), observability (e.g., it is easier to identify a stuck open relief valve, assuming direct valve position indication, than a moderate size loss of coolant break), empirical results (given some indications, operators have been found to search possible explanations in this order).

Coordination among multiple situation analysts.

There can be multiple situation analysts active at one time, each trying to explain different aspects of observed plant behavior. They can share results of their exploration, converge on a single coherent explanation set for the whole pattern of observed behavior, or develop multiple accounts for the whole pattern. Since multiple alternate explanations for a given configuration of evidence may be simultaneously pursued, a criterion is required to decide when an explanation is sufficiently supported and coherent to be adopted (be it provisionally) and acted upon. The responsibility for committing to an explanation resides in the *Decision Strategist.* The explanation selected need not be a single factor that accounts for all of the findings. It can be a set of multiple factors that in combination account for the items to be explained. While multiple criteria for a coherent explanation can be envisioned and alternative approaches should be explored, the Decision Strategist, as currently implemented, employs a simple parsimony criterion for committing to an explanation.

To illustrate the coordination and communication involved with multiple situation analysts, let us consider a second example (Figure 4-5). In this example there are two unexpected findings (P_2 is decreasing and P_3 is increasing) that each trigger their own situation analyst (S_1 & S_2 respectively). S_1 calls to mind that either influence I_2 or I_4 could account for the observed behavior of P_2 (Figure 4-5). However, I_2 has a higher strength value than I_4 as a potential explanation for P_2 behavior, so that, given that evidence alone, I_2 is provisionally the more likely explanation. Similarly S_2 calls to mind that either I_3 or I_4 could account for the observed behavior of P_3, but that I_3 is the more likely explanation for P_3 behavior since it has a higher strength value. Now, if S_1 and S_2 share results, then they can converge on a single account for two unexpected findings since I_4 would account for both P_2 and P_3 behavior. Note that a criterion for what is a good explanation is needed to decide that one explanation for two findings (I_4) is better than two explanations for two findings (even though each explanation is stronger individually for each separate finding) -- a criterion of parsimony.

CES explanation building behavior can be changed by varying when multiple situation analysts are created, how they share results, how they converge on a single explanation for the whole pattern of observed behavior or develop multiple accounts for the whole pattern, and the criterion for what is a coherent explanation. These variations provide performance adjustment factors to model different situations, for example, a fixation prone person or team structure.

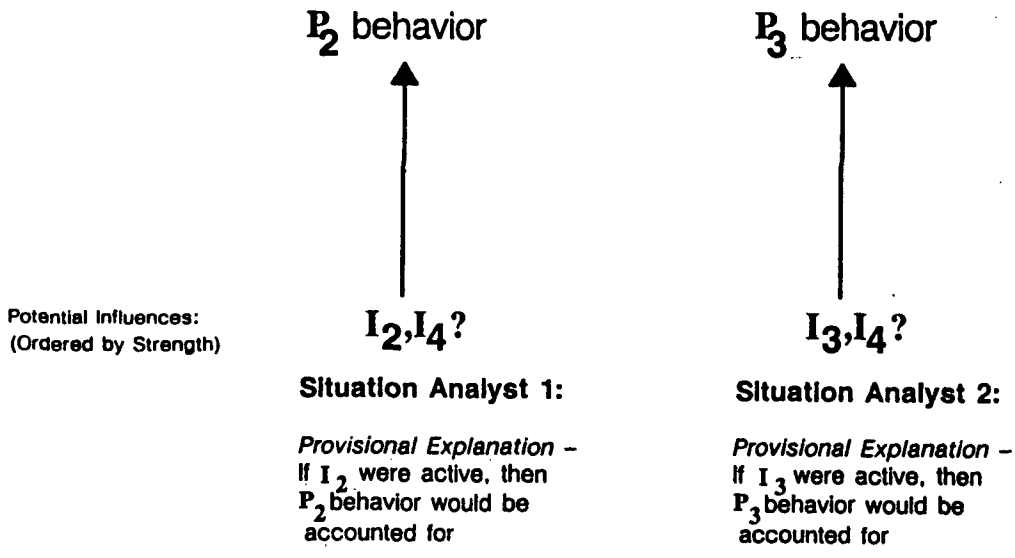
In identifying the fault or set of faults that can account for plant behavior there is always a tradeoff between efficient strategies that quickly converge on a solution but are subject to error, and more thorough strategies that are less error prone but take more time and resources to converge on a coherent explanation. In CES, the explanation building mechanisms were designed to be able to represent three different strategies that range from efficient but fixation prone to thorough but resource intensive.

Single explanation assumption.

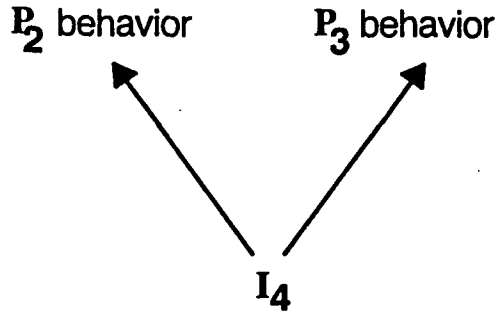
At one extreme, CES can be made to consider only single fault explanations for plant behavior. Under this setting, only the first unexpected parameter behavior will create a situation analyst; in the current example, unexpected finding P_2 would trigger the creation of situation analyst S_1 (Figure 4-5). No other situation analysts will be created. When a second parameter (P_3) also exhibits unexpected behavior, it does not create its own situation analyst. Instead, the relevant behavior analyst reports the finding to the first situation analyst — S_1. If the information is useful in narrowing down the set of potential influences being considered by the one situation analyst, then the limit on multiple situation analysts will not retard performance. This is true in the current example where information on P_3 behavior can be used by Situation Analyst 1 to focus in on I_4 as the most appropriate explanation since I_4 can account for both P_2 and P_3 behavior. However the single situation analyst strategy will only be efficient in single fault situations where all the unexpected findings can be accounted for by a single unknown influence.

This extreme limit on multiple situation analysts leads to extreme fixation-prone, garden-path behavior if CES starts down the wrong explanation or if there are multiple failures (note that under the single explanation assumption I_3 is never considered as a possible explanation in the current example). Unexpected findings that cannot be accounted for by the first situation analyst that happened to be created are ignored in the sense that no explanation is sought for them. This behavior is psychologically plausible under some conditions. Individuals who become fixated on the wrong solution path often miss or ignore evidence that does not readily fit with the hypotheses being entertained (e.g., Montmollin & De Keyser, 1986).

Processing Cycle One:



Processing Cycle Two:



Convergence: If both P_2 and P_3 behavior must be explained by a single influence, then I_4 would account for both

Explanation: I_4 is present

Figure 4-5: Case 2: This example illustrates the coordination and communication among multiple situation analysts. There are two unexpected findings (P_2 is decreasing and P_3 is increasing) that each trigger their own situation analyst (S_1 & S_2 respectively). S_1 calls to mind that I_2 or secondly I_4 could account for the observed behavior of P_2. Similarly S_2 calls to mind that I_3 or secondly I_4 could account for the observed behavior of P_3. By sharing results, S_1 and S_2 converge on a single more parsimonious account for two unexpected findings: I_4 would account for both P_2 and P_3 behavior.

Multiple explanations.

An opposing approach to building explanations is to entertain multiple views of what would account for a set of findings (multiple explanations in the explanation set). Initially, in this strategy, separate explanations are sought for each finding to be accounted for. This is accomplished by having each unexpected finding create a situation analyst with the responsibility of evaluating potential explanations for that finding. Multiple situation analysts work largely independently of each other -- each pursues an explanation for one part of the total puzzle.

Subsequently, places where these separate accounts can converge are sought, if possible. This is done by having multiple situation analysts share partial results in order to recognize where a single explanation can account for multiple findings. Convergence is based on adding the constraint that the "best" explanation of those called to mind by either situation analyst must account for union of the evidence seen by each situation analyst. This convergence process requires criteria for selecting among alternative views (alternative sets of explanations) and deciding when an explanation is sufficiently supported and coherent to be adopted and acted on. Note the difficulties in deciding which of multiple alternatives to pursue or to pass on to response plan analysts to trigger selection of appropriate corrective responses. This responsibility falls on the *Decision Strategist* who coordinates the activities of the multiple situation analysts and commits to an explanation or set of explanations. Changing the criteria for what is a good or coherent explanation set will change which of the possible views of a set of findings will be preferred.

A version of this strategy was built in the CADUCEUS problem solving system (Pople, 1985). The virtue of this strategy is that it is *highly fixation resistant* because it can call to mind and thoroughly consider all of the ways the set of findings can be put together, assuming it has sufficient resources and time to complete the assessment. However, this strategy requires more processing to arrive at an explanation, processing that is vulnerable to breakdowns and that takes resources (time) to complete, when there is only a single fault.

The example given in Figure 4-5 can be used to show how processing occurs under the thorough fixation resistant explanation building strategy. Unexpected finding P_2 triggers the creation of a situation analyst (S_1) which considers I_2 and secondly I_4 as possible influences that could account for P_2 behavior. The observation of second unexpected finding, P_3, triggers a second situation analyst (S_2). S_2 calls to mind and establishes that I_3 and secondly I_4 could account for the observed behavior of P_3. In the convergence process

the possible explanations called to mind by both situation analysts are merged and ordered relative their ability to account for both P_2 and P_3. Under this additional constraint, the presence of I_4 would be the strongest explanation because only it can account for both unexpected findings.

Now add to our example a third unexpected finding (P_4) which leads to the creation of a third situation analyst, S_3 (Figure 4-6). S_3 then calls to mind I_3 as the strongest possibility to account for the observed P_4 behavior. This produces a dilemma because there are two ways these results could be put together: I_4 could be active which accounts for findings P_2 and P_3 and a separate explanation for the unexpected P_4 behavior remains to be found, or I_3 could be active which accounts for findings P_3 and P_4 and a separate explanation for the unexpected P_2 behavior remains to be found. In this kind of situation CES defers commitment and searches for more evidence.

Single explanation bias.

An intermediate explanation building strategy is first to try to build a single explanation that accounts for all unexpected findings. Multiple explanations are entertained only when a single explanation is unable to deal with all of the findings. With this setting, the first situation analyst to be created (S_1) performs all explanation building activities as long as it can offer an hypothesis to account for *all* of the unexpected plant behavior encountered up to that point. In the current example this means that when unexpected finding P_3 is reported, it is "absorbed" by S_1, if it helps distinguish among the possible explanations being considered by S_1. Therefore, no additional situation analyst would be created. This absorption has two consequences. First, it narrows the set of candidate explanations for P_2 behavior. The presence of I_4 is a stronger explanation than I_2 because it accounts for two findings instead of one. Second, it precludes alternative explanations for P_3 behavior from being called to mind (such as I_3). This can lead to fixation effects because alternative explanations for P_3 behavior will be excluded from consideration. For example, when P_4 is observed, a separate explanation will be sought for this finding because it has already "jumped to the conclusion" that I_4 accounts for P_2 and P_3. Thus, the possibility that I_3 accounts for both P_3 and P_4 (Explanation Set B in Figure 4-6) will not be considered and a potential influence on P_3 could be missed.

This strategy of a single explanation bias will perform well in many situations because it is not completely limited to single fault hypotheses. Nevertheless, this strategy is prone to garden path behavior. When plant behavior is detected that is unexpected given the initial explanation

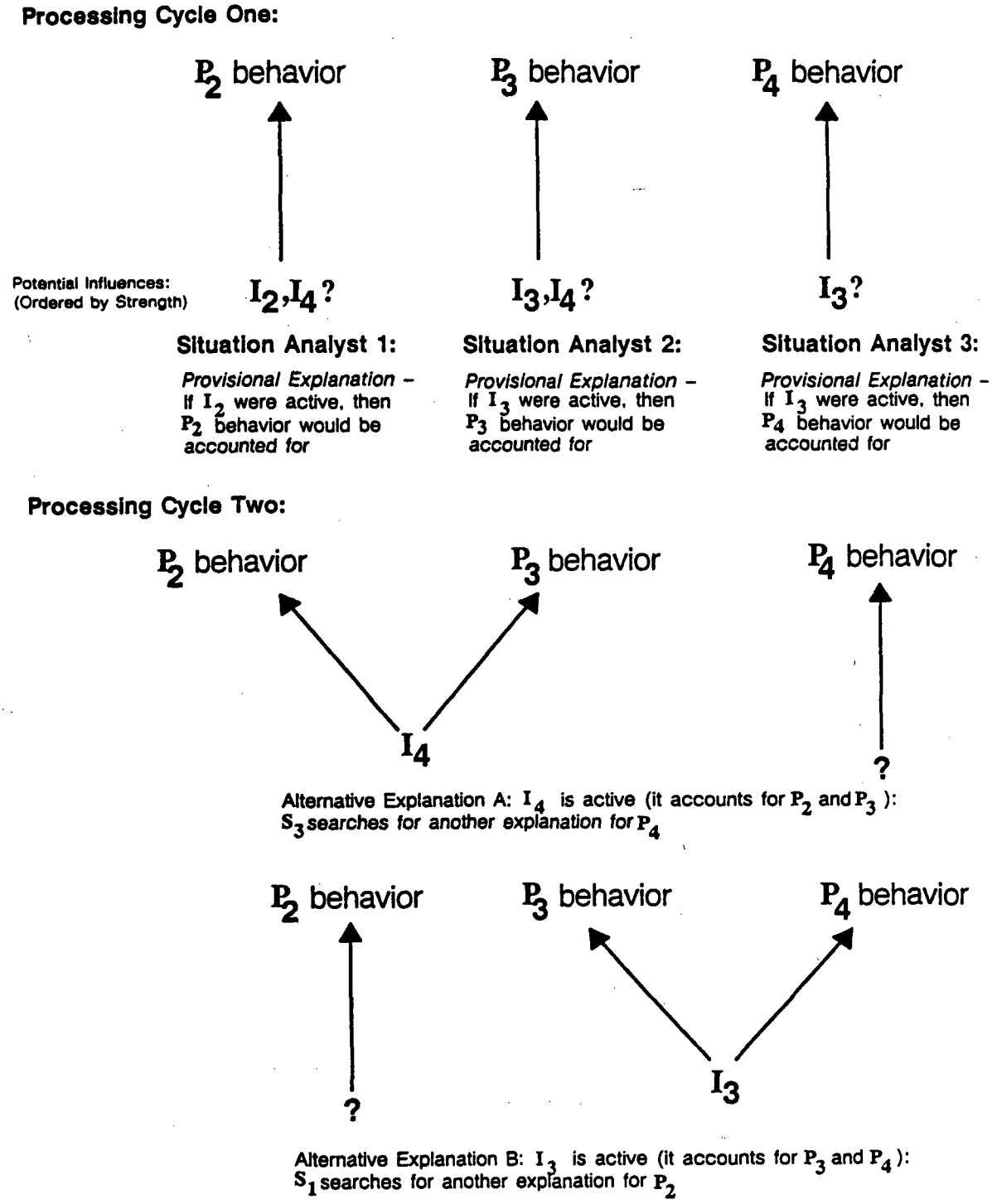


Figure 4-6: Case 3: This example is an extension of Case 2 in Figure 4-5. A third unexpected finding (P_4) is observed that leads to the creation of a third situation analyst, S_3. S_3 then calls to mind I_3 as the strongest possiblity to account for the observed P_4 behavior. This produces a dilemma because there are two ways these results could be put together: I_4 could be active which accounts for findings P_2 and P_3 and a separate explanation for the unexpected P_4 behavior remains to be found, or I_3 could be active which accounts for findings P_3 and P_4 and a separate explanation for the unexpected P_2 behavior remains to be found. In this kind of situation CES defers committment and searches for more evidence.

generated, it will not reconsider its initial interpretation. It assumes that the initial interpretation is correct and seeks an explanation for the residual unexplained findings in light of the initial interpretation. This behavior is very typical of the kinds of fixation effects observed in human problem solving. Once an individual settles on an explanation, he may become committed to it, and seek to "rationalize" all new incoming data that are unexplained by the initial interpretation. The single explanation bias emphasizes processing efficiency, while being able to handle some multiple fault situations.

