Cognitive
Engineering
Laboratory

Applying a formative ecological framework to simulator design challenges

Antony Hilliard

CEL 07-04

Cognitive Engineering Laboratory, Department of Mechanical & Industrial Engineering,
University of Toronto
5 King's College Road, Toronto, Ontario, Canada M5S 3G8
URL: www.mie.utoronto.ca/labs/cel/

Cognitive Engineering Laboratory

Directors: Kim J. Vicente, Ph.D., P. Eng.
Greg A. Jamieson, Ph D., P. Eng.

The Cognitive Engineering Laboratory (CEL) at the University of Toronto (U of T) is located in the Department of Mechanical & Industrial Engineering, and is one of three laboratories that comprise the Human Factors Research Group. CEL was founded in 1992 and is primarily concerned with conducting basic and applied research on how to introduce information technology into complex work environments.

Current CEL Research Topics

CEL has been funded by Atomic Energy Control Board of Canada, AECL Research, Alias|Wavefront, Asea Brown Boveri Corporate Research - Heidelberg, Canadian Foundation for Innovation, Defence Research & Development Canada (formerly Defense and Civil Institute for Environmental Medicine), Honeywell Technology Center, IBM, Japan Atomic Energy Research Institute, Microsoft Corporation, Natural Sciences and Engineering Research Council of Canada, Nortel Networks, Nova Chemicals, Westinghouse Science & Technology Center, and Wright-Patterson Air Force Base. CEL also has collaborations and close contacts with the Mitsubishi Heavy Industries and Toshiba Nuclear Energy Laboratory. Recent CEL projects include:

- Developing advanced human-computer interfaces for the petrochemical and nuclear industries to enhance plant safety and productivity.
- Understanding control strategy differences between people of various levels of expertise within the context of process control systems.
- Developing safer and more efficient interfaces for computer-based medical devices.
- Creating novel measures of human performance and adaptation that can be used in experimentation with interactive, real-time, dynamic systems.
- Investigating human-machine system coordination from a dynamical systems perspective.

CEL Technical Reports

For more information about CEL, CEL technical reports, or graduate school at the University of Toronto, please contact Dr. Kim J. Vicente or Dr. Greg A. Jamieson at the address printed on the front of this technical report.

Applying a formative ecological framework to simulator design challenges. Antony Hilliard, Masters of Applied Science, 2007, Department of Mechanical and Industrial Engineering, University of Toronto

ABSTRACT

Human-in-the-loop interactive simulators are widely recognized as a cost-effective and safe way to train workers in challenging work situations, and to more tractably research human performance. Designing cheaper, lower fidelity and more tractable computer-based simulators has been proposed as a way to widen simulator adoption and extend training and human performance research benefits. There is currently no systematic method to design simplified simulations while retaining psychological fidelity. Design decisions are typically based on common tasks and simulator designers' intuition and expertise. This is difficult to replicate, complicates comparisons between simulators, and presents problems for scientific defensibility of research. This thesis demonstrates how Cognitive Work Analysis (CWA)'s phases can be applied to support designers both in systematic, defensible design of simplified simulators, and also in ongoing scenario design and development. A subset of CWA design methods are demonstrated in a petrochemical domain case study.

Table of Contents

Table of Figures

Chapter 1 Overview

The Cognitive Engineering Laboratory (CEL) has recently acquired new research infrastructure, the Seven-node Human-Automation Interaction SimulaTOr System (HEPHAISTOS). The technical capabilities of the HEPHAISTOS infrastructure support more realistic industrial process control environments than previously available to CEL researchers. Complimenting the increased difficulty level of such environments, Lambton College is collaborating with the CEL to provide access to skilled process control technology students as research participants.

In order to support cognitive engineering research on human-automation interaction, interface design, and team organization and cooperation, the HEPHAISTOS infrastructure must be matched with a software simulation model and interface that can support a suitable range of human behavior. While simulations of complex process control systems are used in industry, they have generally been developed to replicate specific plant instrumentation and used to train operator procedures. Whether such simulations can support valid cognitive engineering research is uncertain. To scientifically develop or evaluate human performance research simulators, design choices must be both documented and theoretically justified. This thesis proposes an adaptation of an existing theoretical framework to the principled, systematic evaluation and design of research simulators.

First, an overview of simulator applications will be presented, and some limits to the scope of this paper explained. Next, existing simulator terminology, taxonomies, and design frameworks will be summarized. Then, several case studies will be used to illustrate the application of simulator design principles, and highlight examples that motivate further work.

Finally, a collection of proposed simulator design methods will be presented, based on the Cognitive Work Analysis (CWA) framework. A subset of these methods will be applied to a candidate HEPHAISTOS simulation, and design implications described. Finally, an evaluation of the method’s usefulness and opportunities for future work will be outlined.

Note on Vocabulary

Wherever possible, this paper will use vocabulary proposed by Vicente (1999). Terms specific to simulators will be adapted from the Simulation Interoperability Standards Organization's working glossary (FISG, 1998). Some key terms are reproduced below.

> **simulation**. 1. A method, software framework or system for implementing one or more models in the proper order to determine how key properties of the original may change over time. See model, representation. 2. An unobtrusive scientific method of inquiry involving experiments with a model rather than with the portion of reality this model represents. (FISG, 1998)
>
> **simulator**. A device or physical system that implements or performs simulation. See simulation, simuland, software model, mathematical model. (FISG, 1998)
>
> **actor**. A worker or automation (Vicente, 1999, p. 4)
>
> **automation**. A mechanical, electrical, chemical, or computerized actor. Automation acts on other automation or on the work domain (Vicente, 1999, p. 5)
>
> **workers**. The people who participate in and act in a sociotechnical system. Synonymous with operators and users. (Vicente, 1999, p. 10)
>
> **participant**. A worker acting in a simulator.

Chapter 2 Introduction

The complex systems that support modern society involve the management of large costs and significant risk. Intervening in their operation to educate or train novice workers or to perform research is usually unacceptable. Computer-based simulations of such systems can provide a more cost effective and less risky approach to training and research.

2.1 Training applications of simulators

Traditionally, simulators have been considered better the more true to life they were. Examples include the full recreation of the Apollo space capsules used for astronaut training and development of full nuclear power plant control room and aircraft cockpit simulators (Kozlowski et al., 2004). When training programs used lower fidelity simulators, it was typically a compromise made to satisfy budget restrictions.

However, full scope high-fidelity simulators may not always be best. Their complexity means that developing the simulator, simulation and scenarios is costly and slow, which can be a serious liability (Elliott, Dalrymple, Schiflett, & Miller, 2004). This is especially apparent when using them to train for adaptability, rather than rote procedures (Gugerty, Schifflett, Elliott, Salas, & Coovert, 2004; Naikar & Sanderson, 1999). In military applications, the need for human adaptability in response to disturbances, what Vicente (1999) calls context-conditioned variability, is recognized as the primary role of workers in responding to enemy activity. Simulated disturbances must be developed by scenario designers, and if high-fidelity simulators' cost and complexity hinders scenario development, the effectiveness of training programs can be reduced, as illustrated by the following example (Bowers et al., 2004).

In a high-fidelity flight simulator used for training commercial airline pilots, an unexpected event was supposed to occur during each simulated flight. However, the simulator was so complex and expensive to program, and FAA approval so lengthy, that only a limited number of scenarios were available and had to be re-used many times. Through experience and by word of mouth, participants were often aware of what 'unexpected event' was going to happen. A simulated flight taking off from Chicago O'Hare would be the scenario involving a #2 engine failure, for example. Such limitations in simulated variety directly contradict the objectives of training for adaptability.

Examples such as these have led some to suggest that the design of training should be guided by a task or work analysis, so that simulator fidelity can be only as high as necessary for training requirements (Kozlowski et al., 2004; Naikar & Sanderson, 1999). Principled simulator design can help ensure that both development and ongoing operation costs are identified and accommodated within organizational budgets.

2.2 Research applications

The development of simulators for research is motivated by both the costs and risks involved in the use of real-world systems, and associated trade-offs in experimental effectiveness (Stanton, 1996). Experimenters often do not have access to or control over systems of research interest due to cost or risk. When such access is obtained, actually performing research in such systems presents many difficulties.

As work has expanded from manual labor to skilled machine tasks to supervising computerized control, worker processes have shifted from external action to internal cognition, challenging the methods available to human performance researchers. Early experiments investigating human control processes were limited by experimental equipment and most particularly by lack of powerful statistical analysis tools. Experiments had to be simple and easily controlled for defensible results to be obtained. Stimulus-Response type psychology experiments provided easily analyzable data and so provided the foundation for most generally accepted models of human motor and perceptual processes (Brehmer & Dörner, 1993). Studies of more complex and dynamic human work in-situ were often limited to qualitative and interpretive results such as those gained through observations of industrial control room activities (Zuboff, 1988).

With technological development the methods of psychological research have also progressed. However even when equipped with dataloggers, eye trackers, and advanced statistics, quantitative research on cognitive work in the field remains a challenging exercise. Using large-scale high fidelity simulators as an alternative to field research allows experiments to be performed without risking real-world disaster, and their completeness can be used to defend the validity of results according to the theory of identical elements (Druckman & Bjork 1991, qtd. in Kozlowski et al., 2004). On the other hand, high-fidelity simulators can be similarly difficult for researchers to use: they still consume large amounts of funding, they introduce the challenge of finding participants capable of operating them, and their complexity means that experimental degrees of freedom remain high and confounding (Brehmer & Dörner, 1993). Identical resemblance is not always helpful for research: as Brehmer put it in his oft-quoted "Cat Problem", the most realistic simulation of the cat is another equally complex cat – which is just as difficult to understand (Brehmer, Schifflett, Elliott, Salas, & Coovert, 2004).

In the mid-1980s, the processing power of microcomputers reached a level that enabled the economical implementation of small, simplified simulators. Microworlds (Brehmer & Dörner, 1993) were envisioned as a means to bridge the gap between controlled laboratory studies and expensive, experimentally un-tractable high-fidelity simulators. This approach suggested that the design of simulators for theoretical research should involve 1) simplification to allow experimental effects to be measured by reducing degrees of freedom, and 2) introduction of scenario development, data collection, and other tools to aid researchers. This approach introduces a new challenge: how to design microworlds to be simple, useful, and scientifically defensible. The variety of approaches to this challenge, and the terminology, theories, and methods used will be discussed in Chapter 3.

2.3 Scope limitations

Simulations can be developed for a wide range of natural systems, and one design process may not be suited to every case. Thus, we will narrow the focus of this paper through three qualifications.

2.3.1 Complex sociotechnical systems

This paper will discuss only human in-the-loop simulators that are developed from, or intended to generalize to, complex sociotechnical systems. Human-in-the-loop simulation indicates that participants are involved in the simulation's operation (Sheridan, 2006). Complex sociotechnical systems are often described in terms of features such as distributed social interactions, challenging dynamics, hazards, tight coupling between elements, extensive automation, uncertainty, and unanticipated disturbances (Vicente, 1999).

A useful distinction can be made between *causal* systems primarily governed by natural laws, such as a process control plant, and *intentional* systems that are intended to follow human intentions, values, and priorities, such as an information management or financial system (Rasmussen, Pejtersen, & Goodstein, 1994). Characteristics of both extremes are found in many systems such as military or medical domains. For example, workers in a hospital must consider both the causal processes of the human body and the intentional values of care giving. While systems with both causal and intentional components will be considered in this paper, the focus will be on systems that tend to have a majority of causal constraints. This is consistent with the HEPHAISTOS research focus.

2.3.2 Correspondence-driven

A key characteristic of complex sociotechnical systems is that they are correspondence, as opposed to coherence-driven (Vicente, 1999). Coherence driven systems, such as a word processor, involve workers interacting with a system in a worker-computer dyad, as pictured in Figure 1a. Such systems can be designed to fit workers' preferences and cognitive constraints without needing to consider an underlying physical or social context. On the other hand, correspondence-driven systems such as industrial process control require workers to manage a physical or social reality with the support of some mediating interface which is often computerized, as shown in Figure 1b.

a)

b)

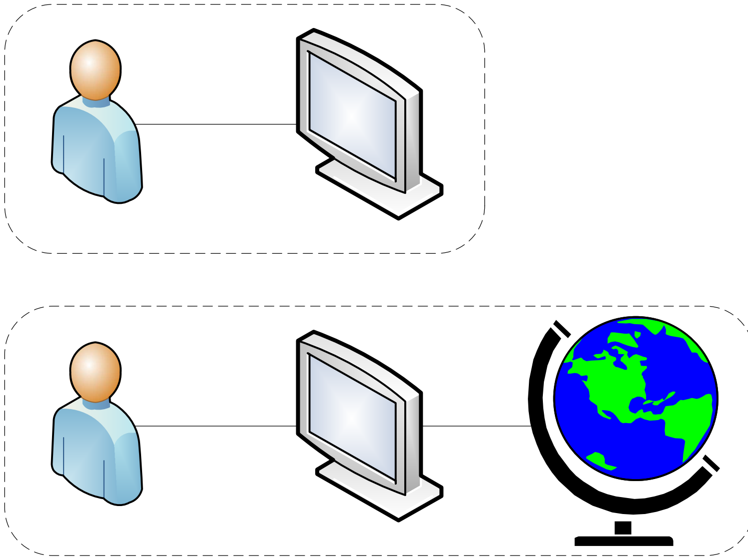


Figure 1 – a) Coherence and b) Correspondence -driven Work Domains (adapted from Vicente(1990))

In coherence driven work systems, simulation and prototyping are very similar, and realism is often only a question of cost and representativeness of participants. Since there are no environmental constraints, the system, a prototype, or a mockup can be used for training, research, or development, without safety risks. An example is the 'beta' and usability testing of graphics processing software or mobile phone interfaces. On the other hand, correspondence driven systems have challenges associated with simulating a physical or social reality, as discussed in sections 2.1 and 2.2, and if the work system is infrequently occurring, such as a hospital disease outbreak, sometimes simulation is the only option.

This paper will address simulations of correspondence-driven systems that continuously develop as a function of participant actions, autonomously by internal processes (Brehmer, Leplat, & Rasmussen, 1991), and in response to disturbances in their environment (Vicente, 1999). Simulation types that will not be discussed include those that either develop only as a result of participant actions, such as puzzle games, or those that develop only as a result of their internal processes, such as an operations research industrial process simulation.

2.3.3 Direct or Computer-mediated

Since correspondence-driven complex sociotechnical systems have properties that cannot be sensed or manipulated directly by unaided humans, some form of interface is typically present between humans and the system, such as mechanical controls, instrumentation, or communication tools. In the case of *computer mediation*, display or control of system components is performed via a computer interface (Bennett, Nagy, & Flach, 2006). Identifying systems as being computer-mediated is of course not always clear-cut. Many systems contain mediated and unmediated elements, even in something as simple as urban commuting: radio broadcasts provide mediated information concerning citywide traffic, and navigation systems can provide an even wider range of computer-mediated data.

Table 1 - Systems and simulations, computer-mediated and not

Natural (real-world) System	Simulated System: **computer-mediated**	Simulated System: **not**
not	Vehicle driving Firearms use City government VFR aircraft flight	Scale model vehicles Military field exercise Operating room surgery Emergency disaster response
comp. mediated	IFR aircraft flight Air traffic control Process control room Financial stock trading	'Handmade' process control

Simulations can be designed with similar or different levels of mediation than their target systems, as shown in Table 1. The four quadrants of this figure present combinations of this distinction for real and simulated systems. In the top right quadrant are systems that are largely directly experienced by workers, and can thus be simulated in a live action format. Props and simulation equipment can be used to safely represent potentially dangerous elements: e.g. blank ammunition and casualty arbitrators for military exercises, and patient dummy dolls for medical procedures (Manser, Dieckmann, Wehner, & Rall, 2007). In the bottom right quadrant are computer-mediated systems that are simulated as direct participation exercises. Exemplars of this category are difficult to find; one hypothetical example could be a centrally controlled industrial bakery that trains its control room operators in home-style baking in order to familiarize them with the physical appearance, tastes and smells of the process. In the top left-hand quadrant are simulations that are implemented with computer mediation, even though the systems they represent are for the most part directly experienced. Examples include vehicle driving simulation (Lee, 2001), computerized military firing ranges (Gula, 1998), simulations of political leadership (Dörner et al. 1983, qtd. in Brehmer et al., 1991), and flight simulators for small aircraft using visual flight rules. Finally, examples from the bottom left-hand quadrant of Table 1 are largely computer-mediated, both in the real world and the simulator. Industrial process control rooms in the petrochemical industry, for example, can sometimes be located kilometers away from the actual equipment, leaving operators completely dependent on computerized control systems and radio communication with site workers.

Computer-mediated systems are both easier and more difficult to simulate. Simulation of computer-mediated systems can seem straightforward, because the information sources and controls that workers use have already been designed and can be duplicated with sufficient budget and ingenuity (Skraaning jr., 2003). Likewise, non-computer-mediated systems have a rich perceptual environment that is impossible to duplicate. Identifying what elements of the real world guide skilled experts and how to include them in a simulation is no easy task (Kirlik, Miller, & Jagacinski, 1993) and while work in virtual reality has approximated many aspects, it is still far from complete (Goldstein, 2006). However, computer mediation can make simulator design more difficult as well: because work in computer-mediated systems is cognitive, especially in supervisory control environments (Sheridan, 2006; Vicente, 1999), overt behavior may not reveal the effects of inevitable differences between a natural system and its simulation.