4.2.3 Response Plan Analysts

Response plan analysts call to mind knowledge about pre-planned sequences of responses (Figure 4-7 over time to disturbances (e.g., decreasing pressurizer level) and faults (e.g., steam generator tube rupture). They are responsible to monitor, adapt, and fill in these plans as needed. The response sequences include both automatic system responses and human operator responses. For example, the script for a decreasing progression in pressurizer level might read "increase/maximize net charging inflow" {"increase charging flow (automatic)," "start additional charging pump (optional manual)," "stop letdown outflow (automatic)"} and, if this response is insufficient, "switch to emergency injection (automatic)."

Knowledge about how to respond to a perceived situation can be triggered or called to mind by (1) observed behavior (e.g., decreasing pressurizer level), (2) accepted explanations, or (3) anticipated effects of hypothesized influences (would x account for decreasing pressurizer level in the current context).

The kind of corrective response activated depends on what response information is stored in the knowledge base – what response scripts and what kinds of plant states they are linked to. This is one kind of information contained in procedures. The kind of corrective response activated also depends on CES processing – what plant behavior has been observed and what explanations have been constructed at that point in the unfolding incident. Thus, the process is one of *plan assembly* from pre-stored response plans.

The responses can be more function (symptom) based or more "event type" based. This depends on the grain at which responses are linked to states in the knowledge base and on the grain of situation assessment at that point in the unfolding incident. For example, early in an incident the response may be more function based because deeper explanations of the situation have not been built yet or because the evidence is not yet available; but this can shift to more "event type" as the evidence or time needed to build deeper explanations becomes available.

Knowledge called forth from these scripts about what actions on the plant to expect can be used by the appropriate behavior analysts to check if the response has happened (e.g., did letdown isolate?) and if the new influence has affected the behavior of the datum in question (e.g., is level now stable or increasing). In this way CES can be set up so that human execution errors, failures of automatic systems, and failures of the machine to respond as demanded can be detected. On the other hand, CES can fail to detect these execution failures because of monitoring failures or because expectations about what should have occurred are substituted for the results of actual data checks, for example, when monitoring resources are limited.

Evoked responses which are executed manually can also be output to people who close the loop from CES to the simulated plant by executing intentions to act.

Responses can be evoked and monitored in parallel with explanation building activities. For example, observation of decreasing level can trigger responses appropriate to this state, at the same time that activities are underway to explain why level was decreasing in the first place (what level influencer exists that is not in the current set). While these activities can go on in parallel, they can also compete for resources so that doing one blocks the other.

The behavior of pressurizer level following the addition of a new (or stronger) influence from the response script (assuming that the response was actually made) provides more information for the Situation Analyst (did the response counteract or dominate the unknown influence or does the unknown influence continue the level decrease). If the unexpected behavior continues unabated, one possibility is that the expected response was not properly carried out.

Currently, the plan assembly capabilities and the plan monitoring capabilities (execution failure detection) of CES are implemented. Another important desired competence for response management is choice under uncertainty and risk, but this capability is not implemented at this time. This means that, when selecting responses, CES cannot yet explicitly take into account factors such as uncertainties across alternative explanations, consequences of actions if wrong, goals, post-conditions to actions, or requirements for actions. More creative plan generation capabilities that go beyond building new plans out of more elemental existing responses (e.g., planning by analogy) would require another step increment.

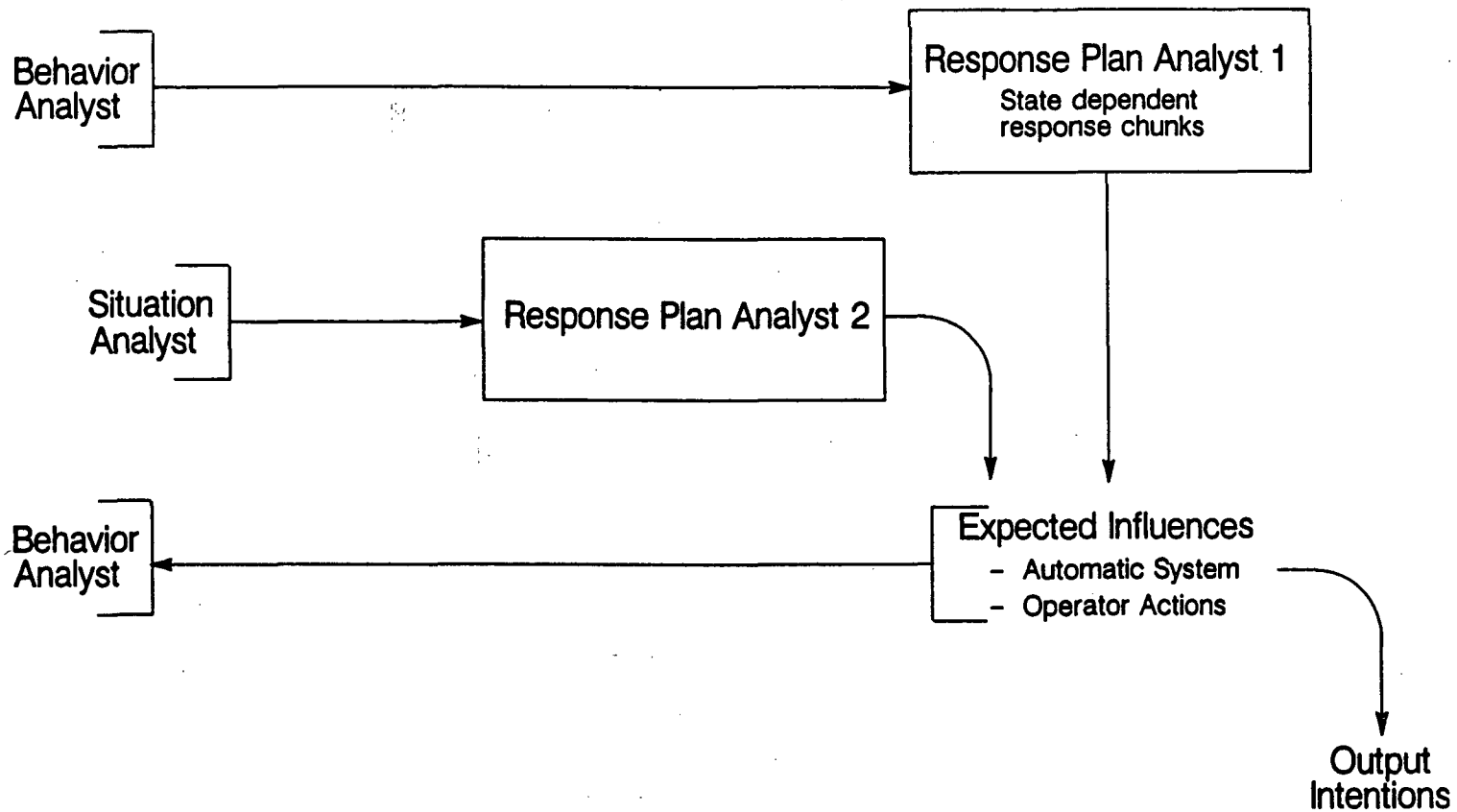


Figure 4-7: Response plan analysts call to mind knowledge about pre-planned sequences of responses over time to disturbances (e.g., decreasing pressurizer level) and faults (e.g., steam generator tube rupture). They are responsible to monitor, adapt, and fill in these plans as needed. The response sequences include both automatic system responses and human operator responses.

4.2.4 Qualitative Reasoning

Couplers in the knowledge representation express relationships between NPP entities. Some of the relationships expressed are that one entity (process) *influences* another entity (parameter). Knowing about what influences are acting on a parameter allows qualitative reasoning about its resulting behavior. Qualitative reasoning attempts to determine high information value landmarks in the behavior of continuous parameters (e.g., direction of change, changes in the direction of change) rather than to completely determine quantitative continuous functions. Simple numerical calculations can play a role in forming expectations about plant behavior e.g., setpoint crossings and categorized continuous values (absolute values or rates of change).

If there is only a single active influence or multiple influences which have the same direction of impact, CES computes expected behavior simply. When the qualitative reasoning bottleneck arises (when there are multiple active influences which have different impacts), CES can respond in two ways. First, dominance relations, directly entered into the knowledge base, can be read out to determine expected behavior. For example, the relief valve pathway open effect to reduce primary system pressure dominates the pressurizer heaters effect to increase pressure. Therefore, if both constitute the active influence set for primary system pressure, then pressure would be expected to decrease. Second, if the overall impact of multiple influences cannot be resolved, then the default position is that any judgment about what is expected or unexpected behavior is postponed, the behavior analyst continues to monitor, and a situation analyst is not activated. However, CES contains a mechanism which learns dominance relations for future use. If the behavior analyst cannot infer expected behavior from the pattern of influences, the resulting behavior is noted in association with the pattern of influences and can be recalled when this pattern occurs in the future.

4.2.5 Uncertainty

In the NPP world uncertainty arises at many points during the problem solving process. One source of uncertainty arises because of the many-to-many mapping between available data and plant physical or functional structures. As a result a given piece of evidence may be accounted for in many ways (e.g., "low pressurizer level" can arise due to an energy imbalance caused by a secondary side disturbance; a leak or break into containment; a leak in the charging line; a sensor failure, etc.). CES handles uncertainty in several ways. First, the couplers explicitly encode the many to many mapping that exists in a given NPP. For example, there would be couplers linking "low pressurizer level" to each of the many disturbances

that it provides evidence for (i.e., there would be couplers explicitly encoding all of the occurrences that could act as influences on pressurizer level). Second, the directional strength of relationship parameters (i.e., evoking strength and frequency) associated with a coupler are used to reflect the degree of uncertainty there is in concluding one item in the coupler given the other. For example, in the coupler that links "steam generator tube rupture" and "low pressurizer level", the evoking strength of "low pressurizer level" for "steam generator tube rupture" would have a low value reflecting the fact that "low pressurizer level", by itself, provides weak evidence for a steam generator tube rupture.

A second kind of uncertainty arises in explanation building. When multiple views of a set of unexpected findings are considered (via a set of activated situation analysts), uncertainty arises in the multiple ways that the pieces of the puzzle can be put together. This is regulated through the Decision Strategist and its criteria for what is a good explanation.

A third kind of uncertainty arises during response planning in selecting among a set of alternative responses (e.g., attempt to reinstate a malfunctioning process or substitute an alternative process that achieves the same goal). Making response choices under risk and uncertainty is the responsibility of the response plan analyst. While CES includes provisions for a choice under risk and uncertainty module, this module is not part of the current implementation.

4.3 The Virtual Display Board

Included in the knowledge representation is a description of what data about plant state are directly available to the model to "see," or what we call the virtual display board. This description represents what plant information would be directly available to the operators to observe. CES monitors the virtual display board to acquire data about plant behavior.

The CES knowledge base includes a list of plant parameters or states that it can directly access (e.g., from a data file or as output from a simulation program). Depending on the plant being modeled these plant parameters can be direct sensor readings, or more integrated information about plant state such as the output of computerized displays or decision aids). Associated with each element on the virtual display board are parameters that reflect characteristics of how that information is presented in the plant being modeled (i.e., characteristics of the *representation* provided to the operator of that NPP).

There is a parameter associated with plant data that reflects the degree to

which the form or content of a data point commands attention or can capture control of processing resources -- prominence or *salience*. For example, in terms of form, changes in a datum displayed as meter in a field of similar meters are less likely to command attention compared to a representation of the changes as blinking lights or audible sounds. For example, in terms of content, changes in some plant data are more "important" than others (e.g., better indicators of system state; an operator's idiosyncratic or favorite data points to check), such as changes in pressurizer pressure are much more salient than changes in pressurizer relief tank pressure, independent of the form of presentation. The salience parameter is a number that controls the priority or order in which behavior analysts report their findings to other analysts (response plan or situation analysts). Therefore, salience is one factor that effects whether an observed finding captures or interrupts ongoing processing (e.g., changes in primary system level, pressure and temperature take precedence, in general, over parameters such pressurizer relief valve position). The specific context (results of other processing) also determines what processing is carried out next.

Another parameter associated with plant data expresses how easy it is to see its value or state -- *observability*. This represents factors such as the cost of acquiring data (physical or virtual distance), the amount of processing of the display to extract the desired information, the potential for misreading. Observability is a number that controls whether any change in a datum can spawn its behavior analyst. If the observability number exceeds a threshold value, then a change in that datum can create a behavior analyst -- data-driven monitoring can occur. If the observability number is below the threshold, that datum would be monitored only if other CES processing analysts decided that information on the behavior of this piece of plant data was needed (knowledge-driven data gathering). For example, pressurizer relief tank (PRT) level is not normally monitored, but would become relevant to monitor in specific contexts such as when pressurizer pressure is high and the pressurizer relief valve pathway is open to the PRT (evidence of increasing PRT contents would be expected) or when pressurizer pressure is low and decreasing (called to mind as a possible explanation).

Currently in CES, data are correctly read out or there is no read out (i.e., it does not look). But data could also be misread or only approximate answers read out (e.g., if the display is at the wrong resolution for the intended judgment -- x is in the area of the limit but whether it has crossed it or not can not be seen). Misreadings and approximate readings are not handled in the initial program (except as no read out).

4.4 Samples of CES Processing

This section contains segments of CES processing behavior over different time samples of plant data. The samples show how the CES architecture functions to monitor plant data, form expectations, build explanations, and select corrective responses. Following each sample there is a commentary about what aspects of CES cognitive competencies are illustrated in that processing segment. The samples are idealized versions of the actual behavior of the current implementation of CES. Those capabilities which have not been implemented are noted. The NPP situations used in the processing samples are not chosen to be relevant to PRA studies; they are chosen from the design incidents used to develop CES.

The first sample simply shows the processing needed to detect and correct a stuck open pressurizer spray valve. The second sample focuses on a difficult diagnostic situation (interfacing system break) where, depending upon the Performance Adjustment Factor settings, the problem solver could fixate on an erroneous diagnosis. In the third sample, there is a fault that leads to excessive steam flow on the secondary side and reactor trip. CES must reason across multiple disturbances to locate the source of the disturbance chain, and it must decide whether the reactor trip has eliminated the source of the trouble (a turbine problem) or whether disturbance continues to be active (a steam break). In the fourth sample there are two faults that both affect primary system pressure (the faults from samples 1 and 3). In this case CES must recognize that disturbances remain after one fault has been eliminated.

4.4.1 CES Processing Sample: 1

Time Step 1

Initial conditions (t_0):
Plant is at steady state with reactor power = turbine power = 50%; major parameters are all on target; pressurizer spray valve has just stuck open (see valve PCV-455B in Figure 4-8).

CES Processing
A behavior analyst (#1) awakes and observes that a pressurizer spray valve (PCV-455B) is open (observation of an "interesting" plant behavior). Behavior Analyst 1 continues to monitor for changes in the pressurizer spray process.