There are many pragmatic reasons to design simulators to be computer-mediated: cost, reconfigurability, and data collection, for example. Unmediated simulators for systems complex enough to be of interest to cognitive engineering can require large teams of participants and extensive equipment. Thus, this paper will focus on the design of computer-mediated simulations, though unmediated simulations have many overlapping design challenges and opportunities.

2.4 Conclusions

Narrowing this paper's focus to computer-mediated simulations of complex sociotechnical systems still encompasses a wide range of simulator applications. In the next chapter, terminology common to microworlds and high-fidelity simulators will be introduced, and some of the wide body of previous work in simulator design and evaluation will be discussed.

Chapter 3 Theory and Current Practice

To place further discussion in context, this chapter will discuss philosophical perspectives on simulation, practical and theoretically-grounded simulator terminology, and review existing approaches to simulator design.

3.1 Philosophy/Psychology of Interactive Simulation

It is common to discuss the ways in which simulations differ from or resemble reality. This takes for granted an agreed definition of reality, and what portions of it are to be compared, something philosophers have yet to agree upon (Hammond, 1998). While we will not directly address this question, for the scope of this work we will substitute the slightly less vague term *natural system*, referring to the complex sociotechnical system(s) being simulated (Vicente, 1999).

3.1.1 Natural Differences

Designing a simulation resembles scientific model-building, which has been discussed at length in the philosophy of science literature, from which we will introduce a few distinctions.

The first distinction is between two classical paradigms: the ideographic tradition of analyzing and interpreting individual events in search of cause and effect, and the nomothetic tradition of seeking laws and regularities across a range of occurrences (Rasmussen, Pejtersen, & Schmidt, 1990). The ideographic approach to simulation can be thought of in terms of historical re-enactments. An ideographic simulation of the Apollo 13 disaster would be developed to follow the exact timeline of the incident and explain in-depth the specific events and decisions related to the event. A nomothetic approach would instead focus on the common factors, both physical and intentional, that can help explain system behavior. For an Apollo spaceflight, a nomothetic simulation would recreate the physical and social environment common to all Apollo missions and evaluate their influence on Apollo 13's fate. Since modern physical and psychological science follows the nomothetic tradition, such simulations will be the focus of this paper.

Within nomothetic approaches, scientific models can be either *descriptive* or *explanatory* (Humphreys, 2004). A descriptive model of phenomena is accurate, but its mechanisms do not provide insight into why or how the phenomena work, or allow predictions to other conditions. Modern approaches such a neural networks, or ancient approaches such as Kepler's epicyclic model of planetary motion exemplify this approach (Humphreys, 2004). An explanatory model, on the other hand, captures underlying regularities that can be manipulated to describe situations beyond those the model was developed for. A planetary motion model using Newton's model of gravitational attraction, for example, could simulate the effect on the solar system of the Earth suddenly splitting in two. While simulations based on either approach can provide indistinguishable end results, those built around explanatory models are much more useful for research as they can be extended beyond their original scope.

To construct explanatory models or simulations of natural systems, designers both abstract and idealize. Abstraction "omits some feature of the modeled system without representing the system as lacking that feature" (Jones, 2005, p. 184), such as by making no mention of the color of a driver's simulated vehicle. Idealizing, on the other hand, involves distortion or approximation of a natural system's properties, such as by neglecting wind effects from a driving simulation. In simulation terms, an abstraction would be making some properties that describe a natural system irrelevant in a simulator, while idealization would include those properties but either change their dependence or hold them constant. Such constant variables do not disappear (Brehmer et al., 1991) and problems can arise from hidden interactions involving such constants (Chapanis, 1967).

Researchers from the ecological psychology community have expressed caution regarding simulations (Goldstein, 2006). Ecological psychology holds that humans' behavior is strongly influenced by their natural environment. Brunswik, a founder of the ecological approach, argued that the extensive idealization of classical stimulus-response experimentation "destroys the very phenomenon under investigation, or at least, it alters the processes in such a way that the obtained results are no longer representative of people's actual functioning in their ecology" (Dhami, Hertwig, & Hoffrage, 2004, p. 962). Abstractions and idealizations resulting from computer simulation are thus a source of concern. While some ecological psychologists agree that simulators can be made to be useful and adequately representative (Hammond, 1998), they insist that design should be guided by theory that supports "hopefully judicious choice over which features of the world are both relevant and irrelevant psychologically to the behavior of interest" (Kirlik et al., 1993, p. 933). This argument would suggest that for human-in-the-loop simulators, neglecting a feature of the natural system would be considered an abstraction if it has both little effect on the natural system itself and also little effect on participants' behavior within that natural system. For example, the noises made by an automobile are irrelevant to simulating the physics of its road handling capabilities. Thus, a simulation that neglected automobile noises would be an abstraction if used to research the behavior of traffic police in spotter aircraft, but an idealization if used to research traffic police stationed at the side of the road.

3.1.2 Human Influences

Inevitable differences between simulations and natural systems are reflected in differences in human behavior. Even in a simulator that replicates the appearance and behavior of a system precisely, emotional experiences are difficult to induce in participants (Skraaning jr., 2003). Examples include the occasional moments of sheer terror typical of hazardous industrial systems or air traffic control domains (Chapanis, 1967). This has been noted in medical domain simulations, where "the under-performance of trainees during simulated scenarios has been ascribed to a lack of stress associated with the real event ('adrenaline gap')" (Maran & Glavin, 2003, p. 26). Social interactions are also distorted: participants perceive their role differently with respect to experimenters than they would with colleagues or supervisors (Chapanis, 1967). Participants readily perceive these differences: when asked to rate the realism of operating room simulations, anesthetists "rated the scenarios themselves as internally realistic but raised concerns regarding the short duration of scenarios, the heightened alertness of trainees knowing they were in a simulation and a failure to convincingly recreate the operating room culture." (Manser et al., 2007, p. 248).

Likewise, psychological constructs determined from laboratory studies may not carry over to more complex settings. Vigilance theory, for example, does not necessarily apply in natural (Chapanis, 1967) or simulated (Moray & Haudegond, 1998) systems, and similar concerns have been expressed about the base rate fallacy (Hammond, 1998), overconfidence bias, and hindsight bias (Dhami et al., 2004). This suggests that using psychological constructs to guide simulator design may be risky if they have not yet been shown to apply in the natural system of interest.

Participants can adapt to simulations' scope reduction, abstraction, and idealization by reducing their problem solving space when in simulator. One example was documented when studying fault diagnosis in a nuclear power plant (NPP) (Beare et al. 1984, qtd. in Moray, 2004). Researchers made probability plots of participant response time on logarithmic scales and found them roughly linear, indicating exponential distribution of problem-solving time. Participant response times in a high-fidelity simulator had a similar distribution, however, they were ten times quicker! The reason proposed for this discrepancy is that operators are taking into account abstractions and idealizations of a simulated environment and limiting their problem-solving scope, confident that the simulator will not incorporate many of the unlikely potential failures in an operational NPP (Moray, 2004).

As Manser (2007) observed, there is no theoretical consensus as to what simulators need to contain to be effective natural system surrogates for research, much less methodological solutions to design such simulators.

3.2 Classifications and characteristics of simulators

As starting points in reviewing simulator design theory and practice, we will discuss some practical terminology for distinguishing classes of simulators and comparing simulators with one another.

3.2.1 Gray's Classes of Simulators

Gray (2002) has introduced a modest taxonomy of some general types of simulators. The taxonomy is not exclusive or exhaustive, and is intended to support a common language for discussion of simulation types. It provides five useful distinctions to compare simulators by their design goals and characteristics:

1. *Laboratory Tasks.* This category encompasses tools of more 'traditional' psychological research such as radar signal-detection equipment. Gray is clear that what appears simple to a cognitive researcher may be highly complex to one studying human visual attention. The term 'simple' is not intended to have universal applicability.

2. *Simulated Environments / Microworlds*. Use of these types of simulators is often intended to produce broadly generalizable results for theory development. Most of the physical form of one or several natural systems is abstracted away, leaving functional relationships that can be understood by novice test subjects. These simulators are often built to continue research of phenomena studied in simpler settings. The DURESS II microworld is an exemplar of this category (Vicente, 1999).

3. *Scaled Worlds*. These simulations have been 'stripped down' to remove aspects of the real system that are not relevant to the research or training goals or are too expensive. Multiple Scaled Worlds can be generated for one system. The simulation still has strong correspondence with the real system, and often requires experienced users. Research is often focused on generalizability to the natural system rather than to general human performance. Some military command and control simulators fall into this category (Elliott et al., 2004).

4. *High Fidelity simulators of "Simple" worlds.* This category is intended to capture detailed simulations of smaller systems – either simpler whole systems or a portion of a complex system, such as the flight management computer of an airliner.

5. *Hi-Fidelity simulators of Complex Worlds*. The stereotypical exemplar of this category is a full nuclear power plant control room simulator. Physical, functional and engineering fidelity[1] is very high, as is cost. Still, social and motivational elements may not be as well reproduced.

We will use these terms with reference to Gray's taxonomy, and use 'simulator' as a generic term to encompass all categories.

3.2.2 Dimensions of Distinction

Gray suggests that the five categories of simulators can be compared most usefully across three dimensions, the first from the researcher's perspective, the second from the simulator's perspective, the third from the test subject's perspective (Gray, 2002). He notes that there are many other dimensions of difference, such as the obvious choice,

---

[1] Defined in Section 3.2.3.

complexity (Berutti, 2007), but argues that absolute complexity is less important to researchers than these three aspects.

1. Tractability: Gray defines this as "ease of working with", from a researcher's perspective. Simulators with high tractability are easily managed, controlled, and trained for, offer data collection of performance measures, not just outcomes, and allow interconnections with required research tools. Tractability, as with validity, is defined relative to the research question. What may be tractable for cognitive workload studies might be very untractable for a visual attention study.

2. Correspondence: Correspondence is a description of how similar or how close of an analogy the simulator is to the real world, incorporating both physical and psychological fidelity. Low correspondence is associated with very highly abstracted microworlds or laboratory tasks that may not even occur in real life. High correspondence is associated with very detailed high fidelity simulators. Correspondence generally seems to be inversely related to generalizability. Gray notes that a flight simulator for an F-16 pilot would never be used to train a Boeing 777 crew, while a microworld that elicits common cognitive processes can be used to investigate many aspects of industrial control.

3. Engagement: This dimension is associated with participants, being the degree to which the simulation encourages them to expend effort in working with the simulator. For highly abstracted microworlds or scaled worlds, the loss of sources of motivation such as professional or social obligations can be partly compensated by making the simulation inherently engaging (i.e. 'game-like'). Gray encourages caution when attempting to modify engagement, since the level of engagement of the user may be a factor inherent to the system under investigation.

These three criteria can be used to compare Gray's five categories of simulators on continuous scales, shown in Figure 2. Continuous scales are useful for illustration but should not be taken as a means to a 'weighted decision matrix' style of simulation selection.

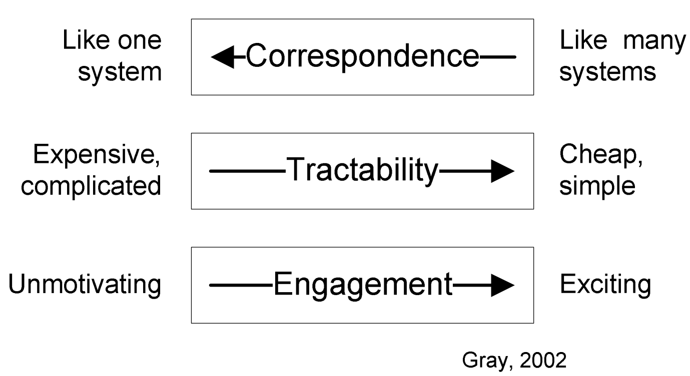


Figure 2 - Correspondence, Tractability, and Engagement (adapted from (Gray, 2002))

Tractability is a ‘more-is-better’ property, to be achieved through good design practice and possibly at the expense of generalizability. However the correspondence property is a continuum between “one aspect of many systems” and “many aspects of one system”, either of which can be desirable depending on the research question. Gray suggests that increasing the ‘fun factor’ of an inherently unexciting task may obscure any effects under study and thus the engagement dimension should not be any higher than that of the real-world system. Thus what constitutes ‘better’ is not evaluated with respect to the same baseline between these three dimensions.

Also, while tractability and engagement contain their *quality* within their dimensions (poorly implemented tractability is low tractability, poorly implemented engagement is either too boring or over-exciting), the quality of correspondence is not represented on the same scale as used to define the term. A system with high correspondence implemented poorly does not behave like a system with low correspondence.

With these caveats, these dimensions of distinction are useful for discussing simulator design and implementation.

3.2.3 Fidelity & Validity

The five categories of simulator discussed in Section 3.2.1 differ greatly in their level of *fidelity*, and their *validity* for research and training. The exact meaning of these terms have been extensively debated in the simulation community, and recognized as needing standardization. We will adopt the definitions produced by the Simulation Interoperability Standards Organization (SISO)’s 1999 Fidelity Implementation Study Group (F-ISG):

> Fidelity: The degree to which a model or simulation reproduces the state and behavior of a real world object or the perception of a real world object, feature, condition, or chosen standard in a measurable or perceivable manner; a measure of the realism of a model or simulation; faithfulness. (FISG, 1998)

The second definition, realism, is the one most commonly used, but the first contains a key distinction – between fidelity in reproducing a simulated real world object, and in reproducing the *perception* of a real world object. We will discuss this further below. Fidelity is considered to be a property of a simulator *and scenarios*, independent of research or training applications (Roza, Gross, & Harmon, 2000). The SISO F-ISG notes that while fidelity is intended to be an objective measure, it “should generally be described with respect to the measures, standards or perceptions used in assessing or stating it.” (FISG, 1998).

> Validity: The quality of being inferred, deduced or calculated correctly enough to suit a specific application,

> with particular application to a model or simulation's representational capability. The logical truth of a derivation or statement, based on a given set of propositions. (FISG, 1998)

Validity is a relationship between a simulator and an application, describing the results' degree of truth and extent of applicability. The two distinctions of validity hopefully present in all experiments are *internal validity*, meaning the experiment actually measures what it purports to; and *external validity*, that its results generalize beyond the human subjects, experimental apparatus, and scenarios used[2]. Two simulators of different fidelities may both have acceptable validity for the same research question or training objective, and vice-versa. If psychological constructs, such as situational awareness are incorporated in an experiment, *construct validity* must also be considered (Skraaning jr., 2003). Other distinctions within validity are content validity, that the simulator contains similar elements to domains where generalizability is desired, and face validity, that the simulator looks plausible to a layperson.

As validity depends on application, we will focus more on simulator fidelity and discuss considerations of validity on a case-by-case basis.

Fidelity Distinctions

A frequent criticism of scaled worlds and microworlds is that they are too "low fidelity". As argued elsewhere (N. J. Cooke, Schifflett, Elliott, Salas, & Coovert, 2004; Elliott et al., 2004; Kozlowski et al., 2004; Lee, 2001; Naikar & Sanderson, 1999) a single notion of "fidelity" is too broad to effectively describe the relation between natural and simulated systems. A distinction between at least two components of fidelity is made by many researchers:

[2] The term ecological validity has been used by some researchers to refer to the simulator's resemblance to real-world situations, overlapping with definitions of external validity (Skraaning jr., 2003) as well as face and content validity. This practice has been deplored by ecological psychologists (Hammond, 1998) as a muddying of the technical meaning of the term ecological validity as defined within the Brunswickian theoretical framework. Ecological validity within this framework does not describe experiments or simulators, but is a property of perceptible features, "the potential utility of various cues for organisms in their ecology (or natural habitat)" (Hammond, 1998). Considered in a simulator design context, the Brunswickian ecological validity can more readily be used as a means of discussing fidelity. In this work we will include 'ecological validity' in the scope of external validity.

Table 2 - Aspects of Fidelity

Paper	(Kozlowski et al., 2004)	(Elliott et al., 2004)	(Lee, 2001)	(Naikar & Sanderson, 1999)
Fidelities Identified	Physical	Physical (incl. performance)	Physical, Engineering	Physical
	Psychological	Functional, Cognitive, Construct	Functional	Psychological means-ends
Other considerations				Performance measures

Researchers appear to agree on the notion of *physical fidelity* as being the degree that the simulation looks like a natural system, including interfaces, spatial layout, sounds and motion. Some argue that the accuracy of the simulated system's dynamics and physical modeling should be included in this category (Elliott et al., 2004), while others suggest that the distinct term *engineering fidelity* be used (Lee, 2001; Maran & Glavin, 2003; R. B. Miller, 1953). Since these two terms seem to be meaningfully different, we will adopt their use. An example of a system with low physical fidelity but high engineering fidelity could be an automobile simulation with few idealizations that can predict to within millimeters a car's stopping distance, but uses a command-line text interface. A converse example would be a state-of-the-art video arcade racing game with full control functionality, force feedback and wrap-around visuals, but with deliberately inaccurate (and hopefully entertaining) representations of vehicle dynamics. Designers often must closely couple engineering to physical fidelity, especially for manual control behavior as exemplified by an aircraft simulator's controls and flight physics.