The valve open indicates the pressurizer spray process is an active influence on primary system pressure (decreasing progression) and an expectation is

posted that a decreasing progression in primary system pressure will be observed.

Behavior Analyst 1 also notes that the pressurizer spray process is active when primary system pressure is below the automatic system setpoint (2260 psig). This is an abnormality and triggers a response plan analyst to consider appropriate responses to this situation. The first response would be for an operator to correctly align the valve. The response plan analyst generates an intention to close the spray valve (for output from CES as well) and it posts the expectation that the valve will be closed as expected influence on the pressurizer spray process. Behavior Analyst 1 monitors for this change.

Time Step 2

Plant state t_0+30 seconds:
Primary system pressure continues to decrease; spray valve remains open.

CES Processing
Behavior Analyst 1 is monitoring the pressurizer spray process. It expects spray valve PCV-455B to be closed based on the expected influence posted at Time Step 1, but observes that the spray valve is still open.

A behavior analyst (#2) awakes and observes a decreasing progression in primary system pressure (observation of an "interesting" plant behavior). This behavior is consistent with the expectation generated for primary system pressure behavior, given that the pressurizer spray process is active.

Behavior Analyst 2 observes that primary system pressure is abnormally low (less than 2210) which triggers a response plan analyst (#2) to consider appropriate responses to correct this. The response plan analyst invokes the response script for decreasing primary system pressure. The first stage of response is automatic activation of the backup heaters, and the response plan analyst posts an expectation that the backup heaters will activate. The second stage of response is automatic reactor trip, and the response plan analyst posts an expectation that an automatic reactor trip will occur. The third stage of response is an automatic safety injection signal (SI) and initiation of emergency core cooling (ECCS), and the response plan analyst posts an expectation that an automatic SI signal and ECCS initiation will occur.

Time Step 3

Plant state t_0+300 seconds:
Spray valve is manually closed; primary system pressure begins to recover.

CES Processing
Behavior Analyst 1 is monitoring the pressurizer spray process. It observes that spray valve PCV-455B has closed. This confirms the expectation posted at Time Step 1.

The pressurizer spray process is cleared as an active influence on primary system pressure.

Behavior Analyst 2 is monitoring primary system pressure and observes a change — it is now in an increasing progression (observation of an "interesting" plant behavior).

This behavior is consistent with the perceived influences acting on primary system pressure (in the absence of other influences primary system pressure will move back to the current equilibrium value for this plant mode, e.g., 2235 psig).

Commentary: Sample 1

Monitoring
There are several aspects of monitoring that are illustrated in this processing sample. First, it shows how adding and removing active influences changes what is expected plant behavior.

Second, it shows how CES is capable of getting feedback on execution failures or failures of components to respond as demanded (CES also needs to be capable of failing to verify expected influences, for example, as a function of resource competition). In this case, there is a verification of a correct manual response to a detected abnormality.

A basic discrimination problem in situation assessment following some disturbance is judging whether the plant is returning to a "normal" state or whether there are any remaining disturbances? One usually thinks about the need to discriminate abnormalities against the background of a previously normal plant; it is also necessary to be able to discriminate the change from a disturbed state back to a "normal" state after actions to correct the perceived disturbances. This is part of being able to detect whether corrective responses have been effective. In this case at Time Step 3, all abnormalities have been eliminated, and CES perceives plant state to be moving back to normal. If the responses are not effective, either because of

execution errors, failures of the machine to respond as demanded, other faults, erroneous situation assessment or erroneous intentions, then CES should be capable of detecting that plant state is not responding as expected and continue to monitor, to work towards an explanation, and to generate corrective responses (see Processing Sample 4 for an example of this). Of course, under some resource settings CES should fail to detect that its intentions have not been effective.

Time Step 1 is an example of situations where a fault has just occurred against a background of a normal plant, but there has been insufficient time for the effects of the fault to propagate. There is only one signal available to indicate the abnormality, in this case, a component (pressurizer spray valve) is misaligned. If a problem solving agent has a complete field of view of changes in the state of the plant (no resource limitation on monitoring), then the abnormality can be detected at this stage. CES can be run with a complete field of view for monitoring. With this setting, the misaligned component is immediately detected (as in Time Step 1 in this example run). There are circumstances where this may be a reasonable assumption about the monitoring behavior of the operational staff. For example, at a shift turnover the new crew often walks the board to review plant status; or in the course of verifying automatic system responses following a reactor trip the crew may walk the board and notice other changes in plant status (e.g., cases in Woods et al., 1982); or new personnel arriving to support the crew after an incident has begun will review plant status; or the functional organization of the Technical Support Center will dedicate several people just to monitor plant status.

However, this is generally not plausible from a human performance perspective due to resource limitations and the goal to minimize unnecessary effort. A particular abnormality can be directly detected even when there are monitoring resource limitations, if the change in state is salient either in form (the representation of plant behavior) or content. Some changes in some parameters may be so perceptually salient (e.g., an associated auditory alarm) that it triggers CES processing to assess what the change means in the current context. Similarly, some parameters may be so important that they are monitored at a sufficiently high rate to assure detection and processing of noteworthy changes in its behavior. This is expressed formally in CES by turning on or off the capability of changes in a datum to awaken its behavior analyst. If the data-driven capability is turned off, changes in the datum can still be observed, if another processing agent is interested in the behavior of this part of the plant.

Changes in pressurizer spray valve position are probably not particularly salient in the typical current control room. If an analyst judged this to be

the case in a particular control room, then a change in this valve position by itself should not be detected immediately and directly. In the CES modeling environment this means that a change in this piece of data by itself cannot trigger CES processing (i.e., it cannot trigger data-driven processing). In Processing Sample 4, the capability of the pressurizer spray valve to activate a behavior analyst is turned off. Thus, initially, the misaligned spray valve is not noticed, but the disturbance that results, decreasing primary system pressure, will be noticed. The the responsible situation analyst invoked to account for this unexplained decrease may then consider the state of the pressurizer spray process as one possible explanation in its investigations.

Differences in the representation of the plant available to the operational staff can make changes in a datum more or less salient, in other words, shift the processing from that in Processing Sample 4 to that in Sample 1. For example, a new computerized alarm system may have a rule that generates a perceptually salient message if a spray valve is open when primary system pressure is less than the automatic spray system actuation setpoint in appropriate plant modes. If the effect of this change were to be investigated, then CES should be run in a variety of NPP situations with the data-driven capability turned off for data on the pressurizer spray valve position (the current control room) and with the capability turned on (the control room plus new alarm system).

Response Management
This processing sample also illustrates the basic mechanism for activating corrective responses. An abnormal finding triggers activation of response plans that are linked to that perceived situation (response plans can also be triggered by the results of explanation building as in Processing Sample 3). This sample illustrates that corrective responses can be either manual human actions or automatic system actions. This sample also illustrates that response plans can be single actions, as at Time Step 1 or stages of responses as a disturbance increases in severity, as at Time Step 2 (response plans can also be sequences of actions).

The corrective response to a detected abnormality can be defined at several levels (see Woods & Hollnagel, 1987). For example in this case, the abnormality is, at the same time, a misaligned component and a process is active which should not be active. The corrective response to the first is to correctly align the component. The corrective response to the second situation assessment is to identify an action or actions that will make the process inactive, this includes correctly aligning the component in question, but it also includes other ways to make the process inactive (e.g., is there another valve which can be closed, i.e., how can the flow path be blocked?

can the pump be stopped, i.e., how can the force which moves the material be stopped?).

To be plausible both from a NPP point of view and a human performance point of view, the corrective response to be attempted first is the response closest to the observed abnormality, e.g., the abnormality is the valve is misaligned so that the first corrective response is to correctly align the valve (see cases in Pew et al., 1981 and in Woods et al., 1982). If this response is not successful, then the response plan analyst should shift its focus from the valve misalignment to the process disturbed by this – the pressurizer spray process is active when it should not be; therefore, consider alternative ways to make this process inactive that it knows about.

The depth of processing for response management to handle disturbances at different levels of abstraction is not currently implemented within CES.

When considering responses to perceived plant states, a response plan analyst can look ahead and anticipate responses that will be needed if the plant continues on its current trajectory. For example, in Time Step 2, when pressure crosses one abnormal limit, Response Plan Analyst 2 calls to mind a series of automatic systems responses that are to be expected as pressure continues to decrease. Alternatively, response plan analysts could be set up to call to mind and output corrective responses only when their specific triggering state has been met. In the case here this setting would mean that the automatic safety injection signal would not be called to mind until the pressure decrease crossed its activation setpoint.

4.4.2 CES Processing Sample: 2

Time Step 1

Initial conditions (t_0):
Plant is post-trip reactor power = turbine power = 0%; major parameters are near and moving towards target regions except pressurizer level has begun to fall (reference level = 25%; actual 19% and decreasing). The fault is a break in reactor coolant pump seal injection line D; charging line flow sensor reads 550 gpm; charging line pressure is less than primary system pressure; seal injection flow loop D reads 550 gpm (see Figure 4-9). CES enters situation with initial situation assessment that the plant is moving towards normal post-trip status.

Note: The fixation resistant explanation building strategy is used in this CES run.

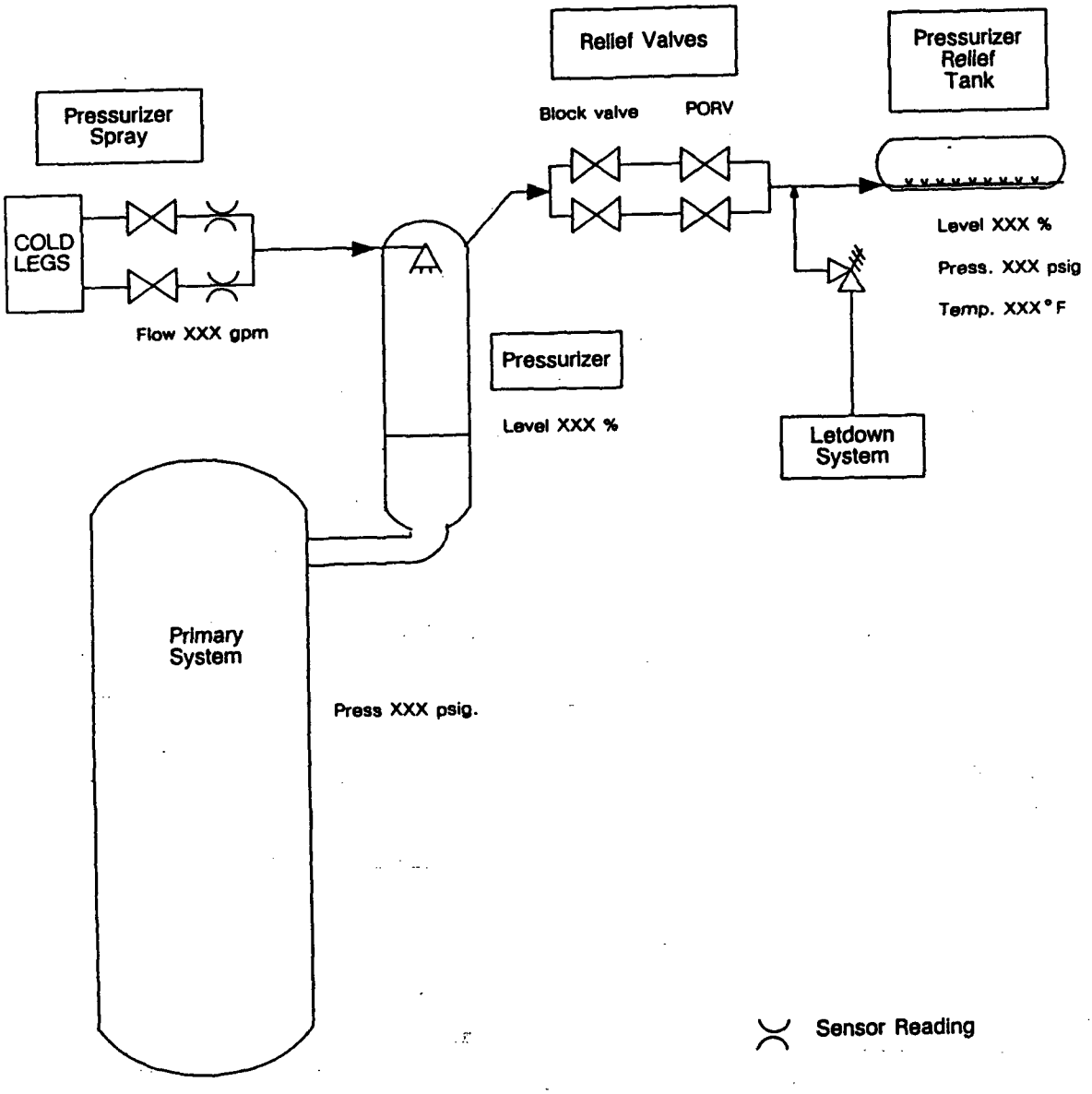


Figure 4-8: Partial representation of the primary system of a pressurized water reactor relevant to CES Processing Samples 1 and 4.

CES Processing

Figures 4-10 and 4-11 graphically illustrate CES processing at this time step. A behavior analyst (#1) awakes and observes that pressurizer level is in a decreasing progression (observation of an "interesting" plant behavior). This behavior is not accounted for by any known influence (net inflow had been slightly positive, i.e., charging flow slightly greater than letdown flow and all other mass in/outflows inactive).

Behavior Analyst 1 notes that pressurizer level is abnormally low (less than 20%). This observed abnormality triggers a response plan analyst (#1) to consider appropriate responses (Figure 4-11). The response plan analyst invokes the response script for decreasing pressurizer level. The first stage of response is automatic increase in charging flow, and the response plan analyst posts an expectation that this will occur. The second stage of response is automatic isolation of the letdown system, and the response plan analyst posts an expectation that the letdown system will automatically isolate. Behavior Analyst 1 continues to monitor for changes in pressurizer level.

A situation analyst (#1) awakes with the responsibility to explain the unexpected pressurizer level behavior. It activates (calls to mind) the knowledge that a decreasing progression in pressurizer level is associated with loss of mass (a break) from the primary system to containment (i.e., this is the strongest possibility given the evidence seen by this situation analyst). It then activates the knowledge that a loss of primary coolant break to containment (or loss of coolant) is associated with a decreasing progression in primary system pressure, an increasing progression in containment pressure, and indication of radiation in containment (Figure 4-10).

Situation Analyst 1 then directs monitoring activities by creating behavior analysts to determine if the behavior of these data match these projected observations (knowledge-driven monitoring). The behavior analysts report that a decreasing progression on primary system pressure is not observed, an increasing progression in containment pressure is observed and indication of radiation in containment is observed.

Time Step 2

Plant state t_0+200 seconds:

Letdown has isolated on pressurizer level less than 18%; pressurizer level is decreasing slowly (the net outflow from the primary system is 12 gpm, i.e., the total seal return flow); primary system pressure is stable; charging line flow sensor reads 550 gpm; charging line pressure is less than primary system pressure; seal injection flow loop D reads 550 gpm.

CES Processing

Figures 4-12, 4-13 and 4-14 graphically illustrate CES processing at this time step. Behavior Analyst 1 is monitoring pressurizer level which is still in a decreasing progression, but now at a very slow rate.

A behavior analyst awakes and observes an increasing progression in charging flow. This finding is consistent with the expected influence posted at Time Step 1 by Response Plan Analyst 1 that an automatic system would act to increase charging flow.

A behavior analyst awakes and observes valve HV8149A is closed (i.e., isolation of the letdown system). This finding is consistent with the expectation posted on the Active Influence Set at Time Step 1 by Response Plan Analyst 1 that an automatic system would act to isolate letdown.