Psychological fidelity is a catch-all term, encompassing the definition of fidelity as reproducing the *perception* of features of a natural system. It has been defined in different ways throughout the literature, which we will discuss in more detail.

Cognitivist approaches to Psychological Fidelity

One definition of psychological fidelity is "the degree to which the system captures functional and cognitive aspects of the performance domain" (Elliott et al., 2004, p. 121). These authors propose three distinctions within psychological fidelity: functional, cognitive, and construct fidelity. Functional fidelity includes both replication of real-world goals, interdependencies, and scenario events, as well as team roles, responsibilities, tasks, and tactics. Cognitive fidelity qualifies each of these identified "functions" in terms of the presumed complexity of cognition required of simulator participants. Construct fidelity is judged by the degree to which each of the previously identified functions and performance constructs are elicited by simulation scenarios.

These terms are compatible with accepted conventions used in social science, but present two challenges for analysis of correspondence-driven complex systems. First, 'functions' are not clearly defined, and do not distinguish between dissimilar elements such as invariant physical laws and worker team assignments. This makes the use of structured work analysis methods more difficult and increases the subjectivity of analyses. Second, designing a simulation based on comparisons of human cognitive constructs seems to introduces added opportunities for loss of correspondence: one, the introduction of analyst bias when assessing workers and participants (Kirlik et al., 1993) and two, the dependence on either predictive human cognitive models (Stanton, 1996) or simulator design prototype-and-test cycles (Vicente 1999). These weaknesses do not impair the design of a laboratory task that is tightly tied to theory and can be readily modified to elicit the desired cognitive response, but could be risky to use for scaled world or high-fidelity design involving significant design cost or that is difficult to alter after delivery. As in designing complex sociotechnical systems, there is often a need in simulator design to 'get it right the first time' (Naikar & Sanderson, 1999; Vicente, 1999).

Ecological approaches to Psychological Fidelity

Ecological psychology holds that complex cognitive constructs are often best explained in terms of interactions between an organism and the environment to which it is adapted. A compatible description of simulator psychological fidelity would thus give primacy to the functional elements and relationships within natural systems that are psychologically relevant to workers' behavior, and later consider cognitive processes (Vicente, 1999). Hammond (1998) insists that such a form of psychological fidelity is essential to external validity: "If experiments are to produce results that will generalize to circumstances outside the laboratory, they must not merely include substantive material that is representative of the outside situation, but the formal, that is, structural, aspects of the situation outside the laboratory as well." (Hammond, 1998).

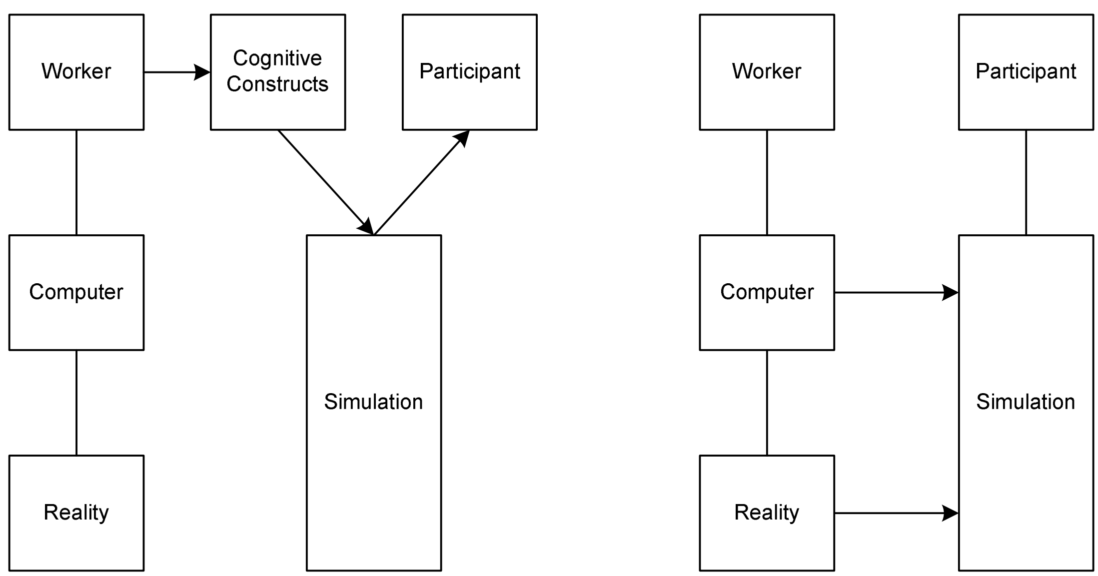


Figure 3 - Cognitivist (left) and Ecological (right) approaches to psychological fidelity

One means to a psychologically relevant structural representation of systems is the Abstraction-Decomposition Space (ADS) (Rasmussen, 1986), which integrates two structures that have been found[3] to describe expert workers' reasoning patterns: means-ends abstraction and part-whole decomposition. The first captures the different levels of abstraction experts use when considering how or why actions should be taken, as well as the languages that can be used to describe the system, ranging from physical form (physical system components' sensory appearance) to functional purposes (the ultimate reasons behind the system's operation). The second involves reducing complexity by grouping parts of a system into meaningful groups.

The degree to which a simulator incorporates more abstract elements of a system, such as functionality, processes, values, and functional purposes has been described as its *functional fidelity* (Lee, 2001). A laboratory task (as defined in Section 3.2.1) will typically capture only physical functions (and perhaps some physical form) of a real system, while a high-fidelity simulation will extend from physical function to functional purposes, with varying levels of detail. Multiple simulations of the same system can be compared by the range of an ADS representation that they cover (Lee, 2001).

Psychological fidelity in an ecological sense must include functional fidelity, but also incorporate the means-ends structure between system elements. This approach has been discussed in the context of simple microworlds, in which researchers note that the explanation given to participants of a simulator (its cover story) should prompt expectations and pre-existing mental models that match the simulator's capabilities and behavior. If a microworld is to focus on some property of real-world systems, such as tight coupling between elements, its cover story should be selected to be analogous with a familiar system that shares similar functions and means-ends relations, or else humans' problem-solving behavior will be distorted (Brehmer et al., 1991). For high-fidelity systems where participants have fixed expectations, the link to psychological fidelity becomes more clear: having "a simulator with parallel means-ends structure to the work domain will allow workers to flexibly exploit means-ends relations that are available in the real environment" (Naikar & Sanderson, 1999, p. 286). While functional fidelity indicates the extent that a natural system's functions, processes, and purposes are found in a simulated system, means-ends fidelity indicates the degree to which participants can employ them to serve the same purposes. Simulator development can result in both omitted means-ends relations (for simplification purposes), or introduction of new ones (sometimes unintentionally[4]). Together, functional and means-ends psychological fidelity provide an ecological foundation for consideration of simulator tasks, tactics, team roles, and operator expertise.

---

[3] This structure for analyzing cognitive behavior has been critiqued and defended in the literature (Lintern, 2006; Naikar, 2005; Vicente & Rasmussen, 1992)

[4] In a microworld simulating an unmanned aerial vehicle command and control environment, a radar function that was intended to be a means to observation and searching was used by several participants as a means to draw the attention of enemy craft and lure them away from vulnerable friendly forces. In this case, the interaction between the terrain properties and the simulated computer opponent's behavior introduced an unintended affordance (Kirlik et al., 1993).

It should be noted that none of these fidelity measures can *specify* the fidelity required of a simulator (Lintern & Naikar, 2000). That question must be answered with respect to the research or training question for which validity is required. A coherent description of fidelity, however, can help structure the design process and meaningfully compare simulators.

Fidelity Summary

The following briefly summarizes the fidelity definitions that will be used in this document:

1. Physical : Similarity in appearance
2. Engineering: Similarity in dynamic behavior
3. Functional: Similar extent of functionality in the natural system
4. Psychological means-ends: Functionality can be used for the same means, and exists for the same ends as in the natural system

3.3 Existing simulator design approaches

Using the vocabulary and distinctions introduced above, we now turn to existing approaches to simulator design. A generic simulator design process is illustrated in Figure 4, adapted from recent literature (N. J. Cooke et al., 2004; Sauer, Wastell, & Hockey, 2000). We will use it as a starting point to discuss two broad categories of simulator design methodologies: those using some form of Cognitive Task Analysis (CTA) (Chipman, Schraagen, & Shalin, 2000) and those using the Cognitive Work Analysis (CWA) framework (Rasmussen et al., 1994; Vicente, 1999).

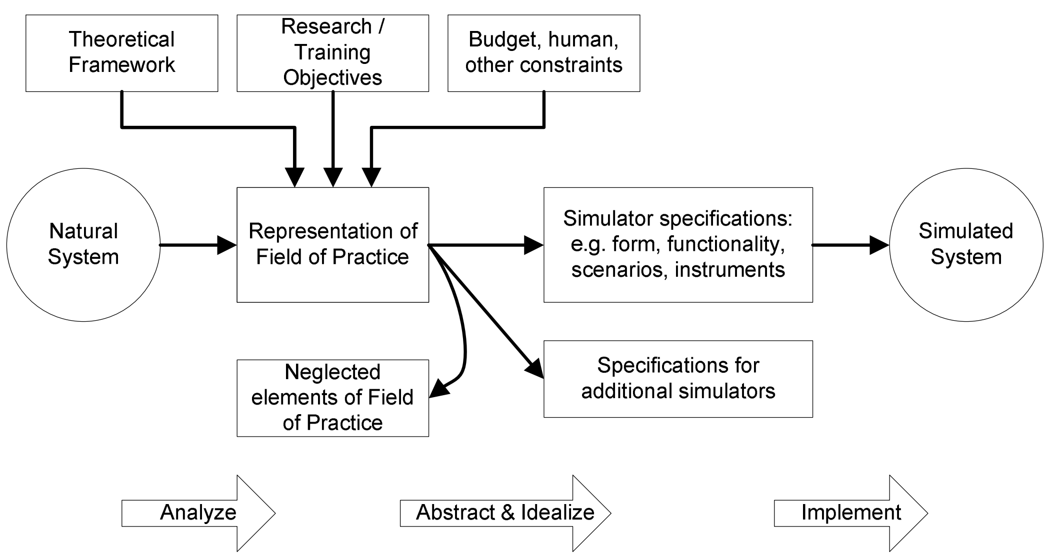


Figure 4 - Generic Simulator Design Process

Three general aspects of the simulator design process shown in Figure 4 are:

1. How to analyze the natural system
2. How to abstract and idealize real-world system properties for the desired level of correspondence, and what to add or remove from the simulator for tractability
3. How to implement the simulator specifications to achieve desired correspondence and engagement

Analysis is the process of generating a representation (or model) of the natural system. It includes knowledge elicitation, tools, research or training objectives, and the theoretical framework that underpins the representation. Both CTA and CWA have been used for this step by different researchers. Abstraction & idealization is the process of generating simulation requirements with the desired physical fidelity, psychological fidelity, tractability, correspondence, and engagement. If the simulator type is a high fidelity simulation of a complex system, a near identical copy of the real world system should be modeled. If a microworld or scaled world is desired, this step will be more crucial in ensuring representativeness. Implementation is less frequently discussed in academic literature, but is a crucial part of a successful design project and often a source of new insight and considerations that must be incorporated into previous stages of development. Prototyping as part of an iterative design process is strongly recommended (N. J. Cooke et al., 2004). Figure 4 does not intend to imply a 'waterfall' design methodology, although in many cases later stages of design are sequentially constrained due to procurement policy, such as in commercial and military work environments. In such situations, correcting simulator specification errors after delivery can be costly.

3.4 CTA-based simulator design

Several variations of CTA methods have been used by researchers in simulator design. While many researchers agree that a CTA of the natural system is useful for Figure 4's analyzing step, and a psychological theory of human behavior necessary to abstract and and idealize (Hinsz et al., 2004; Kozlowski et al., 2004; Sauer et al., 2000; van den Bosch et al., 2004), specifics vary. For example, Hinz et. al (2004) used Wickens' information processing model to design a microworld for research in human memory performance, while Sauer et al. (2000) used Hockey's theory of performance decrement to design a microworld for human attention and cognition research. Gugerty (2004) employs task and goal-focused CTA methods but acknowledges the usefulness of the ecological approach in incorporating constraints inherent to the natural system. He uses a representation tool based on the Goal-Means Hierarchy (C. A. Miller & Vicente, 1998b) to combine task and constraint information. The need to justify an analysis' completeness is recognized: Van den Bosch et.al (2004) emphasize that researchers should consult multiple Subject Matter Experts (SMEs) when analyzing the system, and

Sauer et al. (2000) use operator, trainer, and designer interviews to reduce omissions and inaccuracies in the CTA[5].

Cooke (2004) outlines in somewhat more detail the design process used in her research. She starts her design analysis process using a CTA generated by Gugerty. This is supplemented with site visits, SME interviews, and documentation to form Figure 4's "Field of Practice". Amalgamated with information from research questions and budgetary, technological, and other constraints, this forms what she calls the "Design Feature Space". The next step is to use this body of information to abstract and simplify aspects of the field of practice suitable for generation of a microworld. Cooke admits that

> "The process underlying such decisions, however, is difficult to explicate. That is, the identification of relevant dimensions in the field of practice and the nature of their replication in the [simulator], though critical to the external validity of the design, seems more like intuition or art, rather than the application of well-specified rules. For us, background knowledge in cognitive psychology and human factors and experience in experimentation, task analysis, and cognitive task analysis most likely fueled what looks like intuitive decisions. Although experts may disagree on some relevant dimensions, there is likely more consensus than not and the true success of this abstraction process cannot be determined without the field validation trials." (N. J. Cooke et al., 2004, p. 270)

This ambiguity in proceeding from analysis to design is common to the papers referenced above in this section. With the exception of Gugerty (2004), none have proposed any theory-based design tools to illustrate or guide their analysis, making reproducing or justifying their methods a difficult pedagogical exercise. While all employ a variant of CTA, a model of human performance, and use research questions to guide simulator design, the contribution of each to the simulator's physical form, functionality, instrumentation and scenarios is not clear. The need for a more rigorous method is noted by some:

> "We strongly believe that a systematic theory-based design framework is required on which to found the design of micro-worlds" (Sauer et al., 2000, p. 57)

[5] This experience is consistent with arguments made regarding limitations and incompatibilities in even experts' knowledge of complex systems (Rasmussen et al., 1994; Vicente, 1999).

Examples of the results of some of these design processes, most notably Sauer's CAMS microworld, will be discussed in Chapter 4.

3.5 CWA theoretical framework

Cognitive Work Analysis is a systematic, theory-based design framework developed over three decades for analysis of complex sociotechnical systems and design of information technology support systems (Rasmussen, 1974, 1986; Rasmussen et al., 1994; Vicente, 1999). While we will assume that the reader is familiar with the five phases of the CWA framework as defined by Vicente (1999), we will review some of the philosophical and methodological distinctions of this approach and their suitability for simulator design.

One general distinction is that CWA explicitly describes natural systems in terms of *categories*, in the tradition of the nomothetic approach to scientific modeling as described in Section 3.1.1. This philosophical foundation encourages generalization, specificity in language (Vicente, 1999), and provides support for graphical analysis aids (Rasmussen et al., 1994). Rather than specific tasks, CWA focuses on categories of *constraints*, and describes how these constraints shape what can be done in a natural system. This *formative* approach can help guide design for disturbances or unexpected events that analysts cannot anticipate (Vicente, 1999). A summary of five categories of constraints in CWA (phases) is shown in Figure 5 and explained below.

The first layer of constraints in CWA, shown as the outside grey barrier in Figure 5, is the *work domain*. This covers the physical and social aspects of the natural system, without any control, tools, or actors (either workers or computers)[6]. A work domain is intended to describe possibilities for action in psychologically relevant terms (as discussed in the context of psychological fidelity, in section 3.2.3). This approach may be useful for simulator design, as it explicitly distinguishes between the potentially expensive-to-simulate physical form of a particular system, and functional specifications that may be more tractable to implement.

---

[6] An example of a Work Domain Analysis for car driving would include the purposes of driving (e.g. to arrive at a destination), laws of physics and social laws that determine what cars can and cannot do (e.g. not drive through guardrails), the processes involved in driving (such as braking, or turning), the functionality of a car, and its physical appearance. Such an analysis could be used to study either a human driver or a hypothetical computerized autopilot, since either would have to cope with the same features of the driving environment.