Behavior Analyst 2 is monitoring primary system pressure. It does not observe a decreasing progression as would be consistent with the lead posted by Situation Analyst 1 at Time Step 1. Since the behavior is not seen and given that the time window allotted for its appearance has elapsed (checking for an expected behavior or a confirming observation can have an elapsed time criterion associated with it), this is noted as a finding by Situation Analyst 1.

Based on primary system pressure behavior (stable), Situation Analyst 1 can conclude that the influence of the postulated loss of coolant to containment on pressurizer level (i.e., the unknown level influence) is dominated by the influence of the charging system (i.e., maximum net charging).[6]

A behavior analyst awakes and observes an increasing progression in water flow through reactor coolant pump seal injection line D (Figure 4-13). This behavior is not accounted for by any known influence which awakens a situation analyst (#2) whose responsibility is to explain the unexpected behavior. It activates (calls to mind) knowledge that seal injection flow is associated with the charging system and triggers behavior analysts to check charging flow and charging line pressure. These analysts report back that charging flow is in an increasing progression and equals seal injection flow and that charging pressure is low and in a decreasing progression.

---

[6]This means that Situation Analyst 1 would use its knowledge about maximum charging flow (given this pump and the expected post-trip primary system pressure) to estimate the size of break if it did not check current charging flow, or it would check current charging flow and misinterpret this value as the size of break even though it is abnormally high (beyond the capability of the charging system to deliver at the current primary system pressure).

Based on this pattern of results, Situation Analyst 2 activates the possibility of an interfacing system break in the seal injection area -- charging pressure < primary system pressure; seal injection flow = charging flow (i.e., this is the strongest possibility given the evidence seen by this situation analyst).

Multiple situation analysts can communicate to consider possible convergence, i.e., assume that both findings in this case (decreasing pressurizer level and increasing seal injection flow) have a common explanation (Figure 4-14). For example, in the convergence process Situation Analyst 1 considers accounting for pressurizer level behavior given the possibility of an interfacing system break rather than a loss of coolant to containment. The finding about charging line pressure < primary system pressure changes the perceived current inflow to the primary system to zero; letdown is isolated; seal injection flow is zero (given the hypothesis of a seal line break); seal return flow is 12 gpm. This pattern of influences would account for pressurizer level behavior (a very small net outflow).

Thus, two or more situation analysts can work together to converge on a single explanation which accounts for all observed behavior. In this case, the best single explanation is the interfacing system break in seal injection: it accounts for containment behavior (high radiation; increasing pressure), pressurizer level behavior (low net outflow: inflow=0, outflow=seal return), primary system pressure behavior, charging system behavior (high flow, low pressure, high seal injection flow).

The accepted explanation of an interfacing system break in seal injection can then trigger a response plan analyst who looks up and outputs appropriate responses to this perceived situation, i.e., attempt to isolate the break. Expectations about changes in active influences, assuming isolation, are posted (e.g., charging line pressure increasing progression to normal range; net charging inflow).

Commentary: Sample 2

Explanation Building

For this processing sample CES explanation building was set up so that each unexpected finding generates a situation analyst (the fixation resistant setting on the explanation building strategy PAF). As a result, there are two active situation analysts who possess different views of the current situation (Figure 4-14). Starting with pressurizer level behavior, Situation Analyst 1 interprets the situation as a loss of coolant to containment (pressurizer level decreasing; containment abnormalities). As a result of the decreasing level, automatic systems have produced maximum net charging inflow to the primary system (letdown is isolated; charging flow has increased to maximum). Because primary system pressure is not falling and the level decrease has slowed, it

has concluded that maximum net charging can compensate for the break. Situation Analyst 1 has not seen and does not investigate data on the state of the charging system and reactor coolant pump seal injection and return.

Starting with reactor coolant pump seal injection flow in an increasing progression, Situation Analyst 2 interprets the situation as an interfacing system break. Charging flow is higher than is possible for current primary system pressure, assuming a normal (post-trip) charging system topology. Charging pressure is less than primary system pressure which means that there is no charging inflow to the primary system. Seal injection flow equals charging line flow indicating that all of the flow is going to seal injection. Containment indications (radiation; increasing pressure) show that there is an active path to containment from a break in the seal injection area.

The second interpretation is a better explanation because it covers all of the available findings. The first explanation is coherent only for a subset of the data. The two situation analysts can share results to converge on an explanation that accounts for the full range of findings within both areas of responsibility. This is done by postulating that a single explanation accounts for the findings of each situation analyst. Given this additional constraint, the interfacing system break hypothesis accounts for all of the findings.

Garden Path Behavior Prior to the observation of seal injection flow, CES is focused on a reasonable but erroneous explanation. If the charging/seal injection behavior is not seen or processed properly, then CES continues down this garden path because it already has a good explanation for the perceived state. This incident was given to three human subjects to solve in an unpublished study. One of the subjects went down the garden path of a loss of coolant break for a period of time.

Processing mechanisms and knowledge resources (performance adjustment factors) which would contribute to CES going down the garden path in this case include:

- A situation analyst is not activated to pursue findings in charging/seal injection because
    - the findings are not observed, which could occur because of
        - low salience (turn off data-driven capability of charging/seal data), or
        - monitoring limitations;

- a second situation analyst is not allowed to be created due to a single explanation assumption or a single explanation bias;

- the criterion for a good explanation is satisfied by the initial interpretation (loss of coolant) and further explanation building and/or monitoring is carried out only in light of the initial interpretation.

- the significance of the charging system and seal injection behavior is not recognized due to bugs in the knowledge base, i.e., missing the significance of the fact that charging pressure less than primary system pressure (there is no charging inflow to the primary system) and missing the significance of the fact that charging flow is higher than is possible for current primary system pressure, assuming a normal post-trip charging and primary systems topology;

- the convergence process between multiple situation analysts breaks down.

For example, if this incident is run with charging system and seal injection data obscured (i.e., the capability of changes in the charging and seal injection data to capture CES processing turned off), then this garden path behavior is exhibited. Exhibiting garden path behavior at one point in an unfolding incident does not necessarily mean that recovery from the erroneous situation assessment will not occur. The further evolution of the incident may produce salient evidence of discrepancies which trigger revision if PAF settings allow explanation building activities can be resumed.

Monitoring

This processing sample illustrates knowledge-driven monitoring. At several points, situation analysts create or prompt behavior analysts to monitor parts of the plant in order to pursue a possible explanation for an unexpected finding (knowledge-driven monitoring also can occur to verify expected corrective responses). In some cases, the behavior analyst is looking for a specific behavior, e.g., is primary system pressure in a decreasing progression as some possible explanation suggests.

In other cases, there is no expectation about what specific behavior to look for. Instead, there is a search for interesting behavior in items associated with a data-driven finding (what could this finding mean?). For example, Situation Analyst 2 follows up the seal injection flow finding by examining data on the state of the charging system.

This processing sample also illustrates the role of time in knowledge-driven monitoring. At Time Step 2, Situation Analyst 1 checks to see if, consistent with the possibility of a loss of coolant to containment, primary system pressure is decreasing. The behavior analyst reports back that pressure is stable. This is not inconsistent with the hypothesis because it may take time before expected behavior is manifested in dynamic systems. In other processing samples, the expected behavior does occur on a future time step and is reported back to the relevant situation analyst and adds weight to the hypothesis which set the lead. In the case here, the expected behavior does not occur. At what point does the non-occurrence of the expected behavior constitute a finding (a negative finding)? This is done by associating with items in the knowledge base a time interval which expresses knowledge about the expected time interval for the behavior to be manifested (e.g., after a pump is turned off, flow may continue due to gradual coast down). When the interval has elapsed following an inquiry, the non-occurrence is counted as a finding by the situation analyst. On this basis, Situation Analyst 1 treats the non-occurrence of a primary system pressure decrease as a finding to be used in explanation building.

In this processing sample, response plan analysts post expected influences based on knowledge about corrective responses. Monitoring to determine whether expected responses in fact do occur could be done at three levels. The weakest level is simply to post the expectation; if the expected change is observed (e.g., data-driven), then confirmation will occur (the behavior will be evaluated as expected and this report communicated to the relevant response plan analyst). The strongest level is to post the expected change and to trigger monitoring activities to look for this change (e.g., letdown isolation at Time Step 2). The intermediate case is to post the expected change and to increase CES's receptivity to the relevant data.

Calling to mind possible explanations

The phrase "knowledge is called to mind that x would account for y" is used several times. This means that, when an unexpected finding is noted, the set of possible explanations are ordered based on the strengths of relationships encoded in the knowledge base and on the evidence evaluated by that situation analyst up to that point. Knowledge about one possible explanation is "activated" or "called to mind," i.e., the top item on the ordered list, and it directs current explanation building processing, e.g., knowledge-driven monitoring. For example, there are several possibilities linked to the finding of unexpected pressurizer level decrease in the knowledge base: change in net charging, break to containment, a steam generator tube rupture, a decrease in primary system energy (note that these also can be hierarchically organized). Given no other evidence, the relationship with the greatest strength is "called to mind" and directs further

processing. All of the possibilities encoded in the knowledge base can be entertained eventually. However, processing can stop because a "good" explanation was found or because of limited resources or limited time.

The order that possible explanations direct processing also depends on the order findings are noted. The relative strength of a possibility depends on the evidence processed up to that point. The links and strengths built into the knowledge base permit different knowledge to be activated depending on the path taken.

Convergence between multiple situation analysts involves postulating a single explanation that accounts for all of the triggering unexpected findings. This means that possible explanations common to the set of situation analysts are put into a new list and re-evaluated in light of all of the observations available to all of the situation analysts.

Commitment to an explanation
Operator actions depend not only on building explanations to account for patterns of findings, but also on *commitment* to an explanation. Commitment refers to when an explanation is communicated to trigger response plan analysts to begin selecting appropriate corrective responses. A problem solver can err by committing too quickly (examine too little evidence) or too late (examine too much evidence or wait too long for more evidence to accrue). Varying the criteria governing commitment (in the Decision strategist) varies how much evidence search will go on.

Commitment to an explanation can depend on the strength associated with that potential explanation, whether there are any strong competitors. Interaction with the response management activities is needed so that knowledge of consequences can be brought to bear (part of the choice under uncertainty and risk mechanisms which are currently not built into CES).

4.4.3 CES Processing Sample: 3

Time Step 1

Initial conditions (t_0):
Plant is at steady state with reactor power = turbine power = 50%; major parameters are all on target.

Plant state at t_0+30 seconds:
Turbine throttle valve has failed and is ramping to 100% open over two minutes which takes the turbine to 100% power. Power mismatch due to

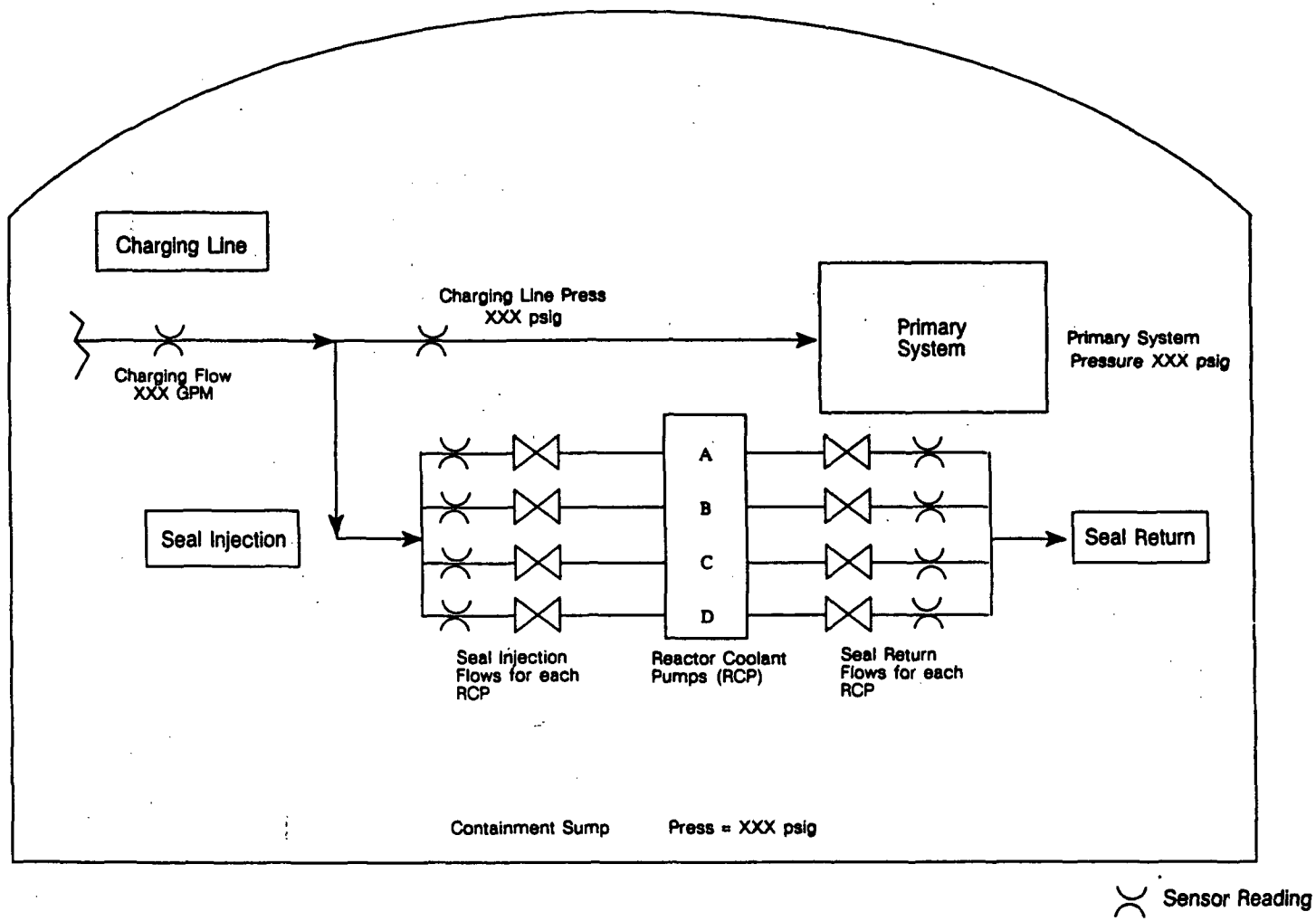


Figure 4-9: Partial representation of the primary system and the charging system of a pressurized water reactor relevant to CES Processing Samples 2.