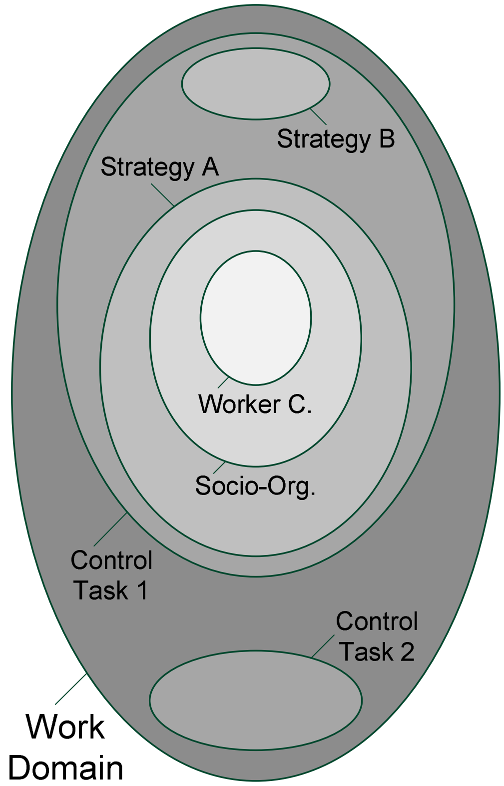


Figure 5 - Five phases of CWA, and their nested constraints (adapted from Vicente (1999))

The second category, *control tasks*, is shown as nested grey ovals in Figure 5. Control tasks describe what can be done in the work domain of a natural system, and when such tasks are appropriate. This type of analysis does not depend on who is carrying out the task, instead capturing both the steps that a novice would have to follow, as well as the cognitive shortcuts that experts can take. This may be useful as simulators must often accommodate both experts and novice participants[7]. CWA also distinguishes between tasks that are required to achieve purposes in the work domain, and *workarounds* that are required by a particular natural system's existing work support tools (Vicente, 1999). This distinction may be useful for simulator designers, as workarounds may not generalize between natural systems[8].

Strategies, the third nested category, represent different ways that actors can mentally perform a control task, in other words different 'angles' that experts use to tackle problems. They are shown within Figure 5's control tasks, because different strategies often cannot be distinguished by task observation. Strategies represent the core of

[7] For example, while a task analysis of an expert driver might observe that they do not use their rearview mirrors when changing lanes, this would not imply that a driving simulator should be built without them!

[8] For example, while the exact steps required to tune in a traffic news station will vary between vehicle radios (or communications technologies), the task of gathering information about remote road conditions is common to urban driving.

expertise, and should be identified if simulator simplifications are not to distort behavior[9].

The same strategies and control tasks may be assigned to different actors, thus the *social-organizational* element in Figure 5. When designing microworlds, simulation designers may need to distinguish the overall work that must be done from the natural system's existing team and automation roles and responsibilities.

CWA places the human worker at the center of these four constraints, indicating that human behavior is constrained by what is possible, when and in what order it can be done, how it can be done, and who can do it, respectively. *Worker competencies* considers human strengths and limitations, such as those addressed in human performance models used by simulator designers in Section 3.4.

3.6 CWA-based simulator design

The CWA framework has been applied to simulator design and research (Lintern & Naikar, 1998, 2000; Naikar & Sanderson, 1999; Sanderson & Naikar, 2000; Vicente, 1995). The first case study of CWA's applicability to simulator design (Vicente, 1995) discussed three cases in which the CWA framework provided useful structure for experiments with the DURESS II microworld. Vicente first proposes that a strategies analysis can be used to categorize the difficulty level of simulator scenarios, both to allow gradual increase in the skills demanded of test subjects and for experimental comparisons. To evaluate performance between subjects, he suggests that more abstract levels of a Work Domain Analysis (WDA) can be used as guides to relevant experimental measures. Finally, he proposes that the phases of CWA can be used to frame performance measure taxonomies: WDA levels represent system state variables, Control Task Analysis (ConTA) results can frame comparisons of task completion times and number of control actions, and so on. This can help ensure a systematic matching of performance measures to experimental questions so that conclusions can be drawn in a scientifically defensible manner (Vicente, 1995).

Benefits of a CWA approach to simulator design have been explored by other researchers (Lintern & Naikar, 1998, 2000; Naikar & Sanderson, 1999; Sanderson & Naikar, 2000), focusing on military training purposes rather than research. The main reasons given for developing a CWA approach instead of existing CTA methods were poor previous experiences with simulator acquisitions, the need to develop simulators that support adaptive behavior, and to design training programs to teach flexible problem solving. In a military training environment, the most important focus of training according to Naikar should be "construction of effective action out of the

[9] Continuing the previous example, expert drivers that change highway lanes without looking at their rearview mirrors might be maintaining a mental model of surrounding traffic flows, or they might be relying on motion cues from their peripheral vision.

resources available for dealing with novel situations" (Naikar & Sanderson, 1999, p. 274). Additional benefits of the CWA approach include a better match with newer functional specification formats for tender contracts (Lintern & Naikar, 2000), a more comprehensible and concise presentation format (Lintern & Naikar, 1998), and a more systematic analysis that is less dependent on known initiation tasks (Lintern & Naikar, 2000).

Experiences from this training simulator application should generalize to design of cognitive engineering research simulators for three reasons. First, designers of both simulator applications aim to induce the same expert cognitive work as in the natural system: training simulators so that participants' expertise will transfer to the natural system, and research simulators so that participants' behavior will generalize to it. Second, both aim to incorporate natural systems' disturbances in scenario development: training simulators so that trainees will become proficient in improvising effective responses, research simulators to reduce learning effects and discourage unrepresentative problem-solving strategies (such as found in the NPP study of Section 3.1.2). Third, this application explicitly recognizes budgetary constraints, which often constrain academic researchers more than industrial or military entities.

Naikar suggests that the results of a WDA can be used to specify simulator hardware requirements so as to have functional and means-ends psychological fidelity (Naikar & Sanderson, 1999), allowing operators to explore the same range of action opportunities available to them in the natural system. When designing simulator functional and engineering fidelity, the Physical Form and Physical Function levels at the bottom of an AH can indicate elements of control functionality and environmental conditions that should be present. She suggests, like Vicente, that the more abstract levels of a WDA, specifically the second-tier Abstract Function / Values and Priorities level should be used as a primary source of performance measurements, with less abstract levels indicating measures that can be used for diagnostic purposes (Naikar & Sanderson, 1999; Vicente, 1995). Well designed simulator scenarios are necessary to enable a simulator with means-ends fidelity to train users to flexibly use the same range of means-ends and experience the same tradeoffs and conflicts as in the real world system. Sanderson suggests that the two most abstract AH levels of a WDA and the four remaining phases of CWA can provide the basis for design of simulator scenarios. In her application (Sanderson & Naikar, 2000), the extent to which the later stages of analysis could be completed was limited as the equipment under study was still in the design and procurement phase.

Table 3 - Previous applications of CWA phases to simulator design choices

	Phases of CWA								
	WDA								
Simulator Design Choices	FP	AF	GF	PFn	PFo	CtrlTA	StrA	SOA	WCA
Physical Fidelity									
Psychological means-ends Fidelity									
Functional Fidelity									
Performance Measures									
Scenario Generation									

Table 3 summarizes the applications of CWA stages to simulator design that have been discussed above. Most design choices have drawn only from the five abstraction levels of Work Domain Analysis (at top), suggesting that opportunities for later CWA stages such as Strategies, Socio-Organizational, and Worker Competencies Analysis have not yet been fully explored (Sanderson & Naikar, 2000).

3.7 Conclusion

In this chapter we have discussed some of the challenges that simulator designers face, and introduced terminology commonly used for simulator discussion. Different approaches to a key simulator attribute, psychological fidelity, have been proposed by researchers using both CTA and CWA paradigms. As a review of the literature shows, CTA-based approaches to simulator design are more widely-used. However, critical analysis of such simulators and comparison of their methods are difficult due to their inconsistent terminology and lack of methodological documentation tools. The CWA analysis framework's structure, terminology, and analysis tools address some of these weaknesses and have been applied to simulator design. While examples of CWA application to simulator design are few, they have shown promising results and may be worthy of development.

Chapter 4 Selected Research Case Studies

This section provides several case studies of simulator designs, with the intent of clarifying distinctions made in the previous theoretical discussion and highlighting opportunities for development of CWA-based simulator design frameworks and tools. Five examples will be presented, spanning the range of simulator types discussed in Section 3.2.1:

- Mass Control Flow (MCF), a laboratory apparatus
- C3FIRE, a microworld
- Cabin Air Management System (CAMS), a microworld
- Terminal Radar Approach CONtroller (TRACON), a scaled world, and
- HAlden Man Machine LABoratory (HAMMLAB)'s high fidelity simulators

Since details of the design process behind simulators are rarely published, much of the discussion will focus on the end product features of the simulator. The selection of case studies is by no means complete: some very interesting simulators were not included due to lack of space, limited development theory content, short research lifespan, or similarity (Elliott et al., 2004; Hess, MacMillan, & Serfaty, 1999; Kirlik et al., 1993; Kozlowski et al., 2004; Reising & Sanderson, 2002b). Other very notable microworlds such as DURESS II (Vicente & Rasmussen, 1990) have been deliberately neglected, since their development has been so closely tied to the CWA framework (Vicente, 1999).

For each of the examples above the simulator will be described, highlights of its development will be noted, and implications for a CWA-based simulator design framework will be presented.

4.1 Mass Control Flow (MCF) Task

This simulator was developed to evaluate computer interface design features such as configural, seperable, and emergent forms (Bennett, Toms, & Woods, 1993). The MCF simulation scope is very narrow, including only a single operator-controlled variable and one control task, which places it in the laboratory task category. Like many laboratory tasks, the MCF evolved from a previous research apparatus (Crossman & Cooke 1974, qtd. in Bennett et al., 1993). Laboratory tasks of this type are a mainstay of human factors experimental psychology, and can be used to illustrate simulator design principles.

4.1.1 Description

The MCF task is loosely based on the behavior of nuclear power plant boilers during system startup, when feedwater flow control presents dynamics that challenge human control (or that of classical PID automation). If feedwater overfills a boiler, the steam turbines downstream must be isolated to prevent water damage, while if the boiler level drops too low, the heat source must be extinguished to prevent thermal damage. Either outcome requires plant shut-down, making skilled execution of this task of interest to researchers.

Dynamics

The key challenge of the MCF system lies in the dynamics of operation. The simulated system consists of a boiler with one feedwater inflow and one steam outflow. The only indication provided for boiler state is the Indicated Steam Generator Level (ISGL), a rough volume measurement that is widely used in industrial applications, but presents challenges for control. This difficulty can be explained in terms of dynamic control properties, thermodynamics, and terms psychologically relevant to workers.

From a control perspective, the boiler system has nonminimum phase dynamics (Bennett et al., 1993), characterized by pure time delays and short-term control reversals. Action to reduce error will initially have an opposite effect, temporarily increasing the error further, making simple PID control strategies difficult to use. For example, increasing feedwater flow if the boiler level is low will first decrease the ISGL, before increasing it. This can also be described in terms of mass and energy relations: the feedwater is a mass and energy source, the steam production a mass and energy sink, and the boiler a reservoir that conserves mass and adds energy through its heating system. Feedwater adds both mass and energy, but because of its low temperature, increases in flow will initially reduce the *energy density* of the reservoir, which will take time to recover through heat addition and steam removal processes. The ISGL indicates volume, which is a function of both mass quantity and energy density (i.e. water level and boiling intensity). The psychologically relevant terms used by workers in power plants to describe this phenomenon are *shrink-swell* effects.

To aid operators in the MCF task, researchers (Haley and Woods, 1988, qtd. in Bennett et al., 1993) developed a compensated steam generator level (CSGL) super-sensor (Stanton 1996). The CSGL estimates future values of ISGL by computing the water mass in the boiler and anticipating “the energy state of the [boiler being] nominal” (Bennett et al., 1993, p. 77). Such predictor indicators have been found to improve human control and are now widely available in industrial automation software (Blevins, McMillan, Wojsznis, & Brown, 2003).

Automation

The only automation in the MCF simulator is alarms for boiler high and low level. If the ISGL falls outside the prescribed safe operating range, the simulation trial is immediately halted.

Social Organization

One operator is tested in the MCF simulator at a time.

Interface

A single computer interface screen is used for the MCF simulator, and at least two variations of interface design have been evaluated. Because of the simplifications made to the system dynamics (see scenarios below), interfaces incorporated only four variables: feedwater inflow, steam outflow, ISGL, and CSGL. A computer keyboard's directional keypad was used for operator input.

Scenarios

Typical of laboratory experiments, the experimental scenarios were tightly constrained. Boiler heating power and feedwater temperature were held constant, leaving steam flow rate as the demand variable and feedwater flow rate to be controlled by the operator.

Scenarios consisted of steam flow rate trajectories generated by adding three forms: a linear ramping function, a sinusoidal pseudorandom function, and a random periodic step disturbance function. Scenarios always started from the same equilibrium initial conditions and could use either an increasing or decreasing ramp function, or none. As the scenario progressed, the step disturbances' magnitude was continuously increased, until five minutes of successful operation, or until the ISGL exceeded an alarm threshold.

4.1.2 Theory / Development

Manual control tasks such as MCF have substantial precedent in psychological research. Three main design objectives are offered to assert the appropriateness of the MCF task for interface design research: 1) similar control characteristics as real world task 2) tractability, and 3) resemblance to a real world problem and displays (Bennett et al., 1993). These three points will be discussed in turn.

To determine the "critical demand characteristics" (Bennett et al., 1993, p. 77) of the real world MCF task, a CTA was performed (Roth and Woods 1988, qtd. in Bennett et al., 1993), presumably following a procedure similar to the CTA approach described in Section 3.3. The dynamics of the water-energy interactions were described in terms of physical processes, operator goals and control theory. The importance of the elements included in the simulation is clear, but the omission of others is not discussed or justified. Possible elements of the natural system that might have strong interaction effects with the MCF task could include control of up and downstream equipment or tradeoffs between system purposes such as safety and financial performance.

Tractability is increased by eliminating elements of the real world task and other reductions of degrees of freedom. Within the simplified task, two performance measures were selected: time on task and memory recall. The first measure indicated how long operators could control the system within the high and low ISGL limits, and the second how accurately and quickly they could report the numerical values of system variables when interrupted during the trials. These performance measures may lend themselves to straightforward statistical analysis, but require justifications to explain how they are meaningful. Time on task is justified by external validity as the "only performance criterion that counts in the real world" (Bennett et al., 1993, p. 91), but differences between the laboratory task and the real world such as the artificial shape of the steam demand scenarios are not addressed. Memory recall's weaknesses are discussed in Section 4.1.3 below. The authors acknowledge the potential weakness of the experimental measures.

The final justification, similarity to real world problem and displays, should be analyzed beyond an appeal to face validity. While physical fidelity of the simulator was low, engineering fidelity was moderate, capturing the need for prediction and cross-coupling of system properties. The dynamics of the simulation, however, were not programmed from first (thermodynamic) principles, and did not model boiler temperature, pressure or other parameters. Other dissimilarities were not discussed thoroughly, but include the limitation of operator control to one variable (the feedwater flow), and the scenarios that increased the disturbance magnitudes without any system-related reason.

Engagement of simulator participants was obtained through the custom of compensation for their time. The much more extensive motivations present in the real world, including time pressures, were not incorporated into the simulation except in the form of increasing difficulty of scenarios.

4.1.3 Implications for a CWA based approach

Three examples from the MCF simulator will be used to illustrate the potential strengths and weaknesses of a CWA-based simulation design approach.

Features of the MCF simulator can be partially captured in an ADS using means-ends relations. The competition between short and long-term control action effects (Bennett et al., 1993) consists of both a "one-to-many" and a "many-to-one" relation, in that manipulation of the feedwater flow provides both mass and energy to the reservoir, but the corresponding reservoir mass and energy balance together affect ISGL. The different time scales that these effects act on are a key part of correspondence in the MFC simulator. Such psychologically relevant temporal relationships are not well captured by the representation tools available in CWA.

A Worker Competencies analysis would suggest that the MCF boiler control task could conceivably be achieved using different modes of human behavior: continuous skill-based control, rule-based discrete impulse control, or knowledge-based open-loop control. However, the MCF simulator's use of a discrete keyboard interface with up or down keys likely discouraged skill-based behavior (SBB), and its randomly generated steam flow demand likely discouraged knowledge-based behaviors (KBB). While these interface properties may be common to some process control natural systems, this illustrates how simulator simplifications may unintentionally distort human behavior.

This example also highlights the difficulty of determining appropriate performance measurements in simulators. For example, one of the performance measures used, memory recall of system state variables, could potentially indicate worse performance for more skilled participants. While it is reasonable that participants using rule-based behavior (RBB) would be expected to use numerical values of system variables to choose their next action, participants using expert skill-based sensorimotor control would be expected to interpret the shape of the MCF display's forms as signals, and therefore would not attend to their numerical values (Olsen and Rasmussen, 1989, qtd. in Vicente, 1999). Thus, participants using a more expert control mode might have poorer recall and corresponding score. The other performance measure, time on task, does not capture 'goodness' of control, and conflates the two simulated system purposes: because of the simulator design, the steam production goal must always be satisfied until the ISGL exceeds alarm limits.

Laboratory tasks based on complex sociotechnical systems must necessarily make drastic simplifications. These three examples show how a CWA-based approach may apply to their design and justification.

4.2 C3FIRE and predecessors

The C3FIRE family of microworlds have possibly the longest active research life of any simulator of comparable fidelity, extending two decades from 1986 to the present (Allard 1986, qtd. in Brehmer & Allard, 1991; Jobidon, Breton, Rousseau, & Tremblay, 2006; Lindgren & Smith, 2006). Developed and primarily used at the Swedish Linköping and Uppsala universities, C3FIRE or its predecessors have also been used by the Danish Risø national research center, and as far afield as the Université Laval, Canada.

Research topics have included effects of feedback time delays, system opacity, and in later versions, team structure and communication (Brehmer et al., 2004). C3FIRE has proven an adaptable research platform.

4.2.1 Description

C3FIRE is a forest firefighting supervisory control microworld, but follows Gray's definition (Section 3.2.1) in that it is not intended to generalize to this domain. Rather, the role of a firefighting chief was chosen as an exemplar of Distributed Dynamic Decision-making (DDD or D3) work environments that would have familiar enough elements for novice participants to use (Brehmer & Allard, 1991). Later, with a shift from theoretical to applied research, the name was changed to Command, Coordination, and Communication (C3)FIRE.

The C3FIRE platform has been developed through several versions, beginning with the single-user Dynamic Environmental Simulator System (DESSY) and continuing with NEWFIRE, the D3FIRE multi-user upgrade, and the significantly redeveloped C3FIRE (Brehmer et al., 2004). These simulators differ in some respects, which will be highlighted here where relevant.

Dynamics

The FIRE family of microworlds simulate an environment consisting of a top down view of 20 by 20 grid landscape, containing flammable timber in which forest fires can spread and fire fighting units (FFUs) can be maneuvered to combat the spread of the fire. The environment contains areas that must be protected from fire, such as a single fire command base in DESSY, and more diverse villages and fueling structures in C3FIRE. Forest fires spread according to a simple engineering fidelity model incorporating the flammability of the neighboring terrain and the direction of the prevailing wind. Wind direction is often manipulated as part of experimental scenarios (see below). The low engineering fidelity of this approach is acceptable, as the simulator is not intended to generalize to better means of firefighting, but rather to provide an collaborative problem solving environment that contains dynamics familiar to a wide range of participants (Brehmer & Allard, 1991).

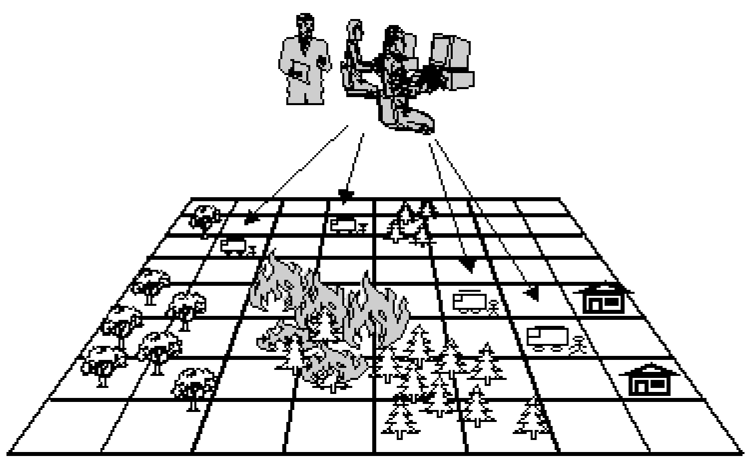


Figure 6 – General overview of C3FIRE microworld (adapted from Granlund (2003))

The original DESSY simulation's eight FFUs have been extended in C3FIRE to a wide range of units including both helicopter and truck FFUs, as well as water and fuel supply vehicles (Lindgren & Smith, 2006). Movement speed of FFUs is only moderately faster than that of forest fires, presenting a challenging time dynamic. This requires operators to predict the future location and extent of fires when dispatching FFUs if they are to be efficient (Brehmer & Allard, 1991). FFUs introduce time delays when mobilizing before fighting fires, demobilizing before movement, and of course take time in fighting fires.

In addition to the time dynamics inherent in the simulated system, communication lags both in reconnaissance and in command transmission have been introduced by researchers.

Automation

The FIRE family of simulators do not contain any automation, with the exception of some of the first experiments with the single-user DESSY. The FFUs in DESSY could be delegated authority to stop and fight any fires they encountered while responding to a participant's movement command. Despite being a powerful option to cope with communication delays, this feature was poorly adopted by participants (Brehmer & Allard, 1991) and was removed in later versions in favor of investigating authority distribution in multiple human participants.

Social Organization

Social organization and coordination has been one of the primary research topics for the FIRE family of simulators, and each version has enabled more complex social structures. While DESSY had only a single operator, it contained FFU automation that allowed some delegation to take place. D3FIRE allowed up to four participants to control the simulation, allowing researchers to reproduce most communication patterns studied in the literature, including hierarchical and distributed arrangements (Brehmer et al., 2004), and C3FIRE permits any arrangement of participants.

The simulators allow social organization to either be imposed by researchers or developed by participants. Communication between participants is by text messages, which increases tractability both by allowing authority structures to be enforced through communication restrictions, and allowing communication to be more easily coded and interpreted for experimental results (Lindgren & Smith, 2006). Criticisms of the validity of this design choice have been anticipated: face-to-face and email communication were compared in an early D3FIRE study and no performance differences were found (Brehmer et al., 2004), which has been used to justify the exclusive use of text communication in subsequent work.

Interface

The physical fidelity of the original DESSY interface was very low with respect to forest fire fighting, being only a monochrome computer screen and keyboard. Control of FFUs and other participant interaction was carried out with command-line syntax. In current revisions, such as C3FIRE, interface elements allow direct manipulation, and command-line interaction has been reduced or eliminated. During the same period, work environments such as military command and control have become more similar to the microworld with the inclusion of computer-mediated situation maps and communication systems. The physical fidelity of the C3FIRE microworld has arguably increased with time, something that may be expected of other computer-mediated simulations of non-mediated systems (Elliott et al., 2004).

C3FIRE's interface is partially computer and partly human-mediated. In studies involving large teams, the staff assigned to planning and strategy are often made dependent on communications from other participants who play the roles of fire fighting unit chiefs. This imposed social organization introduces a layer of human mediation into the strategic planners' interface to the simulation (Granlund, 2003).

Scenarios

Scenarios of the FIRE simulators consist of map and FFU initial conditions, locations and times for fires to start, and times for wind direction and speed to change. Other parameters such as unit firefighting abilities and communication time delays can be set for each trial but remain constant.

C3FIRE scenarios reported in the literature have not focused on unanticipated events; simulated events were limited to fires and wind shifts, whose location and time may be unexpected but are certainly not unanticipated. Research has even used repeated trials of identical scenarios (Brehmer & Allard, 1991), allowing participants to fine-tune their responses with practice. In general, C3FIRE has depended on the behavior of participants to create variation between trials.

Instant replay of scenarios has also been built into C3FIRE. This allows participant review and commentary and has been found extremely useful in increasing participant engagement, for training use, and in place of verbalizations (Granlund, 2003).

4.2.2 Development and Theoretical Justification

Development of the FIRE family of simulators is documented mostly in thesis work and technical reports throughout its 20 year life: DESSY from Uppsala University (Allard 1986 qtd. in Brehmer & Allard, 1991), NEWFIRE from Risø National Laboratories (Løvborg & Brehmer 1991 qtd. in Brehmer et al., 2004), and C3FIRE from Linköping University (Granlund 1997 qtd. in Brehmer et al., 2004). Terminology used in the development of these simulators has shifted with time, from DESSY's dynamic decision-making paradigm (Brehmer & Allard, 1991), to include Situation Awareness in later C3FIRE work (Granlund, 2003). The community of researchers using these simulators have enjoyed unparalleled success in maintaining a microworld research program for such an extended period.

DESSY's development in 1986 preceded both CTA and CWA methods discussed above. The only existing theoretical framework considered was that of classical optimal choice theory, which was rejected due to its inapplicability to continuous dynamic systems. A control theoretic perspective was adopted instead, and the simulation's complexity discussed with respect to the ability of agents to generate appropriate control actions. Some suggestions for simulator complexity measures are number of processes, goals, potential actions, and side effects (Brehmer & Allard, 1991). Dynamic system features proposed for correspondence include rates of change, speed of controlling versus controlled processes, feedback delays, and feedback quality. Challenging human abilities with such temporal features was the main goal; as one researcher commented, "simply involving a large number of persons under stress seems to reflect more real-world problems than the fidelity of the simulation" (Johannson, Persson, Granlund, & Mattsson, 2003, p. 195).

The use of multiple participants seems as if it must reduce tractability through added data collection alone. This concern has motivated many features of D3/C3FIRE, such as text-based communication constraints, automated language coding, and shared graphical maps. C3FIRE's tractable communications analysis has helped in evaluation of work distribution and team effectiveness (Granlund, 2003), but additional performance measures have proven more difficult. Researchers have found that while technological development of features within a microworld is comparatively easy, determining meaningful performance metrics and tractable measurement to support them remains difficult (Granlund, 2003). An example from C3FIRE development illustrates the difficulties to be found in even comparatively simple microworlds.

A seemingly comprehensive performance measure in the C3FIRE microworld is the amount and value of terrain burned by the fire, as this is the main criteria by which firefighters in a natural system are judged (Brehmer & Allard, 1991). However, this measure has been found to be insufficient to indicate team effectiveness (Granlund, 2003). In trials where some FFU situation reports were falsified to simulate C3 environments' imperfect communications, teams that ignored communication accuracy put out fires more quickly. While microworlds' limited degrees of freedom can make these types of strategies effective, in real natural systems they would be considered reckless and counter-productive. As one researcher put it "Our experiences (sic) is that a simple scoring system that is only based on pure effectiveness is not enough … when the environment and the task get as complex as it occurs in the C3Fire environment." (Granlund et al., 2004, p. 187)

The FIRE family of simulators illustrates how tractability can be improved through simulator communication and automation design, how lower correspondence can open up a range of research projects, and how participant engagement in training and research can be fostered.

4.2.3 Implications for a CWA approach

DESSY's designers' control-theoretic viewpoint and rejection of normative task models is shared by CWA's advocates (Rasmussen et al., 1994; Vicente, 1999), so it should not be surprising to find some compatibilities between the FIRE development process and CWA. For example, the terms suggested to describe system complexity (processes, goals, action possibilities, side effects) are comparable to those of a WDA system description.

Similarly, the delegation of firefighting ability to FFUs in DESSY can be modeled using CtrlTA decision ladders. While the author has not found strategies used by C3FIRE participants discussed in detail, the social organization of the simulator participants can be modeled in terms of form, content, and areas of responsibility for the work domain (Vicente, 1999). Finally, feedback effectiveness can be considered in terms of the match between information system features and worker competencies.

On the other hand, some of the temporal characteristics used to describe the simulator, such as the rate of change of system variables, the relationship between control and controlled processes, and system time delays fit well with control theory but are not well captured with existing CWA analysis tools.

The main opportunity for a CWA-based approach is in supporting the development of performance measures, in translating "high-level abstract knowledge of the task and the training to a low-level description of what to measure." (Granlund et al., 2004, p. 189). Dependent measures for early experiments with DESSY are mostly task time-based, including only the physical function-level evaluation of burned forest area and a functional-purpose evaluation of base survival. Later work with C3FIRE highlighted the limits to these performance measures, as discussed above.

4.3 Cabin Air Management System (CAMS)

The Cabin Air Management System microworld, first introduced by Hockey, Wastell and Sauer (1998) is an abstracted and idealized representation of a space station life support system under supervisory control. It has been used to investigate topics including human performance decrements under sleep deprivation (Hockey et al., 1998), effects of long term (135-day) spacecraft simulations (Sauer, Hockey, & Wastell, 1999), and more (Sauer et al., 2000).

4.3.1 Description

CAMS requires operators to maintain a healthy living environment inside a crew cabin, and to minimize consumption of pressurized gas stores. Complexity is presented through system dynamics, automation, and equipment faults. Operator loading is further increased through secondary tasks such as alarm acknowledgement and intermittent tank level recordings.

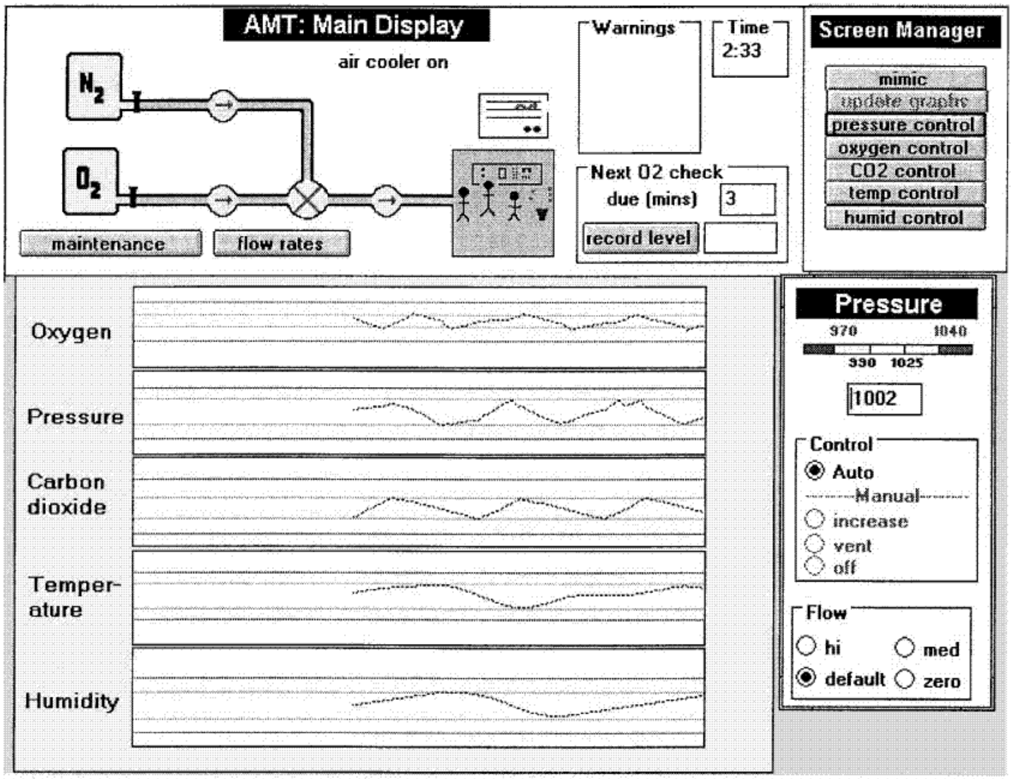


Figure 7 - CAMS microworld main screen (reproduced from Sauer et. al. (2000))

Dynamics

In the CAMS microworld, five state variables describe the cabin interior conditions: oxygen (O_2), cabin pressure, carbon dioxide (CO_2), temperature, and humidity. Each of these state variables must be maintained within a 'safe' range through manipulation of sources and sinks, some of which are coupled. For example, CO_2 is removed from the air with a CO_2 scrubber, which has a side effect of rapidly reducing cabin pressure. Cabin pressure is increased by supplying a blended flow of O_2 and nitrogen (N_2) from storage tanks. The microworld includes alternate means to achieve system functions: for example, O_2 can be supplied either from the main system tank, or from an auxiliary backup cylinder. System lags between control adjustment and system response are roughly 2-3 minutes, long enough to cause manual control difficulties (but not as long as those experienced in many real-world process tasks). Dynamics are idealized, and neglect environmental influences present in real-world cabin air management systems, such as day and night cycles.

Automation

During normal operation, simple set-reset automation maintains system variables in a safe range, leaving the human operator in a supervisory role. In early experiments (Hockey et al., 1998), a 'system-oriented' experimental condition had automation also limiting times at which humans had full access to system information. As discussed below, automation reliability was varied during experimental scenarios.

Social Organization

CAMS is controlled by a single operator.

Interface

The CAMS operator interface uses a standard PC for its display and controls. The screen includes a system mimic diagram, trend charts for each of the five system variables, and 'faceplate' control panels for their respective equipment subsystems. Finally, a maintenance control panel prompts operators for a diagnosis of system faults (described below). A navigation panel provides access to each view and increases tractability by recording users' information polling.

Scenarios

Experimental scenarios for CAMS are comprised of three classes of events: equipment malfunctions, automation failures, and interface faults. Equipment malfunctions include valve blockages, valves failing open, gas leaks and subsystem performance loss, with three varieties of each type. Automation failures disable a system variable's low or high set point, resulting in equipment failing to start or being 'stuck on'. Interface faults involve the disabling of a faceplate control panel, and are normally triggered simultaneously with a fault in the associated subsystem. In total, 20 different faults are possible.

Fault diagnosis is considered a secondary task. When the operator has identified a fault, they can click the appropriate entry on the maintenance control panel to repair it. No feedback is given as to the correctness of a diagnosis.

A tertiary task is provided by generation of arbitrary false alarms, at a rate of one per four minutes.