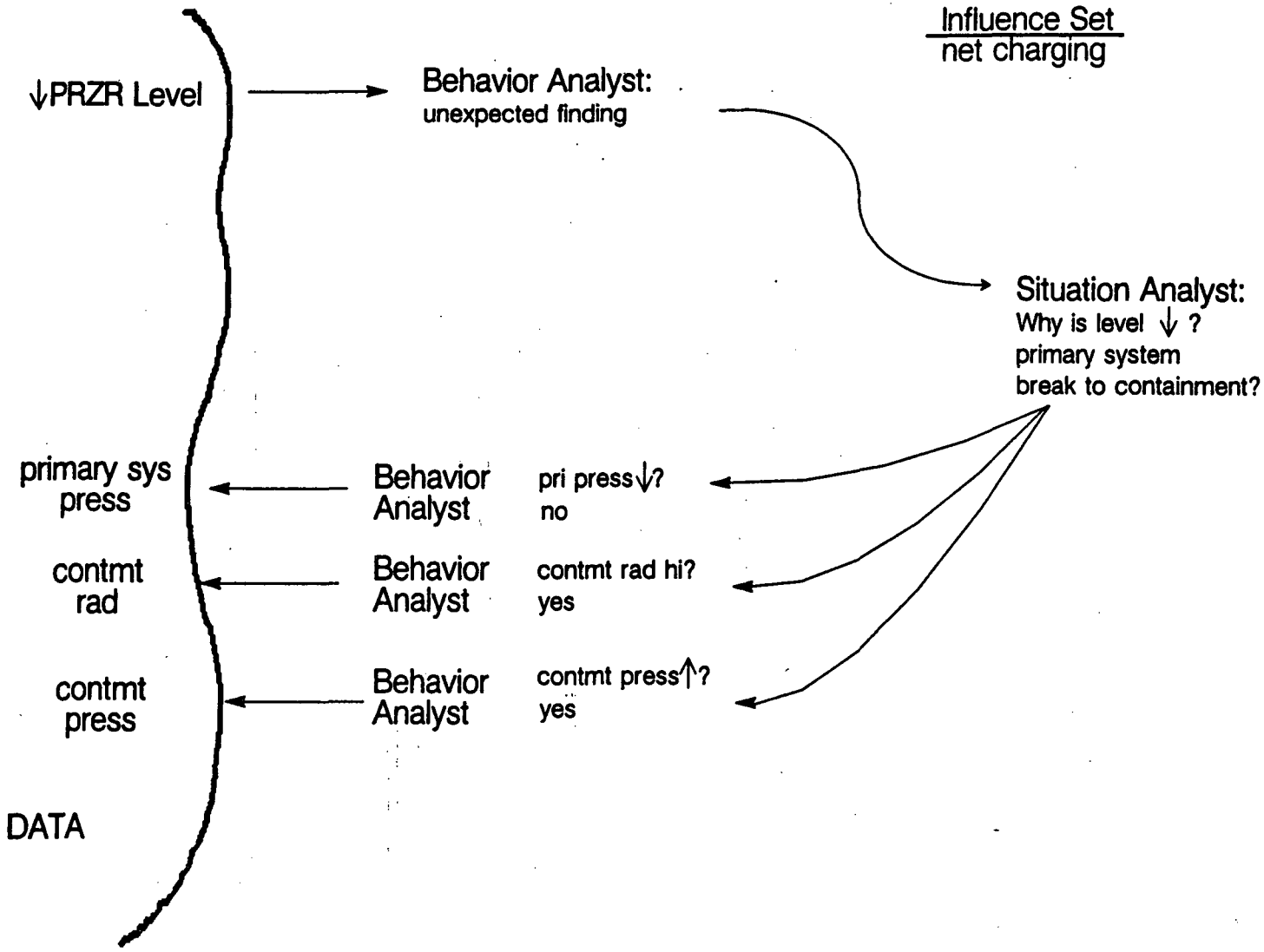


Figure 4-10: CES processing at Time Step 1 of Processing Sample 2: Unexpected level decrease and situation analyst behavior.

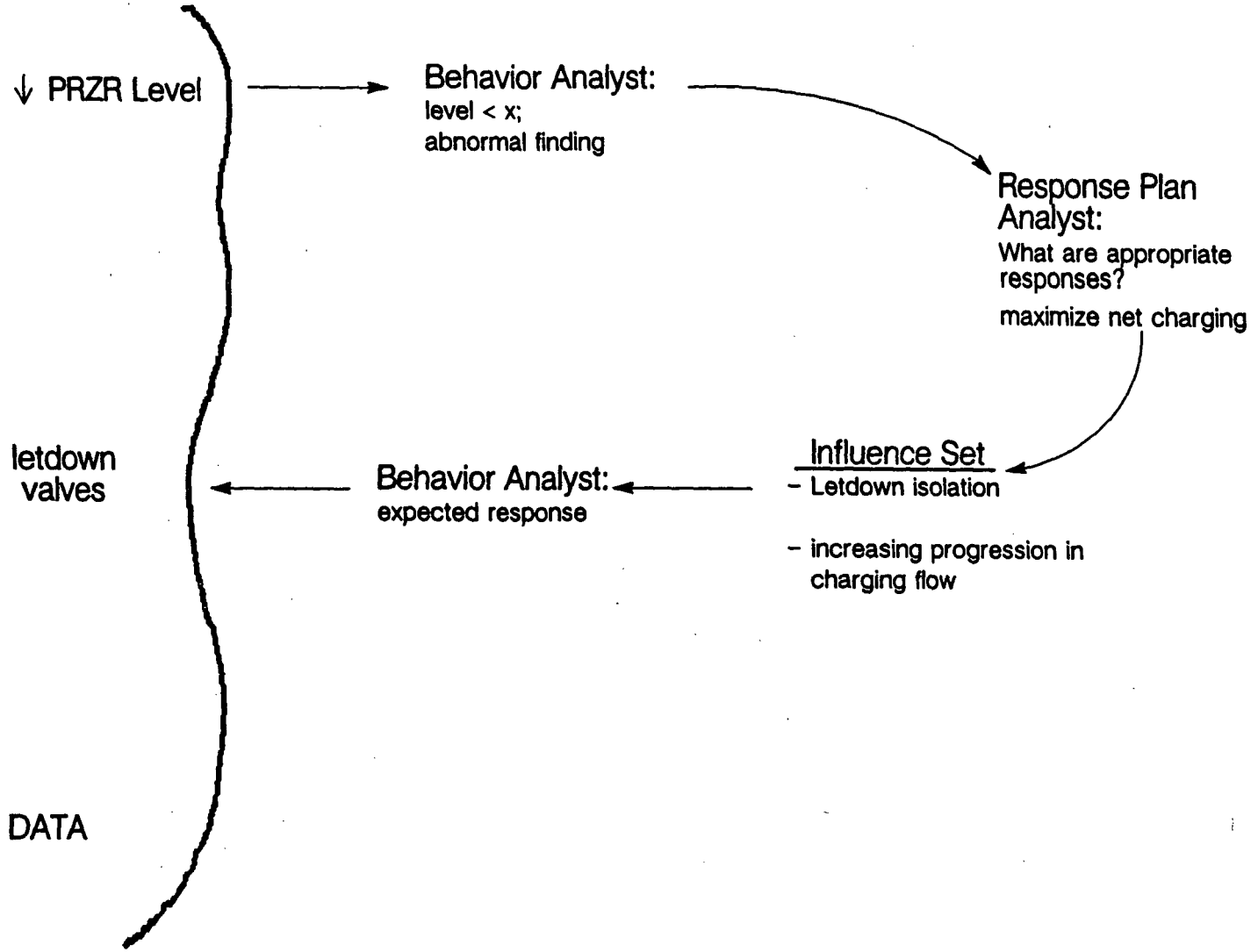


Figure 4-11: CES processing at Time Step 1 of Processing Sample 2: Response plan analysts.

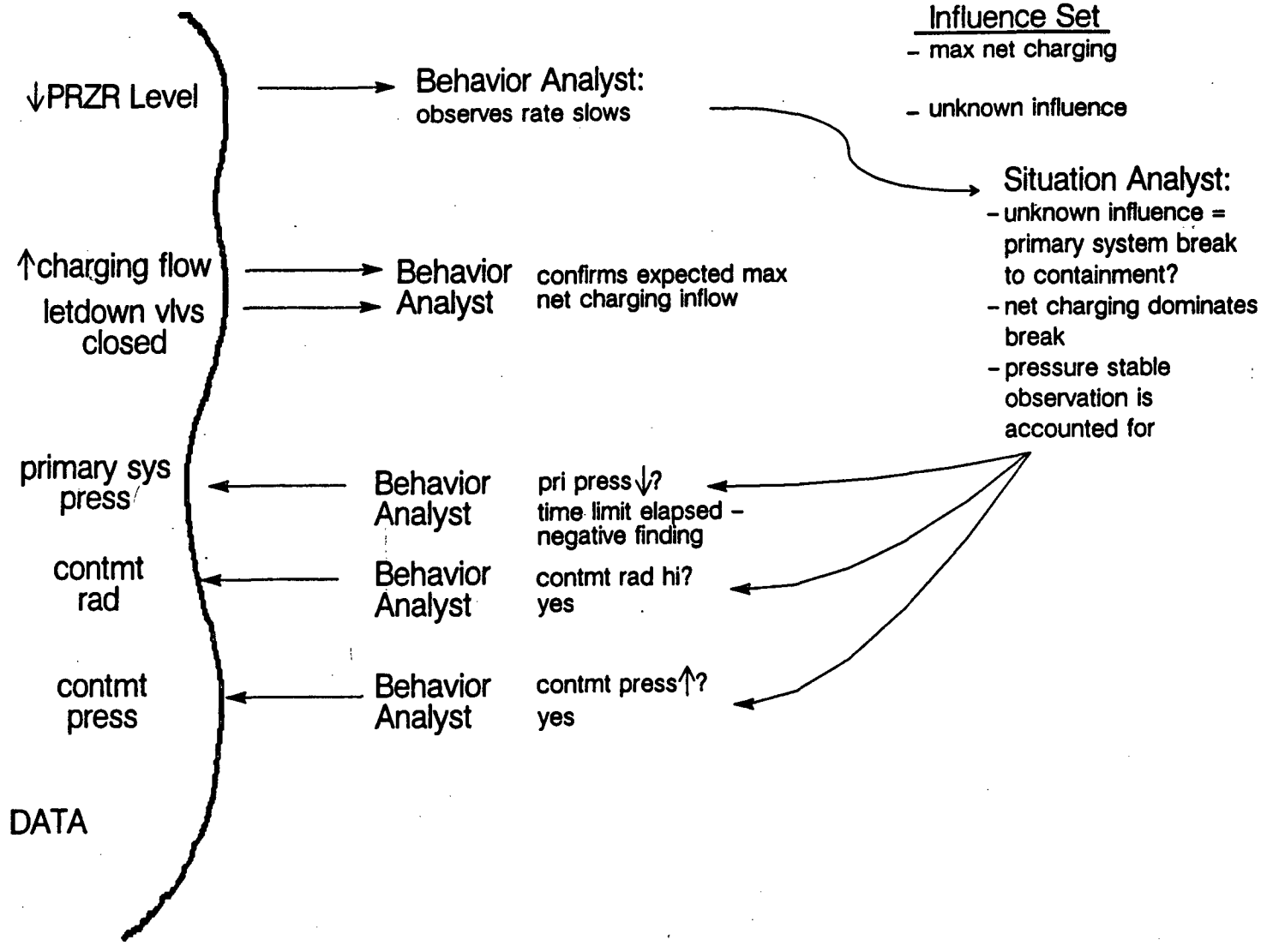


Figure 4-12: CES processing at Time Step 2 of Processing Sample 2: Unexpected level decrease and situation analyst behavior.

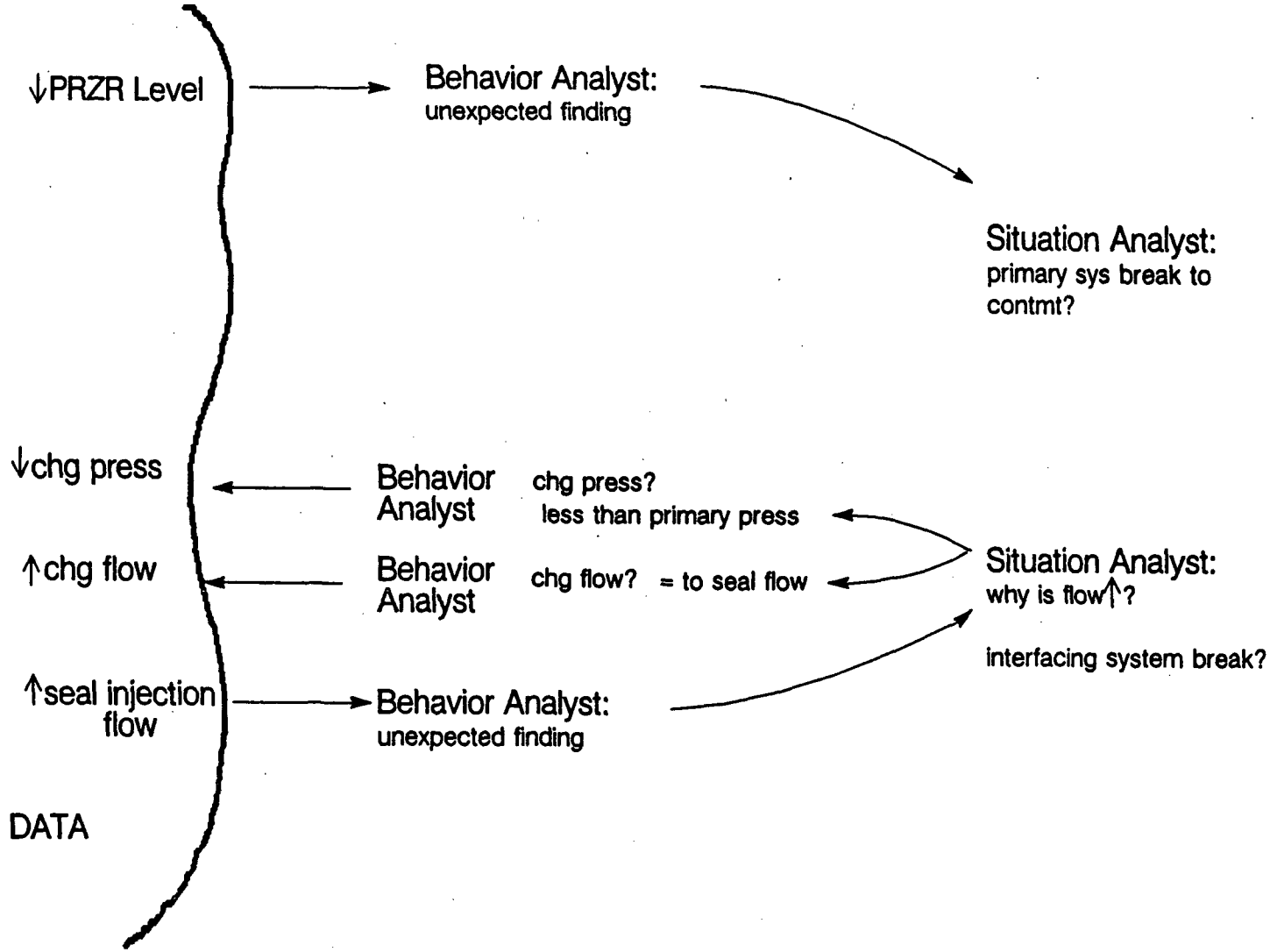


Figure 4-13: CES processing at Time Step 2 of Processing Sample 2: Unexpected flow increase and situation analyst behavior.