4.3.2 Development and Theoretical Justification

The development of CAMS was informed by both research questions and a real-world domain, and followed a CTA approach to simulator design, as discussed in Section 3.3. It initially served as an apparatus to investigate Hockey's theory of performance decrement (Hockey et al., 1998). This model predicts that humans can sustain high workload for short periods of time by increasing effort, after which they adapt by changing strategies and neglecting secondary tasks. This theoretical framing required that the microworld have a cognitively challenging environment, several tasks of varying priorities, and scenarios long enough for performance decrements to manifest. The choice to instantiate the design in the aerospace domain was made largely due to the funding source, a manned space program research initiative with the European Space Agency (Hockey et al., 1998). To capture "specific domain features" of a real CAMS, task analysis, literature review, and research from process control domains was used (Sauer et al., 2000). The results were summarized as requirements for: "closely coupled subsystems, opaqueness, high [level] of automation, restricted access to system controls, considerable monitoring requirements, false alarms, [and] a dynamic autonomous underlying process" (Sauer et al., 2000, p. 49). For external validity, they called for moderate engineering fidelity and physical fidelity as well as internally valid performance measures.

While the resulting design has been applied to a wide range of research, some design choices that limit its applicability to more complex sociotechnical systems can be noted. For one, the possibility of triggering unanticipated events was not included in CAMS. While by definition no simulated system fault can be truly unanticipated, CAMS provided the operator with a check list of every possible system fault in the interface's maintenance panel, for reasons of tractability in recording diagnoses. As discussed in Section 3.1.1, humans have been shown to drastically reduce the scope of their problem solving in simulated tasks, and this design choice limits that scope to only 20 options. Secondly, CAMS does not include unpredictable disturbances, a choice also possibly motivated by tractability so as to reduce performance variation. Thirdly, the set-reset automation in CAMS is extremely simple, so results may not generalize to modern automated environments. Fourthly, false alarms are implemented as a secondary task with no relevance to system condition, in contrast to the diagnostic use of alarms observed in natural systems (Mumaw, Roth, Vicente, & Burns, 2000).

4.3.3 Implications for a CWA approach

Design choices made with CAMS' CTA-based approach provide topics for further elaboration of a CWA-based approach. Interconnection of means and ends, a core WDA analysis focus, is found twice in CAMS: a main and backup oxygen supply are both available, and cabin pressure can be decreased by either the CO_2 scrubber or by a direct vent. The converse relation of many ends accomplished by the same means (i.e. side effects) are also found in the pressure interactions of gas supply and CO_2 scrubbing. These relations can help in highlighting areas of difficulty: cabin pressure, the most coupled variable, was the hardest to maintain in an acceptable range in one CAMS study (Hockey et al., 1998). When simplifying a natural system, CWA may help select which equipment functionality and means-ends relations to include.

Not all features in the CAMS system are easily represented in a WDA. False alarms, for example, are completely unrelated to the state of the system, and the task of manually recording tank levels does not serve any explicit purpose. Their non-contribution to microworld performance may have affected participant engagement, such as in the example of C3FIRE false information verification task discussed in Section 4.2.2. Such tasks cannot be dismissed only because they do not contribute to the functional purpose of the simulated environment – in the real world, secondary tasks such as maintaining record logs are crucial to effective and safe long-term maintenance. However, these secondary tasks must often be motivated by social monitoring and enforcement. Another C3FIRE characteristic not represented in a WDA are time lags, identified as one of the “specific domain features” (Sauer et al., 2000, p. 47) for inclusion. This challenge mimics that for MCF, discussed in Section 4.1.3.

Like C3FIRE, the CAMS microworld was developed from a theoretical framework and a research question, and instantiated in a convenient domain. This suggests that a CWA-based simulator design framework should be capable of supporting simulator design either “bottom-up” from analysis of operational domains or “top-down” from human performance theories and research hypotheses.

4.4 Terminal Radar Approach CONtroller (TRACON)

TRACON is an Air Traffic Control (ATC) scaled world, commercially developed in the mid-1980s from a professional training hardware and software suite (Wesson, 2007). TRACON was modified for PCs and marketed as a video game for home enthusiasts. Both the professional and video game versions of TRACON were successful, the commercial version being adopted by many colleges with ATC training programs, as well as by the US Federal Aviation Administration, National Aeronautical and Space Administration, and Department of Defense (Ackerman, 1992). The videogame version sold well and was updated through three versions. Development ceased in 1992 with TRACON II for Windows, which is still regarded as one of the best ATC simulator video games made. As the video game variant of TRACON has been more widely used in the research literature, it will be described in this section. With minor changes, the general appearance and behavior of the TRACON family of simulators has remained the same throughout its commercial development.

The combination of fidelity and low cost found in the video game version of TRACON proved very attractive to human factors researchers. TRACON II continues to be used for research, as shown by published work in human performance prediction, fatigue and adaptation, and situation awareness (Ackerman, 1992; Farbos, Mollard, Cabon, & David, 2000; Kilgore, St-Cyr, & Jamieson, in press; Nunes & Matthews, 2002; O'Brien & O'Hare, 2007).

4.4.1 Description

In the simulated environment, participants are responsible for air traffic within a sector containing several airports and radio navigation beacons. Participants must safely and efficiently direct aircraft: safety requires aircraft be separated at all times by at least 3 miles or 1000 feet of altitude, while efficiency refers to aircraft arriving at their respective destinations either at the edge of the sector or at one of the airports within it.

Dynamics

Airplanes in the TRACON environment range from light single-engine propeller planes to heavy commercial passenger jets, each with varying levels of maneuverability. Depending on their type, aircraft have allowable ranges of speed and altitude as well as rates of turning, climbing, and descent.

For aircraft to be effectively directed, their constraints must be respected by participants when issuing course, altitude, or speed commands. For example, a Boeing 747 jet will take more room to turn, but can climb or descend quickly. Engineering fidelity of aircraft dynamics is medium: while turns, speed changes, and descents are modeled as simple linear transitions, important couplings between aircraft variables are preserved. For example, aircraft cannot maintain both a high speed and rate of climb at the same time.

While TRACON's dynamics are relatively simple, the interactions that arise from many aircraft in a confined area are not. All interactions with aircraft are time sensitive, so additional aircraft add baseline communication requirements proportional to their number. More significantly, potential interactions between aircraft scale geometrically. Even more complexity is added by variations in aircraft handling properties, which interact with time pressure to limit safe and effective choices in resolving conflicts. The emergence of this complexity is reminiscent of C3FIRE researchers' observations, as discussed in Section 4.2.2.

Unlike simulations such as MFC, C3FIRE, and CAMS, TRACON has non-deterministic components, such as aircraft pilots not responding to commands, requiring multiple landing attempts, or having in-flight emergencies that require special attention. These are scenario-specific, and described below.

Automation

Because the TRACON simulation simulates aircraft pilots, airport control tower staff, and other ATCs as computer actors, automation in TRACON overlaps with social organization. For example, participants can delegate responsibility for an aircraft to control tower staff when the plane is on the approach path of its destination airport.

The only non-computer-actor automation present in the simulation is notification of participant errors, such as aircraft near-misses or if an aircraft is lost by being directed out of radar range. These notifications are only for post-hoc performance feedback and do not provide any advance guidance to participants.

Social Organization

As discussed above, TRACON uses computer actors to simulate the role of humans in an operational ATC system[10]. To make this substitution, communications must be constrained to be machine-readable. Thus, TRACON's communication content is limited to a simple command syntax. Responses from computer actors are either acknowledgements, justified refusals (e.g. "Sorry, your speed request is below my aircraft's minimum speed"), requests for command re-transmittal, or requests for

---

[10] A feature offered by TRACON II was to data-link with a popular flight simulator game over a serial cable, allowing two friends to play the roles of pilot and air traffic controller. This feature has not been used in research, to the author's knowledge.

emergency assistance. Likewise, the form of communication is changed by computer mediation from radio-mediated voice conversations to mouse-selected and typed commands[11]. Computer actors can make errors in interpreting commands, either forgetting to follow them, or following them incorrectly. However, these errors are probabilistic (see scenarios below) and cannot be reduced by participants (e.g. by typing commands more slowly and clearly).

Interestingly, as ATC systems become more integrated with aircraft autopilots, their communications may grow to more closely resemble TRACON, yet another example of microworlds anticipating the computerization of their natural systems (Elliott et al., 2004).

Interface

A circular radar scope occupies most of the TRACON interface. Airports, radio beacons, and aircraft are represented iconically, as are land features such as mountains. Like operational radar, aircraft information is updated in roughly one second intervals, so aircraft move in small steps rather than continuously. Achieving moderate physical fidelity in the TRACON simulation is a fairly straightforward process since operational TRACONs' aircraft interactions are already mediated by radar and radio workstations.

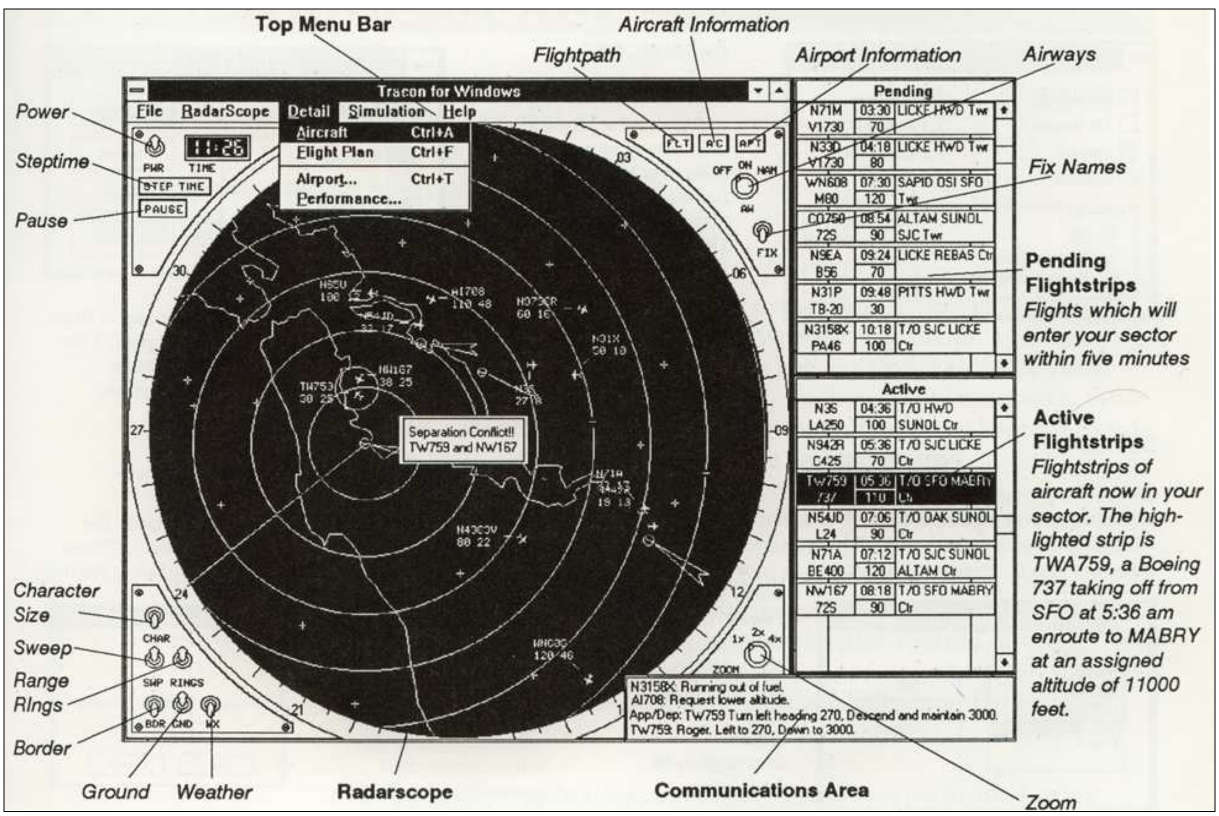


Figure 8 - Screenshot of TRACON for Windows (Wesson, 2007)

[11] Recognizing this difference, speech emulation of computer actors' communications was an early feature available to users owning powerful machines, a novel feature for the late 1980s.

An exception to the physical fidelity of TRACON is the ubiquitous paper flight information strips used in operational ATC to account for aircraft and ensure transfer of responsibility between workers. Paper strips contain a summary of a flight's critical information and are physically passed to and from controllers as each aircraft enters or leaves their zone of responsibility. These physical tokens have persisted in operational ATC despite calls for their computerization (Berndtsson & Normark, 1999). As physical objects, they have advantages over computer-mediated entities: they are difficult to accidentally duplicate or destroy; they are directly manipulated; when stacked in racks, they provide an indicator of controller work load that managers can perceive from across the room; and they are easily customizable with a pencil. The TRACON simulator replaces these physical strips with computerized mock-ups similar in appearance, but which cannot be modified by participants. The effects of virtualizing paper strips ability have not been definitively investigated, but in one study preventing note-taking on paper strips was shown to reduce recall of past commands issued by ATCs who had little video game experience (Zingale, Gromelski, Ahmed, & Stein, 1993).

Scenarios

TRACON scenarios are defined by static and dynamic parameters. Some of the static parameters include:

- The map of the local area, including terrain features like mountains or water, radio beacons, airports, and the shape of the ATC's area of responsibility
- Wind velocity and weather conditions such as clouds and rain.
- Chance of pilot error
- Radar equipment performance and reliability

These invariant parameters constrain the behavior of aircraft in the scenario. Within the area of responsibility, mountainous terrain limits safe flight, radio beacons limit efficient navigational commands, and airports define supply and safe disposal of aircraft. Wind direction determines the approach vectors for airports, and limits the ground speed that aircraft can attain. Visibility and storminess increase the likelihood that pilots will miss landing approaches and require redirection. Pilot competence determines how frequently pilots miss commands or misinterpret them. Finally, equipment reliability can be decreased so that aircraft disappear from radar when at low altitudes or, worse yet, radar contacts can vanish sporadically.

The remaining components of scenarios are scripted, either by a random scenario generator or by handcrafting scenario files.

- Aircraft types, arrival times, locations, velocities, altitudes, and destinations
- Thunderstorm extents, paths, and resulting airport closures
- Probability of pilot emergencies

As discussed above, the number and properties of aircraft largely determine the difficulty of scenarios. Aircraft that are simply flying through the sector often need no intervention beyond acceptance and handoff, while aircraft that are landing at an airport require much more attention from participants. If the arrival time, altitude, location, and velocity of aircraft are such that they will intersect or lose their required separation, intervention is mandatory. Thunderstorms aggravate the difficulty of scenarios, requiring that aircraft be directed around the moving stormfront and arrivals held aloft if airports close due to storm activity. Pilot emergency requests for altitude changes, speed reductions, or low fuel priority landings add more demands.

Creating scenarios that are unmanageable for even expert air traffic controllers can be very easily done in TRACON simply by adding more aircraft. Even without complications such as storms or poor pilot behavior, handling 30 aircraft over a 45 minute session generally exceeds the ability of expert participants (Ackerman, 1992). However, as discussed above, scenario difficulty depends on more than just the number of aircraft, so scenarios with randomly generated flight patterns can vary greatly in the demands they place on participants. This is confounding for both training and research. TRACON researchers have typically overcome this obstacle by handcrafting scenario files so as to be equally matched, but this approach can be quite laborious (Loft, Hill, Neal, Humphreys, & Yeo, 2004).

4.4.2 Development and Theoretical Justification

The TRACON simulation has been designed with "a substantial reduction of rules and operational demands in comparison to the real-world job of an air traffic controller" (Ackerman, 1992, p. 601). While operational ATC trainees take 3 years or more to be certified (Sells et al. 1984 qtd. in Ackerman, 1992), novice participants in TRACON can achieve stable performance with only 20 hours of practice time (Backs, Navidzadeh, & Xu, 2000). Despite this simplification, the designers' intended audience was ATC trainees, and some researchers have opted to use such skilled research participants (Nunes & Matthews, 2002).

Unlike the other simulators described here, TRACON was developed commercially as an operational ATC trainer. Researchers have thus contributed much less to its development and have not described or justified simulator design choices. In fact, some evade the question altogether by asserting that a simulator is "not a scientific experiment" and "formal experimental design is not considered important, and may not always be applicable" (Farbos et al., 2000, p. 204). While modified versions of TRACON have been provided to researchers by special request (Ackerman, 1992), the added features seem related only to more extensive data logging.

As a result, while researchers have made theory-driven insights into the contributions of TRACON simulation components to difficulty (Nunes & Matthews, 2002), the simulator's data-collection capability has not always leveraged such distinctions to improve analysis tractability. For example, as discussed in Section 4.4.1, aircraft are not equal in their demands on participants: aircraft overflying the sector have much simpler ATC guidance requirements than those landing at an airport within the sector. These distinctions could not be fully used in one study (Ackerman, 1992) because TRACON's data collection routines did not record the origins of aircraft that were 'lost' due to mis-navigation or collisions. Thus, performance measures used in TRACON research tend to be either simple outcome-based measures such as the number of flights successfully handled minus the number of failures (Ackerman, 1992) or unjustified re-use of the video game's own scoring system (Farbos et al., 2000).

4.4.3 Implications for a CWA approach

Three examples from this discussion of TRACON will briefly illustrate opportunities for extending simulator design methods. Determining comprehensive performance measures are again important: for example Ackerman (1992) modified TRACON's scoring system to "encourage subjects to develop an appropriate strategy for task component emphasis" (Ackerman, 1992, p. 602). While useful for tractability, these approaches warrant caution: the experience of C3FIRE researchers (see Section 4.2.2) suggests that performance measures with a narrow scope can encourage participants to take more risks than warranted in natural systems and may distort correspondence. Again, this may be an opportunity for CWA-based approaches to aid in systematically determining as comprehensive performance measures as possible, so that participant engagement can be boosted without hidden tradeoffs.

Previous CWA applications suggest that the Abstract Function level of a WDA can indicate performance measures (see Section 3.6). An analysis carried out by Kilgore, St-Cyr, and Jamieson (in press) determined six abstract functions for the TRACON simulation: 1) maintenance of ATC aircraft responsibility, 2) maintenance of aircrafts' field of safe travel, 3) aircraft scheduling demands, 4) passenger comfort limits, 5) performance limits of individual aircraft, and 6) pilots' situation awareness. TRACON's scoring system evaluates 1), 2) and 3), but only partially incorporates 4). Since TRACON's simplified communications do not allow course correction instructions to have different urgencies, 4) and 5) are less meaningful as performance measures, since pilot actors always respond to ATC instructions using the full limits of their maneuvering functionality. Because pilots in TRACON are automated, 6) is also less meaningful as a performance measure. Thus, the performance measures suggested by a CWA not only corroborate those used in the simulation, but highlight areas where simulation design choices have reduced correspondence with the natural system.[12]

[12] Two possible design changes to address these missing performance measures could be: 4&5) change communications and engineering fidelity to allow ATCs to request normal or urgent course corrections, which automated pilots would respond to with either a gentle or abrupt aircraft manoeuvre. Abrupt manoeuvres could be considered more disruptive to passenger comfort. 6) at a cost to tractability, allow

When simulating highly social environments such as ATC, substituting computer actors for humans can greatly increase tractability. However, the validity and ease of development of such actors depends on the types of interactions required of them. In a simulation with simple world state and a very small communication vocabulary, such as TRACON, design of these actors may be conducted in an ad-hoc manner. But for more complex systems that require more extensive communication between participants and actors, the design of such computer actors grows much more complex. While TRACON's constraint of form and content of social interaction leads to greater tractability, in operating environments rich in uncertainty and ambiguity, strong communication restrictions may introduce risks, as illustrated by military friendly-fire incidents (Snook, 2000). A CWA-based simulator design framework should provide a systematic method for considering both effects of changing communication forms and requirements for communications content, either for communication between humans or for design of computer actors.

Finally, as discussed above, even in simplified microworlds such as TRACON, scenario development can be laborious and time-consuming. One way of addressing this problem is with scenario generation tools that allow experimenters to specify scenarios in terms of relevant higher-order properties, and delegate the computation of suitable initial conditions to software. This approach has been adopted by researchers developing ATC-Lab, a microworld intended to supplant TRACON for research (Loft et al., 2004). Their script developer allows experimenters to specify aircraft conflicts in terms of the point of aircraft intersection, distance of minimum separation, aircraft speeds, and the desired time of conflict. The software then calculates several flight plan options that satisfy these requirements and presents them to the experimenter (Loft et al., 2004). A useful output from a CWA-based simulator design framework would be a systematic means to identify potentially relevant higher-order properties, and the means by which they can be induced through scenario design.

4.5 Halden Man-Machine Laboratory (HAMMLAB)

HAMMLAB is a high-fidelity nuclear control room research simulator, part of the Safety Man-Technology-Organization group of the OECD Halden Reactor Project (IFE, 2006). Research conducted since its inception in 1983 includes evaluation of a variety of automation, alarm, procedure, interface, and diagnostic systems, as well as development of methods and measurements for human performance assessment (IFE, 2006).

---

participants to communicate verbally with simulated pilots and have researchers evaluate the relevance and quantity of situation-related information communicated, such as "storm clouds bearing 315, change course 045, maintain 10000, traffic above and below you"

4.5.1 Description

Dynamics

HAMMLAB's simulator currently supports two distinct high engineering fidelity simulations of nuclear power plants (NPP): the boiling water HAMBO reactor, and the pressurized water FRESH reactor. While the simulation dynamics for each of these simulators have been developed from specific NPPs (the Swedish Forsmark-3 and French Fessenheim-1, respectively), their general configurations represent reactor types found around the world. NORS, an older simulation model used in HAMMLAB's inception in 1983, has been retired; it was based on a Finnish/Russian pressurized water reactor.

Engineering fidelity for HAMMLAB's simulations is very high: the stated goal of the research program is to sacrifice as little operational correspondence as possible while maintaining scientific defensibility (Skraaning jr., 2003). As a high-fidelity simulator of a complex world, both HAMBO and FRESH use professional NPP operators from their respective plants as research participants.

Automation

Both simulation models are accompanied by the extensive automation found in their respective NPPs. Automation types found in HAMBO and FRESH include continuous control loops and step-based procedural automation. More recent automation developments such as Model Predictive Control (MPC) and cost optimizers available in modern Digital Automation Systems (Blevins et al., 2003) are not present.

Many automation effects have been studied, some examples are differences in failure modes (electromechanical failure vs. functional mismatch), and performance and trust in diagnostic (knowledge-based) versus procedural (rule-based) systems (Miberg et al., 1999, qtd. in Skraaning jr., 2003).

Social Organization

HAMMLAB is configured with three distinct work stations: two for process operators, and one for a shift supervisor, an arrangement representative of NPPs (Skraaning jr., 2003). Teamwork and communication studies typically use a full crew of three, while evaluation of more specialized operator support tools could use one of two stations.

Experimenters occupy a mirrored glass observation room behind the work stations, where process data collection, expert observations, and scenario control can be conducted simultaneously.

At full capacity, up to eleven people can be involved in the simulator's operation: Three participants, three eye-movement equipment operators (one for each participant), an experimental leader, two process experts, a system control technician, and a lab technician. An upper limit on participants is roughly 30, ten crews of three, and experiments can last seven days. A large experiment can therefore consume 287 person-days for data collection alone - a sizeable cost.

Interface

HAMMLAB's control room has 23 monitors – nine for each operator, and five for the shift supervisor station. A projection wall display screen is visible from every station. Different screen graphics are specific to the simulation used, and are often modified to include experimentally varied features (such as modified alarm displays).

Scenarios

The duration of scenarios is typically half to one and a half hours, a range that has been found by experienced HAMMLAB researchers to be efficient (Skraaning jr., 2003). Presumably shorter scenarios are burdensome in terms of fixed administrative costs, and longer scenarios provide diminishing returns for useful information about participant performance. Developing scenarios in HAMMLAB is costly for experimenters, since scenarios are specific to the research questions and enough need to be created so that they are not repeated by participants, due to learning effects reducing external validity. Because HAMMLAB requires certified NPP operators, participants often return for multiple experiments, so scenarios cannot even be reused between studies.

Scenarios typically begin with the simulated process in stereotypical initial conditions, e.g. running at full power, half power, or shutdown. Then, after preset times, faults of some kind will occur. Depending on research questions, these have included automation failures (as discussed above), equipment failures, large disturbances, or combinations of the above. The large scope and high engineering fidelity of the simulations allow a wide range of scenarios to be created, which increases the challenge of creating equal difficulty scenarios (the significance of which will be discussed further below). Scenario designers must also consider scenarios' dependence on experimental manipulations. For example, an experiment comparing two alarm management interfaces would be fairly scenario-independent since a wide range of scenarios could (and should) be used, while an experiment comparing characteristics of troublesome plant faults would be scenario-dependent since the faults in question are part of the scenarios (Skraaning jr., 2003). A research question can thus require generation of many new scenarios.

4.5.2 Development and Theoretical Justification

Development of HAMMLAB has taken place in several iterations, from inception in 1983 to the present day. In addition to the phasing in of the HAMBO and FRESH simulations (and deprecation of the original NORS), there have been several upgrades to simulation equipment, including computers and monitors. Details of the associated physical and software design decisions may be found in Halden Reactor Project technical reports, but are not accessible to the general public.

HAMMLAB has intentionally developed as high fidelity a simulator as possible while maintaining re-configurability. Emphasis is placed on functional over physical fidelity; hardware instruments and controls, for example, are reproduced on reconfigurable computer displays. Contrary to the premise that experimental control and fidelity must be traded in equal amounts (Brehmer & Dörner, 1993; Stanton, 1996), Halden researchers (Skraaning jr., 2003) assert that this relationship is non-proportional, and that if appropriate experimental methods are used, only a small decrease in fidelity is required to achieve acceptable levels of control. Following this hypothesis, experimental design methods and human performance evaluation tools, rather than simulator simplification, have been the focus in maintaining experimental control in high-fidelity experiments. Halden researchers agree that achieving high fidelity with respect to emotional responsibilities, communications, or long term adaptability to environments is difficult in any simulation (Hopkin, 1995, qtd. in Skraaning jr., 2003).

One aspect of experimental control that Halden researchers have attempted to address is the problem of 'task variance', the variation in participants' performance between different scenarios with the same experimental manipulations. Its magnitude in HAMMLAB experiments is roughly similar to that attributable to individual differences, presenting a data analysis problem. HAMMLAB's complex problem-solving research focus means that scenarios cannot be repeated to average out variability. Instead, many novel scenarios must be developed and categorized into difficulty ranges, a laborious reduction of tractability.

The means chosen to evaluate participants' performance vary depending on the research questions. Eye trackers are only occasionally used in HAMMLAB, as they require additional operations staff and produce volumes of data. Most experiments use a combination of simulator process logs, expert evaluators, and questionnaires delivered to participants (Skraaning jr., 2003). This variety of performance measures have been found to be required, because automation and the simulation's autonomous development tend to obscure human contributions to raw process-based performance measures (Skraaning jr., 2003). A key consideration HAMMLAB researchers make in choosing performance measures is their reactivity, meaning how much they distort what they purport to measure. For example, questionnaires that focus on part of the simulation can have cueing effects, directing participants' attention and resulting in inflated scores in later measurements.

4.5.3 Implications for a CWA approach

The experiences collected above are only a small portion of the challenges faced by HAMMLAB researchers over the past two and a half decades. While not comprehensive, they may contain potential opportunities for a CWA-based approach to support scenario design and performance measures.

Addressing variations in performance between scenarios, participants, and their interactions may be a challenge suited to an ecological approach such as CWA. This potential has been recognized by some HAMMLAB researchers such as Skraaning, who notes that "It is tempting to interpret [performance variation] findings in the light of … theoretical positions claiming that human behavior is always a function of person × situation interactions." (Skraaning jr., 2003, p. 28). Some HAMMLAB researchers have proposed examining scenarios for 'complexity factors' that can be used to design scenarios of equal difficulty, and work is ongoing in this area (Harbord, 1998 qtd. in Skraaning jr., 2003). Existing CWA tools may be adaptable for this purpose: as discussed in Section 3.6, Strategies Analysis has already been used to classify scenario difficulty in the DURESS II microworld. The other four phases of CWA may also provide guidance in design or classification of scenarios.

The experiences described above further emphasize the importance of performance metrics. As discussed in Section 3.6, WDA has been previously used for this purpose with some success, focusing on quantitative system-based indicators of performance. However, the experiences at HAMMLAB suggest that additional performance measures that evaluate the quality of problem-solving behavior are necessary. Subject matter experts (SME) can provide a valid and tractable 'catch-all', when their evaluations are guided with analysis-based templates (Skraaning jr., 2003). Social performance metrics are also required: communication, coordination and cooperation within participant teams can greatly influence performance, accounting for 20-40% of variation in one comprehensive empirical study (Rouse, Cannon-Bowers, & Salas, 1992, qtd. in Skraaning jr., 2003). As noted in C3FIRE and TRACON discussions, other simulator designers have recognized the opportunity to design simulators to make such analysis more tractable. A CWA-based approach may be useful for constraining SME evaluations, and both classifying and interpreting communication and cooperation data.

4.6 Summary

The challenges faced and lessons learned by researchers using the simulators discussed above indicate several simulator design choices that could be informed, described, and justified by use of a theory-based framework. We summarize these here and in Table 4, beginning with those found in lower physical fidelity microworlds.

Challenging dynamics are typical of many complex sociotechnical systems. In MCF, system dynamics are central to the research interest, while in C3FIRE and CAMS they are used to define classes of systems the microworld is intended to generalize to. A theoretical framework for simulator design must help designers justify choices of relevant dynamics and the *engineering fidelity* a simulation must possess to be representative.

The power of *simulator cover stories* provided to participants has been noted by C3FIRE and CAMS researchers. Microworld designers must develop descriptions (such as forest firefighting or life support control) that are both familiar to participants and consistent with the simulator's behavior in order to provide context for participants' understanding. No theoretical guidance has been provided for this task, presenting an opportunity for a simulator design framework.

Simulator interface choices that *constrain human behavior* are evidenced by the MCF simulator's discouraging skilled analog manual control behavior through restriction to discrete keystrokes, in C3FIRE by the forcing of typed rather than spoken communications, and to a lesser extent in TRACON's use of virtual flight strips rather than physical, manipulatable ones. While these choices increase tractability, the extent to which they harm correspondence has not been analyzed or theoretically justified during simulator design, only in after-the-fact studies.

Table 4 – Simulator design choices found in each example.

	Case Study Example				
Simulator Design Choices	MCF	C^{3}FIRE	CAMS	TRACON	HAMMLAB
Selecting engineering fidelity					
Devising cover stories					
Constraining human behavior					
Adapting social organization					
Justifying simulation scope					
Generating scenarios					
Finding performance measures					

Design possibilities for *social organization* include the constraint of inter-participant communication in C^{3}FIRE to enforce a desired authority structure, and the substitution of other air crews with computer actors in TRACON. Both of these design choices contribute greatly to simulation tractability, but can reduce correspondence by encouraging use of different problem-solving strategies. A theoretical framework that could both document and justify such trade-offs would be helpful for claims of external validity.

Inclusion or omission of elements in *simulator scope* is perhaps the strongest determinant of simulator correspondence. While the MCF task omits or holds constant all but a handful of elements, HAMMLAB's simulations correspond closely with operational NPPs, with associated operating costs. Even when elements are included, they may be loosely coupled to the rest of the simulation, such as in CAMS' tank measurement and false alarm acknowledgement tasks. While MCF, CAMS and HAMMLAB researchers have presented some theoretical justifications for design choices, systematically informing and documenting such decisions is an opportunity for simulator design frameworks.

Scenario generation becomes dramatically more complex from microworld to high fidelity simulator. While participants in C3FIRE repeat identical scenarios and CAMS provides only a relatively small number of potential events, researchers at HAMMLAB expend considerable effort (and cost) in generating novel scenarios of roughly equivalent difficulty. As suggested by research with TRACON, partially automated scenario generation software may help alleviate this difficulty, but requires theoretically-derived heuristics that distinguish difficulty factors inherent in the natural system and those arising from human performance limitations. CWA may provide a framework to determine such complexity factors.

Finally, the challenge common to researchers using all simulators discussed here is determining *performance measures* that are tractable, comprehensive, and valid. Narrow definitions of effectiveness can lead to misevaluation of participant behavior, such as in C3FIRE, as may attempts to evaluate responsibilities such as record-keeping and alarm management without links to other simulated purposes, as in CAMS. Theoretical analysis can contribute to determining performance measures, as in TRACON and HAMMLAB, but to be tractable, simulators must support such measures and ideally as part of the original design. A simulator design framework that could suggest comprehensive performance measures during the simulator design process would be useful.

Table 5 - Comparison of existing CWA applications and opportunities identified in case studies

<table>
<tr><th rowspan="3">Simulator Design Choices</th><th colspan="9">Phases of CWA</th></tr>
<tr><th colspan="5">WDA</th><th rowspan="2">CtrlTA</th><th rowspan="2">StrA</th><th rowspan="2">SOA</th><th rowspan="2">WCA</th></tr>
<tr><th>FP</th><th>AF</th><th>GF</th><th>PFn</th><th>PFo</th></tr>
<tr><td>Physical Fidelity</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Psychological means-ends Fidelity</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Functional Fidelity</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Performance Measures</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Scenario Generation</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Selecting engineering fidelity</td><td rowspan="5" colspan="9">?</td></tr>
<tr><td>Devising cover stories</td></tr>
<tr><td>Constraining human behavior</td></tr>
<tr><td>Adapting social organization</td></tr>
<tr><td>Justifying simulation scope</td></tr>
</table>

Table 5 shows the overlap between previous applications of CWA to simulator design (Table 3) and theoretical challenges identified in the preceding case studies (Table 4). While the two most commonly encountered design choices in this section, performance measures and scenario generation, have been previously addressed through CWA, the remaining items in Table 4 have not. In the following section, we will propose applications of the CWA framework to address a subset of these issues.

Chapter 5 CWA for Research Simulator Design

Motivated by theoretical and methodological weaknesses identified in both Chapter 3's review of current practice and Chapter 4's case studies, we will now outline a number of proposals for application of the CWA theoretical framework to simulator design. Application proposals will be linked to corresponding simulator design challenges, such as those compiled in Table 5. Following the scope limitations outlined in Section 2.3, this discussion will focus on design challenges associated with design of computer-mediated simulations based on complex sociotechnical systems with predominantly causal constraints.

This approach is theory-driven, and for this reason these proposals will be presented in the order of CWA's nested constraints (Vicente, 1999), beginning with the environmental constraints of WDA and proceeding to the cognitive constraints of Worker Competencies Analysis (WCA).