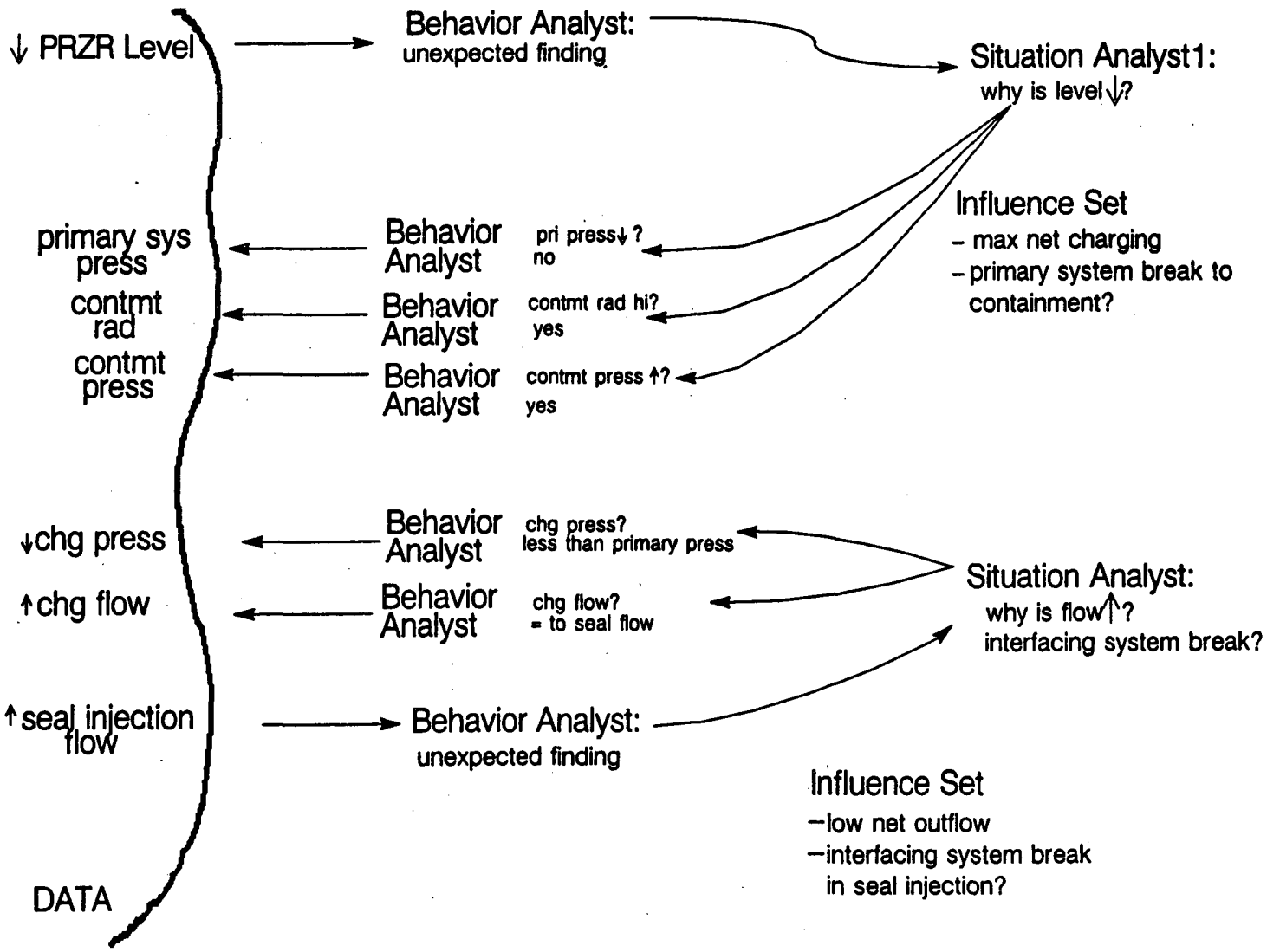


Figure 4-14: CES processing at Time Step 2 of Processing Sample 2: Multiple views of plant state.

excessive energy outflow (reactor power = 56%; turbine power = 60%) results in steam generator pressures less than normal and decreasing, steam flows rapidly increasing, primary system temperature (t_{hot}, t_{cold}, and t_{avg}) less than normal and decreasing, primary system hotleg-coldleg delta t greater than normal (t_{hot} minus t_{cold}), and primary system pressure less than normal and decreasing.

CES Processing
A behavior analyst (#1) awakes and observes a decreasing progression in primary system temperature, t_{avg}, (observation of an "interesting" plant behavior). This behavior is not accounted for by any known influence (the plant had been steady state at 50% power).

Behavior Analyst 1 notes that t_{avg} is abnormally low (less than t_{ref}). This abnormality triggers a Response Plan Analyst (#1) to consider appropriate responses. Response Plan Analyst 1 generates an expectation that an automatic system (rod control) will act (pull rods) and posts the expected influences on automatic rod control (moving out), nuclear power (increasing progression) and t_{avg} (increasing progression). Behavior Analyst 1 continues to monitor t_{avg}.

A situation analyst (#1) awakes with the responsibility to explain the unexpected t_{avg} behavior. It activates (calls to mind) the knowledge that a decreasing progression in t_{avg} is associated with a decreasing progression in primary system energy. Situation Analyst 1 then activates the knowledge that, if this is the case, then it would be expected that pressurizer level is in a decreasing progression and that primary system pressure is in a decreasing progression.

Situation Analyst 1 then directs monitoring activities to determine if the behavior of these data match these projected observations (knowledge-driven monitoring). It creates a behavior analyst (#2) to monitor primary system pressure (if a behavior analyst was already active for a parameter, the situation analyst would only prompt it for the currently observed behavior). Behavior Analyst 2 reports decreasing primary system pressure. This observation is consistent with decreasing primary system energy. Situation Analyst 1 creates another behavior analyst (#3) to monitor pressurizer level behavior. This behavior analyst does not observe a decreasing progression on pressurizer level. This is not inconsistent with decreasing primary system energy because it may take more time before this behavior is manifested.

Behavior Analyst 2 also observes that primary system pressure is abnormally low (less than 2210) which triggers a response plan analyst (#2) to consider

appropriate responses to correct this. The response plan analyst invokes the response script for decreasing primary system pressure. One stage of response is automatic reactor trip, and the response plan analyst posts an expectation that an automatic reactor trip will occur. Another stage of response is an automatic safety injection signal (SI) and initiation of emergency core cooling (ECCS), and the response plan analyst posts an expectation that an automatic SI signal and ECCS initiation will occur.

Situation Analyst 1 deposits decreasing progression in primary system energy as a plant behavior. Since there is no explanation to account for this behavior, a situation analyst (#2) is set up to identify the unknown influence (a behavior analyst mediates the chaining from Situation Analyst 1 to Situation Analyst 2). It activates the knowledge that a decreasing progression in primary system energy is associated with an increasing progression in energy transport from the primary system to the secondary system (or pri-sec energy transport).

Situation Analyst 2 then activates the knowledge that, if this is the case, then it would be expected that the temperature difference between the hotleg and coldleg on each loop in the primary system (or primary system delta t) is in an increasing progression and that steam flow from the steam generators is high. Situation Analyst 2 creates one behavior analyst to monitor the primary system delta t, who observes an increasing progression, and another to monitor steam flow, who observes that steam flow is higher than feed flow and increasing.

Situation Analyst 2 deposits increasing progression in pri-sec energy transport as a plant behavior. Since there is no explanation to account for this behavior, a situation analyst (#3) is set up to identify the unknown influence (again, a behavior analyst mediates this chaining). The situation analyst activates the knowledge that an increasing progression in pri-sec energy transport is associated with increased energy/steam outflow from the secondary system. It then activates the knowledge that there are two avenues for this (excessive turbine demand and steam break) and begins to drive monitoring activities to distinguish these possibilities.

Time Step 2

Plant state t_0+150 seconds:
Just prior to automatic reactor shutdown. Throttle valve is full open; turbine power = 100% reactor power = 87%. The pattern of the previous time step continues; the deviations are larger. Pressurizer level is now low and decreasing.

CES Processing
Behavior Analyst 1 still observes a decreasing progression in primary system temperature, t_{avg}. There are two influences acting on this parameter a decreasing progression from an unknown influence and an increasing progression from increasing nuclear power (automatic rod control). The behavior analyst notes and stores the dominance pattern, i.e., the unknown influence on t_{avg} is stronger than the automatic rod control system.

Behavior Analyst 2 is monitoring pressurizer level behavior (knowledge-driven by Situation Analyst 1). It observes a decreasing progression in pressurizer level. This is communicated to Situation Analyst 1 who notes that this behavior is consistent with and reinforces the postulated decreasing progression in primary system energy.

The evidence continues to support (and the increase in the expected direction reinforces) the conclusions of the situation analysts about plant behavior -- decreasing progression in primary system energy, increasing progression in pri-sec energy transport, and increasing progression in energy/steam outflow from the secondary system.

Time Step 3

Plant state t_0+500 seconds:
An automatic reactor shutdown (due to low primary system pressure) and turbine shutdown have occurred. The turbine shutdown effectively isolates the effect of the stuck open throttle valve. Plant parameters begin to return to normal hot shutdown state.

CES Processing
A behavior analyst (#4) awakes and observes an automatic reactor shutdown (data-driven observation of an "interesting" finding). This finding triggers expected influences on a number of plant parameters and systems including: control rods, nuclear power, reactivity, turbine trip, turbine power, auxiliary feedwater pumps, auxiliary feedwater valve alignment, auxiliary feedwater flow, feedwater isolation, and primary system temperatures (moving to post-trip targets). This finding confirms the expectation posted earlier based on the observed decreasing progression in primary system pressure.

This observed behavior triggers a response plan analyst (#3) to consider appropriate responses in this situation. The response plan analyst invokes the response script for reactor trip. This script initiates a strong emphasis on observing plant state both generally (increased sensitivity to interesting changes) and specific items to be checked including control rods, nuclear power, reactivity, primary system, turbine trip, turbine power, auxiliary

feedwater pumps, auxiliary feedwater valve alignment, auxiliary feedwater flow, feedwater isolation, primary system temperatures moving to post-trip targets, primary system pressure, containment conditions (pressure, temperature, radiation, etc.), reactor coolant pump trip criteria.

Behavior analysts report (among other things) primary system temperatures moving to post-trip targets (e.g., t_{avg} increasing progression towards post-trip target) and primary system pressure increasing progression towards target.

As t_{avg} is now in an increasing progression towards target, Response Plan Analyst 1 ends the response script for low t_{avg} and clears the associated expected influences for automatic rod control and nuclear power.

As primary system pressure is now in an increasing progression towards target, Response Plan Analyst 2 ends the response script for low pressure and clears the associated expected influences, i.e., an automatic safety injection signal and initiation of emergency core cooling.

The situation analysts now conclude from the evidence of temperatures and pressure moving towards their post-trip targets that primary system energy is in an increasing progression, pri-sec energy transport is in a decreasing progression, and energy/steam outflow from the secondary system is in a decreasing progression.

At Time Step 1 CES had concluded there was an unknown influence acting on t_{avg}, primary system energy, pri-sec energy transport, and energy/steam outflow from the secondary system and that the unknown influence dominated the automatic rod control system response. At this time step, influences are removed (i.e., the reactor and turbine trips) and the behavior reverses (the temperatures, pressure are moving towards their post-trip targets). This means that the unknown influence has been removed (there is no new influence that could account for the reversal) and the plant is returning to a "normal" state.

Commentary: Sample 3

Explanation Building
This processing sample illustrates several aspects of explanation building.

At Time Step 1 the situation analysts build a higher level (i.e., not directly observable) characterization of the behavior of the plant out of the pattern of directly observable evidence. Thus, Situation Analyst 1 concludes that there is a decreasing progression in primary system energy because t_{avg} and primary system pressure are decreasing. This characterization then activates

knowledge that directs processing activities in a particular direction, in this case, energy flow disturbances.

The inferred plant behavior can then be treated as any other plant behavior – it can be tracked by a behavior analyst, changes can be judged expected or unexpected, explanations sought for unexpected changes, and abnormal changes can trigger responses via a response plan analyst.

The result of the explanation building at Time Step 1 is that there is an unknown influence acting on t_{avg}, primary system energy, pri-sec energy transport, and energy/steam outflow from the secondary system. The unknown influence dominates the automatic rod control system response. The source of this influence is being sought in the source of the increased energy/steam outflow from the secondary system.

Note that CES has linked several disturbances into a single disturbance chain, i.e., they have a common source. It also has made a judgment about where in the disturbance chain to pursue the common source (this is related to the concept of topographic search in diagnosis; Rasmussen, 1986).

At Time Step 3, the challenge for explanation building is what accounts for the reversal of the abnormal trends: has the unknown influence been eliminated by the changed pattern of influences (the turbine trip) as is actually the case here or has another influence acting in the opposite direction dominated the first unknown influence.

Monitoring
In this case, there are no resource limits on monitoring, e.g., there is no limit on the number of behavior analysts that can be active simultaneously. This case illustrates how quickly the number of active behavior analysts can grow and how limits on monitoring resources can have a large impact on problem solving behavior.

There is an issue that arises when feedback is sought to verify expected actions (either automatic or manual). The feedback can come at several levels; for example, in this case the automatic system response to low t_{avg} could be seen in an automatic system activation signal, control rod movement, the effect on nuclear power, the effect on t_{avg}. Operational personnel do not always have unlimited resources to check for all of these effects to confirm expected responses. The question then is which, if any, of these potential signs will be checked? Relative salience of the different pieces of data is one factor that would affect which data are checked. Monitoring actions encoded in response scripts are another as illustrated for reactor trip

response in Time Step 3. If an inconsistency among evidence is noted, then the complete set (or a fuller set of evidence) could be pursued more deeply (this relates to sets of evidence for issues, the possibility of inconsistent evidence, and how completely an evidence set is examined).

Not only are expected responses added to the influence set, they sometimes must also be removed from the influence set. In Time Step 3 there is a case where an expected response needs to be removed from the influence set because the situation has changed before the response could occur.

Response Management
Response scripts can be tied to the observable datum or to the abstract issue behind the data or to both, e.g., responses to abnormal primary system energy and responses to abnormal t_{avg}.

Note that the change in t_{avg} and the reactor trip at Time Step 3 both change influences on automatic rod control and nuclear power.

4.4.4 CES Processing Sample: 4

Time Step 1

Initial conditions (t_0):
Plant is at steady state with reactor power = turbine power = 50%; major parameters are all on target; pressurizer spray valve has just stuck open (see valve PCV-455B in Figure 4-8).

Plant state at t_0+30 seconds:
Turbine throttle valve has failed and is ramping to 100% open over two minutes which takes the turbine to 100% power. Power mismatch due to excessive energy outflow (reactor power = 56%; turbine power = 60%) results in steam generator pressures less than normal and decreasing, steam flows rapidly increasing, primary system temperature (t_{hot}, t_{cold}, and t_{avg}) less than normal and decreasing, primary system hotleg-coldleg delta t greater than normal (t_{hot} minus t_{cold}), and primary system pressure less than normal and decreasing.

Note: In this CES run, the capability of pressurizer spray valve position indications to trigger data-driven processing has been turned off and the single explanation bias setting on explanation building strategy is used.

CES Processing

A behavior analyst (#1) awakes and observes a decreasing progression in primary system temperature, t_{avg}. This behavior is not accounted for by any known influence (the plant had been steady state at 50% power).

Behavior Analyst 1 notes that t_{avg} is abnormally low (less than t_{ref}). This abnormality triggers a Response Plan Analyst (#1) to consider appropriate responses. Response Plan Analyst 1 generates an expectation that an automatic system (rod control) will act (pull rods) and posts the expected influences on automatic rod control (moving out), nuclear power (increasing progression) and on t_{avg} (increasing progression). Behavior Analyst 1 continues to monitor t_{avg}.

A situation analyst (#1) awakes with the responsibility to explain the unexpected t_{avg} behavior. It activates the knowledge that a decreasing progression in t_{avg} is associated with a decreasing progression in primary system energy. Situation Analyst 1 then activates the knowledge that, if this is the case, then it would be expected that pressurizer level is in a decreasing progression and that primary system pressure is in a decreasing progression.

A behavior analyst (#2) awakes and observes a decreasing progression in primary system pressure. Given that CES is being run with the single explanation bias strategy for explanation building, this finding is "absorbed" by Situation Analyst 1 because it is relevant to the possible explanations it is considering.

Situation Analyst 1 then directs monitoring activities to determine if the behavior of these data match these projected observations (knowledge-driven monitoring). It prompts for primary system pressure behavior and Behavior Analyst 2 communicates its observation of a decreasing progression. This observation is consistent with decreasing primary system energy. Situation Analyst 1 creates another behavior analyst (#3) to monitor pressurizer level behavior. This behavior analyst does not observe a decreasing progression on pressurizer level. This is not inconsistent with decreasing primary system energy because it may take more time before this behavior is manifested.

Behavior Analyst 2 also observes that primary system pressure is abnormally low (less than 2210) which triggers a response plan analyst (#2) to consider appropriate responses to correct this. The response plan analyst invokes the response script for decreasing primary system pressure. One stage of response is automatic reactor trip, and the response plan analyst posts an expectation that an automatic reactor trip will occur. Another stage of response is an

automatic safety injection signal (SI) and initiation of emergency core cooling (ECCS), and the response plan analyst posts an expectation that an automatic SI signal and ECCS initiation will occur.

Situation Analyst 1 deposits decreasing progression in primary system energy as a plant behavior. Since there is no explanation to account for this behavior, a situation analyst (#2) is set up to identify the unknown influence (a behavior analyst mediates the chaining from Situation Analyst 1 to Situation Analyst 2). It activates the knowledge that a decreasing progression in primary system energy is associated with an increasing progression in energy transport from the primary system to the secondary system (or pri-sec energy transport).

Situation Analyst 2 then activates the knowledge that, if this is the case, then it would be expected that the temperature difference between the hotleg and coldleg on each loop in the primary system (or primary system delta t) is in an increasing progression and that steam flow from the steam generators is high. Situation Analyst 2 creates one behavior analyst to monitor the primary system delta t, who observes an increasing progression, and another to monitor steam flow, who observes that steam flow is higher than feed flow and increasing.

Situation Analyst 2 deposits increasing progression in pri-sec energy transport as a plant behavior. Since there is no explanation to account for this behavior, a situation analyst (#3) is set up to identify the unknown influence (again, a behavior analyst mediates this chaining). The situation analyst activates the knowledge that an increasing progression in pri-sec energy transport is associated with increased energy/steam outflow from the secondary system. It then activates the knowledge that there are two avenues for this (excessive turbine demand and steam break) and begins to drive monitoring activities to distinguish these possibilities.

Time Step 2

Plant state t_0+150 seconds:
Just prior to automatic reactor shutdown. Throttle valve is full open; turbine power = 100% reactor power = 87%. The pattern of the previous time step continues; the deviations are larger. Pressurizer level is now low and decreasing. Pressurizer spray valve remains stuck open.

CES Processing
Behavior Analyst 1 still observes a decreasing progression in primary system temperature, t_{avg}. There are two influences acting on this parameter a decreasing progression from an unknown influence and an increasing

progression from increasing nuclear power (automatic rod control). The behavior analyst notes and stores the dominance pattern, i.e., the unknown influence on t_{avg} is stronger than this the automatic rod control system.

Behavior Analyst 2 is monitoring pressurizer level behavior (knowledge-driven by Situation Analyst 1). It observes a decreasing progression in pressurizer level. This is communicated to Situation Analyst 1 who notes that this behavior is consistent with and reinforces the postulated decreasing progression in primary system energy.

The evidence continues to support (and the increase in the expected direction reinforces) the conclusions of the situation analysts about plant behavior — decreasing progression in primary system energy, increasing progression in pri-sec energy transport, and increasing progression in energy/steam outflow from the secondary system.

Time Step 3

Plant state t_0+500 seconds:
An automatic reactor shutdown (due to low primary system pressure) and turbine shutdown have occurred. The turbine shutdown effectively isolates the effect of the stuck open throttle valve. Plant parameters begin to return to normal hot shutdown state except primary system pressure because the pressurizer spray valve remains stuck open.

CES Processing
A behavior analyst (#4) awakes and observes an automatic reactor shutdown. This finding triggers expected influences on a number of plant parameters and systems including: control rods, nuclear power, reactivity, turbine trip, turbine power, auxiliary feedwater pumps, auxiliary feedwater valve alignment, auxiliary feedwater flow, feedwater isolation, and primary system temperatures (moving to post-trip targets). This finding confirms the expectation posted earlier based on the observed decreasing progression in primary system pressure.

This observed behavior triggers a response plan analyst (#3) to consider appropriate responses in this situation. The response plan analyst invokes the response script for reactor trip. This script initiates a strong emphasis on observing plant state both generally (increased sensitivity to interesting changes) and specific items to be checked including control rods, nuclear power, reactivity, primary system, turbine trip, turbine power, auxiliary feedwater pumps, auxiliary feedwater valve alignment, auxiliary feedwater flow, feedwater isolation, primary system temperatures moving to post-trip targets, primary system pressure, containment conditions (pressure, temperature, radiation, etc.), reactor coolant pump trip criteria.

Behavior analysts report (among other things) primary system temperatures moving to post-trip targets (e.g., t_{avg} increasing progression towards post-trip target) but primary system pressure continues in a decreasing progression although at a slower rate.

As t_{avg} is now in an increasing progression towards target, Response Plan Analyst 1 ends the response script for low t_{avg} and clears the associated expected influences for automatic rod control and nuclear power.

The situation analysts now conclude from the evidence of temperatures moving towards their post-trip targets that primary system energy is in an increasing progression, pri-sec energy transport is in a decreasing progression, and energy/steam outflow from the secondary system is in a decreasing progression.

At Time Step 1 CES had concluded there was a single unknown influence acting on t_{avg}, primary system energy, primary system pressure, pri-sec energy transport, and energy/steam outflow from the secondary system and that the unknown influence dominated the automatic rod control system response. At this time step, influences are removed (i.e., the reactor and turbine trips) and the behavior reverses (the temperatures, pressure are moving towards their post-trip targets). This means that the unknown influence has been removed and the plant should be returning to a "normal" state. The situation analysts set up to account for unexpected behaviors in primary system energy, pri-sec energy transport, and energy/steam outflow from the secondary system are cleared (note no situation analyst had been set up previously whose responsibility was to pursue primary system pressure behavior).

Given removal of the influence producing a decreasing progression in primary system energy and the known influences acting on primary system pressure, Behavior Analyst 2 notes that the continuing decreasing progression in pressure is now unexpected.

A new situation analyst awakes with the responsibility to explain the unexpected primary system pressure behavior. It activates (calls to mind) the knowledge that the pressurizer spray system and a primary system break are associated with the observed pressure behavior. It then activates the knowledge that, if the former were active, then a pressurizer spray valve would be open and activates a behavior analyst to monitor these valves.

A behavior analyst awakes and observes that a pressurizer spray valve (PCV-455B) is open. The valve open indicates the pressurizer spray process

is an active influence on primary system pressure (decreasing progression) and accounts for the observed pressure behavior.

This behavior analyst also notes that the pressurizer spray process is active when primary system pressure is below the automatic system setpoint (2260 psig). This is an abnormality and triggers a response plan analyst to consider appropriate responses to this situation. The first response would be for an operator to correctly align the valve. The response plan analyst generates an intention to close the spray valve and it posts the expectation that the valve will be closed on the Active Influence Set for the pressurizer spray process and the relevant behavior analyst monitors for this change.

Time Step 4

Plant state t_0+700 seconds:
Spray valve is manually closed; primary system pressure begins to recover.

CES Processing
A behavior analyst is monitoring the pressurizer spray process. It observes that spray valve PCV-455B has closed. This confirms the expectation posted at Time Step 3.

The pressurizer spray process is cleared as an active influence on primary system pressure.

Another behavior analyst (#2) is monitoring primary system pressure and observes a change -- it is now in an increasing progression.

This behavior is consistent with the perceived active influences on primary system pressure (in the absence of other influences primary system pressure will move back to the current equilibrium value for his plant mode, e.g., 2235 psig).

Commentary: Sample 4

In this incident two faults are present simultaneously. In particular, the incident results from combining the faults present in Processing Samples 1 and 3. Thus, this sample illustrates the kinds of reasoning that can arise in dealing with different kinds of multiple fault situations.

Monitoring
The capability of pressurizer spray valve position indications to trigger data-driven processing was been turned off for this run (i.e., changes in these valve positions cannot capture CES processing). As a result and in contrast to Processing Sample 1, the misaligned valve and the disturbance in the

pressurizer spray process as a method of pressure control (i.e., spray is on when pressure is below the automatic activation setpoint) are not noticed directly. The next avenue to find this fault is to detect the effect on primary system pressure (a resulting disturbance). However, there is another fault present that results in an influence on primary system pressure in the same direction as well as influences on other parts of the plant.

CES at Time Step 1 notices and pursues the unexplained t_{avg} behavior first. In the process of considering potential explanations for this behavior, Situation Analyst 1 triggers monitoring of primary system pressure. The observed decreasing progression is consistent with a decreasing progression in primary system energy. This explanation accounts for both observations, and CES continues to pursue explanations that would account for the inferred decreasing progression in primary system energy.

CES could easily have taken a different tack (depending on the relative salience values attached to each observable piece of data) where decreasing pressure was noticed and pursued as an unexpected finding. If this occurred, then the strength values between the observed pressure behavior and possible explanations would govern the search order. Decreasing primary system energy is linked to both decreasing pressure and decreasing energy in the knowledge base but the relative strength values are different; thus, the order in which it would control processing is different). For example, three human subjects were given this incident to solve in an unpublished study. One noticed the pressure decrease and began his search by checking the spray system status. As a result, he first discovered the stuck open spray valve (the other two first pursued explanations for decreasing primary energy).

This alternative flow illustrates how the order of observation can be affected by the relative salience of incoming data and how the order of observation can affect later processing (especially if there is resource competition).

Differences in the representation of the plant available to the operational staff can make changes in a datum more or less salient and more or less observable. For example, a new computerized alarm system may have a rule that generates a perceptually salient message if a spray valve is open when primary system pressure is less than the automatic spray system actuation setpoint in appropriate plant modes. This change in the representation of the plant would shift CES monitoring behavior from something like that in Processing Sample 4 to something like that in Sample 1, depending on the particular incident and other PAF settings.

Explanation Building
CES first pursues factors that affect primary system energy (e.g., steam line

break). Following the reactor trip, the pattern of influences change because the effects of the stuck open turbine governor valve are eliminated. In Processing Sample 3, CES must detect that the influence of the turbine fault is removed and that all monitored parameters are trending back towards the normal values for this plant mode. In this sample, CES must detect that disturbances persist after some automatic or manual action has been taken. In this case, it detects that the turbine fault is removed, but the behavior analyst set up earlier to monitor primary system pressure notices that primary system pressure continues to decrease.

This observation is unexplained (the primary energy influence has been removed) and triggers a situation analyst to pursue what could account for it. Note that, because CES first pursued the primary system energy issues and because CES was run with the single explanation bias setting on explanation building strategy, no situation analyst was set up earlier to pursue the decreasing progression in primary system pressure as an unexplained finding.

Also note that, when the decrease in primary system pressure continues, an alternative view to the one above is that the initial interpretation of the pattern of evidence (decreasing primary system energy, etc.) is wrong and the entire pattern of evidence needs to be re-evaluated. But in this case, there is no single possible explanation that would account for the entire pattern of findings (e.g., temperature behaviors and the persistent pressure decrease). Whether CES re-evaluates earlier interpretations and, if so, how it settles on a set of explanations to account for the entire pattern of findings depends on the settings of a variety of PAFs and the further evolution of the incident.

4.5 Current Stage of Implementation

The target capabilities for the Cognitive Environment Simulation are formalized based on the architecture and processing mechanisms currently available in the EAGOL AI problem-solving software system. The basic architecture for CES has been set up and implemented via EAGOL to exhibit some part of the major target cognitive competencies that were specified. In many areas further work is needed to evolve CES mechanisms within the current architecture to deal with the full range of target cognitive competencies and to better capture current knowledge about human cognitive processing in complex, dynamic worlds. In the longer term, further development will occur as more is learned about human cognitive processing from exercising CES in the NPP context and from new empirical results.

The current CES base of knowledge about the NPP that operators may possess is very limited at this time. It addresses primarily the primary

system thermodynamic functions and can address significant portions of primary break incidents against a background of non-primary breaks such as cooldown incidents (i.e., secondary breaks).

The mechanisms for interacting with CES (setting up plant input, modifying Performance Adjustment Factors) are currently very limited, as are the mechanisms for watching and recording CES behavior when it is stimulated by dynamic sequence of plant data. As a result, at this stage of development CES can be effectively used only by people who have behavioral science expertise, particularly in cognitive processes and human error, and intimate knowledge of the AI computer structures used to implement CES (i.e., the EAGOL software system). Mechanisms for interacting with CES can be expanded and enhanced to improve productivity and accessibility.

CES currently runs on Symbolics-3600 class computers and is planned to run on Sun-3 class computer as well. It is built on top of the EAGOL AI software "shell" developed by H. Pople of Seer Systems which uses Commonlisp and the Flavors object oriented programming environment. The basic software engine and knowledge base is transportable to several machines that can support the Commonlisp/Flavors software environment such as Symbolics, Sun, or Vax computers (however, access to the model software is done through software tools that are generally machine specific (e.g., window packages) so that some, one-time customization will probably be required to set up access to the model software on machines other than Sun or Symbolics).

Computer files of plant data in particular incidents are needed to stimulate CES. These data files consist of a series of "snap-shots," at some sampling rate, of the set of alarms and sensor values that would be *potentially* available for operational personnel to examine as the incident unfolds. CES is being set up for testing purposes to receive some of the plant data from a plant simulation model used for engineering studies and operator training on Westinghouse type pressurized water reactors. The current CES knowledge base is set up with knowledge about some parts of this type of reactor. It is possible to stimulate CES with data from other plants if it is encoded in a format which can be read by the model software. Note that setting up CES to run incidents on another reactor type assumes that the knowledge base has been adjusted to reflect plant specific changes in setpoints and equipment as well as differences in control board, procedures, etc.

The plant data for input to CES can come from data files or from a direct hookup to some type of plant simulation model. A dynamic plant model is preferred because CES is a dynamic model of the control of the NPP. Potential sources of the data on plant behavior include actual reactor

behavior, test reactor behavior, training simulation models, thermodynamic codes. The CES user needs to decide what input source is appropriate for a particular application of CES. Input sources to CES will vary in their ability to validly capture plant behavior for the kinds of incidents of interest and in the level of effort required to be interfaced to CES.

5. Conclusions and Recommendations

5.1 Benefits of the CES Modeling Environment

CES is the first cognitive process simulation tool that allows exploration of plausible human responses in different emergency situations. By simulating the cognitive processes that determine situation assessment and intention formation, it provides the capability to establish analytically what actions an operator is likely to take under different accident conditions. This means one can investigate the ability of humans to *recover* from equipment failures, execution errors or intention failures to stop or mitigate their consequences. Similarly, one can investigate *errors of commission* due to misperception of plant state or other forms of cognitive error.

The ability of CES to predict errors of commission is particularly important since misapprehension of plant state by the operator can result in multiple actions which can have broad systemic effects. Intention failures are a major source of *human related common mode failures* – multiple failures that are attributable to a common element (namely, the erroneous intention). For example, cases where the situation is misperceived, and the operator deliberately decides it is appropriate to turn off multiple, otherwise redundant and diverse systems as occurred at Three Mile Island and Chernobyl. The PRA community generally recognizes the importance of identifying common mode failure events because they can have large and widespread effects on risk. CES represents the first cognitive process model able to predict the wide spread consequences that can follow from an intention failure.

There are other benefits that derive from the modeling capabilities of CES. One can investigate the sources of cognitive processing breakdowns and intention failures. Because CES encompasses the factors that affect the available problem solving resources such as the specific form and content of displays, training, and procedures, it provides an analytic tool for investigating the effects of changes in NPP person-machine systems including new instrumentation, computer-based displays, operator decision aids, procedure changes, training, multi-person or multi-facility (e.g., technical support center) problem solving styles. This means that proposed changes/enhancements to NPP person-machine systems can be analytically checked before they have been implemented. The cognitive model can be used to filter which changes are sufficiently likely to improve performance that prototype construction and empirical tests are justified. As such, it should provide a cost-effective complement to difficult and expensive high-fidelity empirical evaluations.

5.2 Recommendations

The next steps which are needed to take advantage of the capabilities of the CES cognitive model:

- empirically validate the correspondence between CES and human behavior,
- evolve the model's capabilities and its accessibility to the potential user community.

The usefulness of the CES cognitive model depends on the ability of CES to behave like people do, for the same situation and with the same external and internal resources. In other words, the key question to be answered is the validity of CES as a modeling tool for human intention formation. An initial empirical evaluation and validation study is planned for Phase III of the research project.

Analogous to the situation with analytical computer codes which model reactor behavior, there needs to be an ongoing cycle of model evolution and change as our state of knowledge changes. The Cognitive Environment Simulation is the repository of the current state of knowledge on operator cognitive activities and is the best source for interpolating or extrapolating what human behaviors are likely in cases where there is no or limited experience -- including evaluating changes to the human-machine system and hypothetical situations that arise in postulated incidents for which there is no or insufficient empirical data. To fulfill this function CES needs to evolve as new empirical data are gathered and as our understanding of human error evolves.

The current implementation of CES does not exhibit all of the target cognitive competencies specified for CES, and it addresses only a small portion of the ideal scope of NPP tasks. The full range of cognitive competencies needs to be incorporated into CES and the NPP scope covered by CES needs to be broadened.

The mechanisms for interacting with CES (setting up plant input, modifying model performance adjustment factors) are currently very limited, as are the mechanisms for watching and recording CES behavior when it is stimulated by dynamic sequence of plant data. As a result, CES can be effectively used only by people who have behavioral science expertise, particularly in cognitive processes and human error, and intimate knowledge of the AI computer structures used to implement CES (i.e., the EAGOL software system). Mechanisms for interacting with CES should be expanded and enhanced to improve productivity and accessibility.

5.3 Conclusion

As a result of the model development work in Phase II of this research project, there exists, for the first time, a simulation model of the cognitive processes that affect operator intention formation in NPP emergencies. Reactor thermodynamic models are essential tools for design and risk assessment of the physical NPP. Similarly, the CES cognitive model will be an essential tool to assess human performance for the evaluation of human-machine systems in the NPP and, via the CREATE methodology, for assessment of the human contribution to risk.

Enough knowledge about operator cognitive activities in emergency situations and enough knowledge about parts of the NPP have been incorporated for CES to begin to be a useful tool to explore what would people do in NPP situations of interest and to identify situations prone to intention failures. The process of using CES will then provide useful information on human performance and reliability at the same time that CES undergoes further evolution, extensions and refinement.

6. References

Baron, S. A control theoretic approach to modeling human supervisory control of dynamic systems. In W. B. Rouse (Ed.), *Advances in Man-Machine Research, Volume 1.* JAI Press, 1984.

Bayless, P. D. & Divine, J. M. *Experiment Data Report for LOFT Large Break Loss-of-Coolant Experiment L2-5.* Springfield, VA: National Technical Information Service, 1982. (NUREG/CR-2826).

Bransford, J., Sherwood, R., Vye, N. & Rieser, J. Teaching and problem solving: Research foundations. *American Psychologist*, 1986, *41*, 1078-1089.

Bray, M. A. *The Conduct of Loss-of-Fluid (LOFT) Experiment L2-5: A Lesson for Accident Management in Nuclear Power Plants.* Springfield, VA: National Technical Information Service, 1982. (NUREG/CR-xxxx).

Brown, J.S., & VanLehn, K. Repair theory: A generative theory of bugs in procedural skills. *Cognitive Science*, 1980, *4*, 379-426.

Brown, W. & Wyrick, R. (eds.). *Analysis of Steam Generator Tube Rupture Events at Oconee and Ginna.* Institute of Nuclear Power Operations, 1982. (82-030).

Cohen, P. et al. Management of uncertainty in medicine. In (Ed.), *IEEE Conference on Computers and Communications.* IEEE, 1987. (Also University of Massachusetts COINS Technical Report 86-12).

Dorner, D. Heuristics and cognition in complex systems. In R. Groner, M. Groner & W. F. Bischof (Eds.), *Methods of Heuristics.* Erlbaum, 1983.

Gallagher, J. *Disturbance Analysis and Surveillance System Scoping and Feasibility Study.* Palo Alto, CA: Electric Power Research Institute, 1982. (NP-2240).

Hollnagel, E., Mancini, G. & Woods, D. D. (Eds.). *Intelligent Decision Support in Process Environments.* New York: Springer-Verlag, 1986.

Joksimovich, V. A Review of Plant Specific PRAs. *Risk Analysis*, 1984, *4*(4), 255-266.

Lane, D. M. Limited capacity, attention allocation and productivity. In W. C. Howell & E. A. Fleishman (Eds.), *Human Performance and Productivity: Information Processing and Decision Making.* Hillsdale, NJ: Erlbaum, 1982.

Levine, S. & Rasmussen, N. C. Nuclear Plant PRA: How far has it come? *Risk Analysis*, 1984, *4*(4), 247-254.

Mancini, G. Modelling Human and Machines. In Hollnagel, E., Mancini, G. & Woods, D. D. (Eds.), *Intelligent Decision Support.* Germany: Springer-Verlag, 1986.

Michalski, R. S. & Winston, P. H. Variable precision logic. *Artificial Intelligence*, 1986, *29*, 121-146.

Montmollin, M. de & De Keyser, V. Expert Logic vs Operator Logic. In G. Johannsen, G. Mancini & L. Martensson (Eds.), *Analysis, Design, and Evaluation of Man-Machine Systems.* CEC-JRC Ispra, Italy: IFAC, 1985.

Moray, N., Lootsteen, P. & Pajak, J. Acquisition of process control skills. *IEEE Transactions on Systems, Man, and Cybernetics*, 1986, *SMC-16*, 497-504.

National Transportation Safety Board. *Eastern Airlines L-1011, Miami, Florida, December 29, 1972.* Washington, D. C.: National Transportation Safety Board, 1973. (NTSB-AAR-73-14).

O'Brien, J. N., Luckas, W. J., Jr., & Spettell, C. M. *Team-Enhanced Evaluation Method (TEEM) Procedures: An Enhanced Human Reliability Analysis Process* (Informal Report BNL-38585). Brookhaven National Laboratory, December 1986.

Perkins, D. & Martin, F. Fragile Knowledge and Neglected Strategies in Novice Programmers. In E. Soloway & S. Iyengar (Eds.), *Empirical Studies of Programmers.* Norwood, NJ: Ablex, 1986.

Pew, R. W. et al. *Cockpit Automation Technology* (Tech. Rep. 6133). BBN Laboratories Incorporated, 1986.

Pew, R. W. & Baron, S. Perspectives on Human Performance Modelling. In G. Johannsen & J. G. Rijnsdorp (Eds.), *Analysis, Design and Evaluation of Man-Machine Systems.* London: Pergamon Press, 1983.

Pew, R. W., Miller, D. C. & Feehrer, C. E. *Evaluation of Proposed Control Room Improvements Through Analysis of Critical Operator Decisions.* Palo Alto, CA: Electric Power Research Institute, 1981. (NP-1982).

Pople, H. Jr. Evolution of an Expert System: from Internist to Caduceus. In I. De Lotto and M. Stefanelli (Ed.), *Artificial Intelligence in Medicine.* Elsevier Science Publishers B. V. (North-Holland), 1985.

Pople, H. E., Jr. Heuristic Methods for Imposing Structure on Ill-Structured Problems: The Structuring of Medical Diagnostics. In P. Szolovits

(Ed.), *Artificial Intelligence in Medicine.* Boulder, CO: Westview Press, 1982.

Rasmussen, J. *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering.* New York: North-Holland, 1986.

Rasmussen, J., Duncan, K. & Leplat, J. *New Technology and Human Error.* Chichester: John Wiley & Sons, 1987.

Reason, J. & Mycielska, K. *Absent Minded? The Psychology of Mental Lapses and Everyday Errors.* Englewood Cliffs, NJ: Prentice-Hall, 1982.

Swain, A. D. & Guttmann, H. E. *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications.* Springfield, VA: National Technical Information Service, 1983. (NUREG/CR-1278).

Trager, E. A., Jr. *Case Study Report on Loss of Safety System Function Events* (Report AEOD/C504). Office for Analysis and Evaluation of Operational Data, U. S. Nuclear Regulatory Commission, December 1985.

U. S. Nuclear Regulatory Commission. *Loss of Main and Auxillary Feedwater at the Davis-Besse Plant on June 9, 1985.* Springfield, VA: National Technical Information Service, 1985. (NUREG-1154).

U. S. Nuclear Regulatory Commission. *Loss of Power and WAter Hammer at San Onofre, Unit 1 on November 21, 1985.* Springfield, VA: National Technical Information Service, 1985. (NUREG-1190).

U. S. Nuclear Regulatory Commission. *Loss of Integrated Control System Power and Overcooling Transient at Rancho Seco on December 26, 1985.* Springfield, VA: National Technical Information Service, 1985. (NUREG-1195).

U. S. Nuclear Regulatory Commission. *Reactor Risk Reference Document* (Tech. Rep. NUREG-1150). Office of Nuclear Regulatory Research, U. S. Nuclear Regulatory Commission, February 1987. (Draft for Comment).

Woods, D. D. Operator decision making behavior during the steam generator tube rupture at the Ginna nuclear power station. In W. Brown & R. Wyrick (Eds.), *Analysis of Steam Generator Tube Rupture Events at Oconee and Ginna.* Institute of Nuclear Power Operations, 1982. (Also Westinghouse Research and Development Center Report: 82-1C57-CONRM-R2).

Woods, D. D. Some results on operator performance in emergency events. In D. Whitfield (Ed.), *Ergonomic Problems in Process Operations.* Inst. Chem. Eng. Symp. Ser. 90, 1984.

Woods, D. D., Elm, W. C. & Easter, J. R. The Disturbance Board Concept for Intelligent Support of Fault Management Tasks. In *Proceedings of the International Topical Meeting on Advances in Human Factors in Nuclear Power.* American Nuclear Society/European Nuclear Society, 1986.

Woods, D. D. Coping with Complexity: The Psychology of Human Behavior in Complex Systems. In L. P. Goodstein, H. B. Andersen & S. E. Olsen (Eds.), *Mental Models, Tasks and Errors: A Collection of Essays to Celebrate Jens Rasmussen's 60th Birthday.* London: Taylor & Francis, in press.

Woods, D. D. & Hollnagel, E. Mapping Cognitive Demands in Complex Problem Solving Worlds. *International Journal of Man-Machine Studies*, 1987, *26*, 257-275. (Special Issue on Knowledge Acquisition for Knowledge Based Systems).

Woods, D. D. & Roth, E. Operator Performance in Simulated Process Control Emergencies. Unpublished study, 1982.

Woods, D. D. & Roth, E. *The Role of Cognitive Modeling in Nuclear Power Plant Personnel Activities: A Feasibility Study.* Washington D. C.: U. S. Nuclear Regulatory Commission, in preparation. (NUREG-CR-4532).

Woods, D. D., Wise, J. A. & Hanes, L. F. *Evaluation of Safety Parameter Display Concepts.* Palo Alto, CA: Electric Power Research Institute, 1982. (NP-2239).

Worledge, D. H., Chu, B. B., & Wall, I. B. Nuclear Plant Systems Analysis Research at EPRI. *Risk Analysis*, 1884, *4*(4), 299-311.

NRC FORM 335 (2-84) NRCM 1102, 3201, 3202

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

SEE INSTRUCTIONS ON THE REVERSE.

1. REPORT NUMBER (Assigned by TIDC, add Vol. No., if any)
NUREG-CR-4862
Volume 2

2. TITLE AND SUBTITLE
Cognitive Environment Simulation: An Artificial Intelligence System for Human Performance Assessment. Volume 2: Modeling Human Intention Formulation

3. LEAVE BLANK

4. DATE REPORT COMPLETED
MONTH: September YEAR: 1987

5. AUTHOR(S)
D. D. Woods, E. M. Roth, H. Pople, Jr.

6. DATE REPORT ISSUED
MONTH: November YEAR: 1987

7. PERFORMING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)
Westinghouse Research and Development Center
1310 Beulah Road
Pittsburgh, PA 15235

8. PROJECT/TASK/WORK UNIT NUMBER

9. FIN OR GRANT NUMBER
FIN D1167

10. SPONSORING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code)
Division of Reactor and Plant Systems
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555

11a. TYPE OF REPORT
Technical

b. PERIOD COVERED (Inclusive dates)
May 1986 through June 1987

12. SUPPLEMENTARY NOTES

13. ABSTRACT (200 words or less)
This report documents the results of Phase II of a three-phase research program to develop and validate improved methods to model the cognitive behavior of nuclear power plant (NPP) personnel. In Phase II a dynamic simulation capability for modeling how people form intentions to act in NPP emergency situations was developed based on techniques from artificial intelligence. This modeling tool, Cognitive Environment Simulation (CES), simulates the cognitive processes that determine situation assessment and intention formation. It can be used to investigate analytically what situations and factors lead to intention failures, what actions follow from intention failures (e.g., errors of omission, errors of commission, common mode errors), the ability to recover from errors or additional machine failures, and the effects of changes in the NPP person-machine system.

The Cognitive Reliability Assessment Technique (CREATE) was also developed in Phase II to specify how CES can be used to enhance the measurement of the human contribution to risk in probabilistic risk assessment (PRA) studies.

The results are reported in three self-contained volumes that describe the research from different perspectives. Volume 1 provides an overview of both CES and CREATE. Volume 2 gives a detailed description of the structure and content of the CES modeling environment and is intended for those who want to know how CES models successful and erroneous intention formation. Volume 3 describes the CREATE methodology for using CES to provide enhanced human reliability estimates. Volume 3 is intended for those who are interested in how the modeling capabilities of CES can be utilized in human reliability assessment and PRA.

14. DOCUMENT ANALYSIS - a. KEYWORDS/DESCRIPTORS
Artificial Intelligence
Human Reliability
Probabilistic Risk Assessment
Cognitive Model
Nuclear Power Plant
Human Error
Problem Solving
Human Factors
Cognitive

b. IDENTIFIERS/OPEN-ENDED TERMS
CES
Cognitive Environment Simulation
Cognitive Reliability Analysis Technique
CREATE

15. AVAILABILITY STATEMENT
Unlimited

16. SECURITY CLASSIFICATION
(This page) Unclassified
(This report) Unclassified

17. NUMBER OF PAGES

18. PRICE

*U.S. GOVERNMENT PRINTING OFFICE: 1987-202-292:60296

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, D.C. 20555