5.1 System Boundary

The first step of applying CWA to simulator design is the same as for any analysis: determining the extent of the system to be considered (Burns & Hajdukiewicz, 2004; Vicente, 1999). Large system boundaries increase analysts' workload, while narrow system boundaries can lead to missed opportunities for insight. Accepted guidelines for analysis are to look for boundaries in the system that have relatively weak interactions (Vicente, 1999), and if in doubt start with a broader system boundary, as it can be reduced later if appropriate (Burns & Hajdukiewicz, 2004). This philosophy seems appropriate for design of all classes of simulators. If designers later desire to reduce the scope of a simulator, an analysis of a broad range of the natural system can be used to systematically consider and document scope exclusions.

Another choice in determining boundaries for analysis is whether to apply CWA in a formative or descriptive manner (Sanderson & Naikar, 2000; Vicente, 1999). In a formative application, the analysis is intended to permit revolutionary design of how work *could* be performed. In this case, the system boundary would deliberately exclude existing instrumentation, displays, and controls so that they may be redesigned. When applying CWA in a descriptive manner, existing system work support tools, tasks, and social organization are included in the analysis. As the formative approach is one of CWA's primary distinctions, it is generally recommended. However, if designing a simulator to correspond strongly with a specific natural system (such as a high-fidelity simulator for training), applying CWA in a descriptive manner may be desired. Specific features of the plant and the workarounds (Vicente, 1999) that workers must be fluent in performing will likely be of interest and must be preserved. If designing simulators that are intended to correspond to a wider range of work environments, such as microworlds, the formative approach may prove more useful in generalizing beyond specific details of the natural system.

5.2 Work Domain Analysis

CWA's first phase, WDA, is intended to capture constraints that can describe the goal-directed behavior of workers in a natural system. These constraints range from the system's purposes to the physical condition of equipment and environmental conditions that limit the system's capabilities. We will assume that the reader is familiar with the analytical tools used for WDA, specifically the Abstraction-Decomposition Space (ADS) introduced in Section 3.2.3.

5.2.1 Abstraction levels for simulator description

The different terminology that analysts use to describe the natural system at each abstraction level of an ADS can be useful to communicate simulator specifications, and to categorize aspects of simulator design (Sanderson & Naikar, 2000).

Functional Purposes

Conflicting functional purposes are almost ubiquitous in complex sociotechnical systems. The resulting goal conflicts can be characteristic of complexity, and so should be preserved in simulations where possible.

Participant engagement can be discussed in terms of functional purposes as well. In natural systems, personal responsibility and accountability for consequences often motivate workers to align their actions with the system's purposes. Since such motivations cannot be simulated, researchers should consider linking functional purposes to motivations already held by participants. For example, monetary compensation is commonly accepted as motivation in laboratory task research but this approach can be expensive for simulators with larger numbers of participants or frequent use by fewer highly-trained participants. Translating the natural system's functional purposes into intelligible forms in the simulator, either incorporated into simulator displays as feedback, or as part of the cover story given to participants, may be another approach.

Abstract Functions / Values and Priorities

Elements at this layer can be considered for use as objective performance measures, as discussed in Section 3.6 (Naikar & Sanderson, 1999; Vicente, 1995). Also, if expert participants' knowledge-based behaviors (KBB) are to transfer from the natural system, designers must achieve high engineering fidelity of physical laws and intentional constraints identified at this layer.

Generalized / Purpose Related Functions

The language used at this layer is often used in routine system operations (Naikar & Sanderson, 1999). If a simulation is intended only to research routine, procedural behavior, this level of abstraction may be the most useful for simulator design. The economical, context-situated language used at this level of abstraction can be useful for generating cover stories to brief participants, and to describe scenarios.

Physical Functions

Since most computer-mediated simulations do not simulate the actual physical form of the natural system, this is often the most concrete description language common to system and simulation. This makes physical functions quite useful in determining functional fidelity, physical fidelity, and simulator scope, as functions are easily identified in the natural system, and easily specified in the simulation.

Physical Form

Descriptions at this level of abstraction can guide physical fidelity and simulator scope choices. Although this level is often neglected when designing computer-mediated simulations (Jamieson & Vicente, 2001; Vicente, 1999), descriptions of the concrete, perceivable form of the work domain can be essential if research questions requiring representative experiments are to be addressed (Goldstein, 2006). In computer-mediated natural systems, workers can often ignore the physical appearance and location of 'field' system elements[13], but in non-mediated systems such as driving or flight the wide range of physical cues and signals obtained from the physical form of the work environment are crucial to enabling natural expert behavior.

The choice of whether to include the natural system's existing interfaces in this description of physical form depends on the research question, the type of simulator, and the type of human behavior that will be of interest. If the research will be applied to a specific natural system, the simulator is to be high physical fidelity, and participant transfer of skill based behavior (SBB) is required, then the natural system's controls and displays should be included in Physical Form analysis. An aircraft cockpit simulator is an example of these characteristics; both the physical environment 'out the window' and the layout of aircraft controls should be reproduced for experienced pilots to behave naturally. If a laboratory task, microworld, or scaled world is to generalize more broadly and focus on rule and knowledge based behaviors, then letting the simulator's physical form be guided by functional requirements may improve tractability.

5.2.2 Abstraction-Decomposition Space for Simulation Scope

On its own, the ADS matrix has been used to illustrate differences in scope between simulations of the same natural system (Lee, 2001). To make simulator design decisions,

---

[13] Though not always, as shown by field observations of complex process plants (Mumaw et al., 2000)

however, the means-ends and part-whole structure of the system must be considered. This type of analysis seems to be hinted at by researchers using Cognitive Task Analysis methods: Gray refers to “paring away” (Gray, 2002, p. 208) parts of a system, Cooke to simplifying while preserving “functions that led to interdependencies” (N. J. Cooke et al., 2004, p. 271) , and Gugerty to “eliminating elements that don’t relate to tasks” (Gugerty et al., 2004, p. 257).

Means-ends links

Means-ends relations describe action opportunities and achievement dependencies between elements in the natural system, and may be useful for reducing simulator scope in microworlds. The shape of the hierarchy formed by means-ends relations may suggest key functions and couplings that must be preserved, or help justify the exclusion of loosely connected elements.

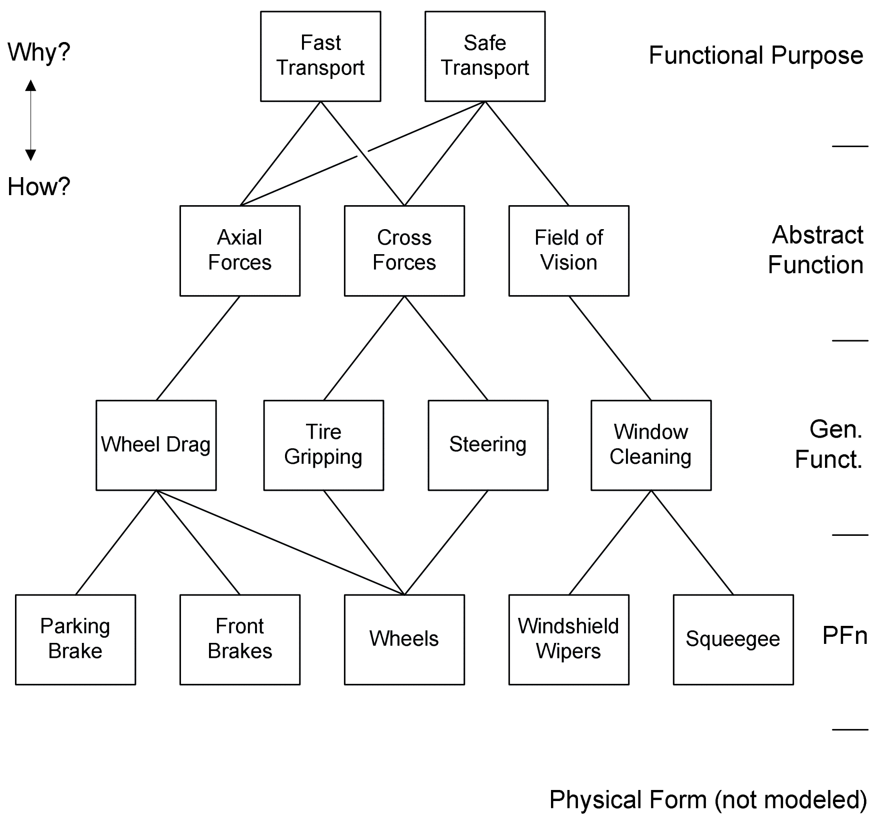


Figure 9 - ADS excerpt for automobile: Means-Ends links example

Proposing firm rules for such an analysis is difficult, but we will specify some hopefully useful heuristics. As an example for illustration, an abstraction hierarchy for an automobile is shown in Figure 9.

- Elements that have only one means-ends relation in vertical columns through three or more levels of abstraction may be more isolated aspects of the work domain that can be removed or simulated in less detail without widespread effects. Causal links to other system elements should be cross-checked. For example, Window Cleaning in Figure 9 is isolated from car handling processes and may be neglected.

- Elements that can be accomplished by numerous means may indicate that some of the supporting elements can be omitted to simplify the simulation. More caution may be warranted when considering elements with only two or three means. For example, wheel friction, the front brakes, and the parking brake can all contribute to car wheel drag in Figure 9. Eliminating the parking brake, however, would remove an important opportunity for adaptive problem solving: if the car's disc brakes fail on a hill, quickly applying the parking brake is a key expert behavior.
- Contrarily, if two means accomplish the same end, one used frequently and one as backup, and if the research question is not concerned with management of alternative resources, then such backup means can possibly be neglected. For example, if Window Cleaning was retained in Figure 9's simulation, the backup Squeegee could be neglected.

Complete means-ends hierarchies often form nests of interconnecting links that make examining individual elements difficult. Analysts may find transition diagrams[14] the most suitable representation for examining means-ends relations.

If analysts use this approach to prune either system elements, or means and ends from a simulation, they should consider the effects on the cover story provided to participants. The description of the simulation may need to be altered to make any changes to the intentional and causal structure of the natural system's work domain clear to participants and ensure that they do not attribute incorrect contextual properties to it.

Part-whole relations

Analysts may use part-whole relations to make decisions concerning the size of the simulation's functional fidelity, or scope/resolution (FISG, 1998). A part-whole analysis can be used in two ways, depending on the type of simplification desired:

- Amalgamation by part-whole relations, simulating the system only at an aggregated level of decomposition. This would reduce functional fidelity.
- Neglecting of parts, simulating the system with components and some omissions. This would reduce simulator scope.

Amalgamation can be thought of as incorporating many small components into 'black box' subsystems. Using amalgamation, a simulation's correspondence can be shifted from being specific to one particular instantiation of a natural system to include functional groups common to many systems. This approach is typified by the CAMS

[14] A transition diagram shows only the connections between two levels of abstraction at similar decomposition levels.

simulation discussed in Section 4.3, which simulates mostly subsystems, of which only one is decomposed into components.

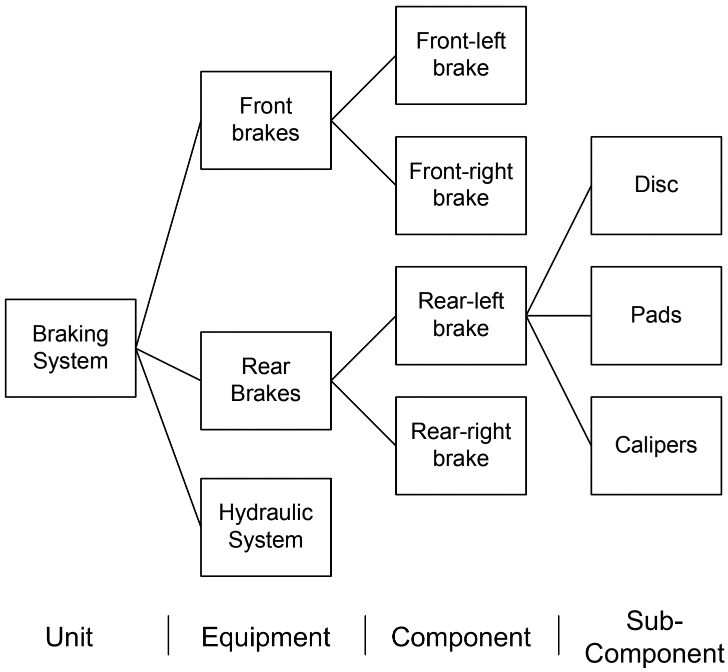


Figure 10 - ADS excerpt for automobile: Part-Whole Example

Another example, from Figure 10 could involve simulating an automobile's braking system at different levels of amalgamation: If brakes were amalgamated to the Component level, the simulation would not specify whether the car's brakes were drum brakes or disc brakes. If they were amalgamated at the Equipment level, they would be functionally equivalent to all car and even motorcycles' braking systems.

An example of neglecting subcomponents would be omitting duplicate 'standby' equipment in a process control simulation. Simulating at a fine level of decomposition, and choosing to neglect some subsystem components should maintain closer correspondence to the operational domain. The TRACON simulation follows this design principle, in allowing the number of aircraft to be varied, and retaining only a subset of the airspace restrictions present in natural ATC systems. Using Figure 10's example, a mechanic diagnosis simulation could neglect the brake caliper sub-component, which would reduce the difficulty of diagnosing brake faults.

However, analysts should check against means-ends relations when amalgamating or neglecting parts of a system. Using the example of a tactical engagement system, if units within a squadron serve different means (for example supply units and combat units), amalgamating them may affect higher-level balances that contribute to system complexity, such as the balance of protecting vulnerable supply units while projecting force with combat units.

Causal relations

Causal relations within levels of abstraction and decomposition show how system elements influence each other or are topographically linked. The networks formed by these causal links may also provide insight for simulator scope decisions. Some work domains, such as tactical unit control environments like C3FIRE and TRACON, tend to form radial networks where central units interlink surrounding system elements. Complexity in these environments emerges from the dynamic network of influences between elements. Analysts may attempt to simplify such systems by eliminating an outlying element of the network, for example wind effects in TRACON, or by part-whole aggregation of a more central element in the network, such as decreasing the number of aircraft.

Other domains, such as manufacturing, have more linear causal structures with processes sequentially influencing each other and occasionally feeding back to previous stages. Narrow choke points in a causal chain may indicate boundaries between loosely coupled sections of the natural system. If these areas of sparse connections are consistent across multiple levels of abstraction, they may be useful dividing points for generation of multiple smaller simulations.

Causal structures can also be used to characterize scenario difficulty. In systems with linear causal relations, perturbations often cascade along the causal chain. For some systems, perturbations at the end of the system are simpler to manage, because their ill effects are short-lived, can be tolerated, and do not spread to previous stages. In other systems, perturbations at the beginning of the system are simpler to manage, because later stages provide opportunities to mitigate ill effects. Whatever the details, the location of a scenario-induced disturbance along causal chains may be a useful comparator of scenario difficulty.

5.3 Control Task Analysis

Both ConTA in Work Domain Terms (Rasmussen et al., 1994) and Decision-Making Terms (Vicente, 1999) may be useful for simulator design. The former is concerned with common recurring states of the system and with common types of activities that must be performed, while the latter considers the sequence of cognitive processes and actions that can be used to accomplish said activities. We will discuss possible applications of each in turn.

5.3.1 Using Work-Domain Terms

ConTA in Work Domain Terms has two steps. The first is identifying prototypical work *situations*, or modes, of the work domain. A classic example for a process control plant could be “start-up, normal operation, disturbance management, and plant shutdown” (Vicente, 1999). For a hydroelectric power system, these could include the season (summer or winter), the amount of water inflow (high or low), and the grid electricity demand (peak or baseline) (Li, Sanderson, & Memisevic, 2006).

Since prototypical work situations emphasize different areas of the work domain, choosing a subset of situations to include in a simulator can suggest parts of the work domain that are outside their range and may be more safely omitted from the simulator scope. (N. J. Cooke et al., 2004; Gugerty et al., 2004; Loft et al., 2004)

The second step is identifying prototypical work *functions*, more overt action sequences that are clearly defined in time (Rasmussen et al., 1994). For a hydroelectric power system, these could include water inflow, water storing, power generating, water pumping, or water spilling (Memisevic, Sanderson, Choudhury, & Wong, 2005). Prototypical work functions can be mapped onto work situations, to indicate when functions *could* be performed (Naikar, Pearce, Drumm, & Sanderson, 2003). In hydroelectric domains, for example, the function of reversing turbines to pump water to reservoir storage would normally not occur during a peak electricity demand situations due to the financial implications (Li et al., 2006).

A less formalized version of this approach has been used to generate scenarios (Li et al., 2006), by mapping a collection of faults in work functions (e.g. a generator failure) onto work situations (e.g. in winter, with high reservoirs). This allowed a library of individual faults to be assembled by common work situation into engaging and tractable scenarios with plausible cover stories.

5.3.2 Using Decision-Making Terms

Prototypical work functions can be modeled as a series of information processing steps and knowledge states that describe actors' control of the system. ConTA in Decision-Making Terms uses the decision ladder representation (Rasmussen, 1974; Vicente, 1999) to represent both the normative sequence of information processing steps, and expert shortcuts such as information processing shunts or associative leaps between steps.

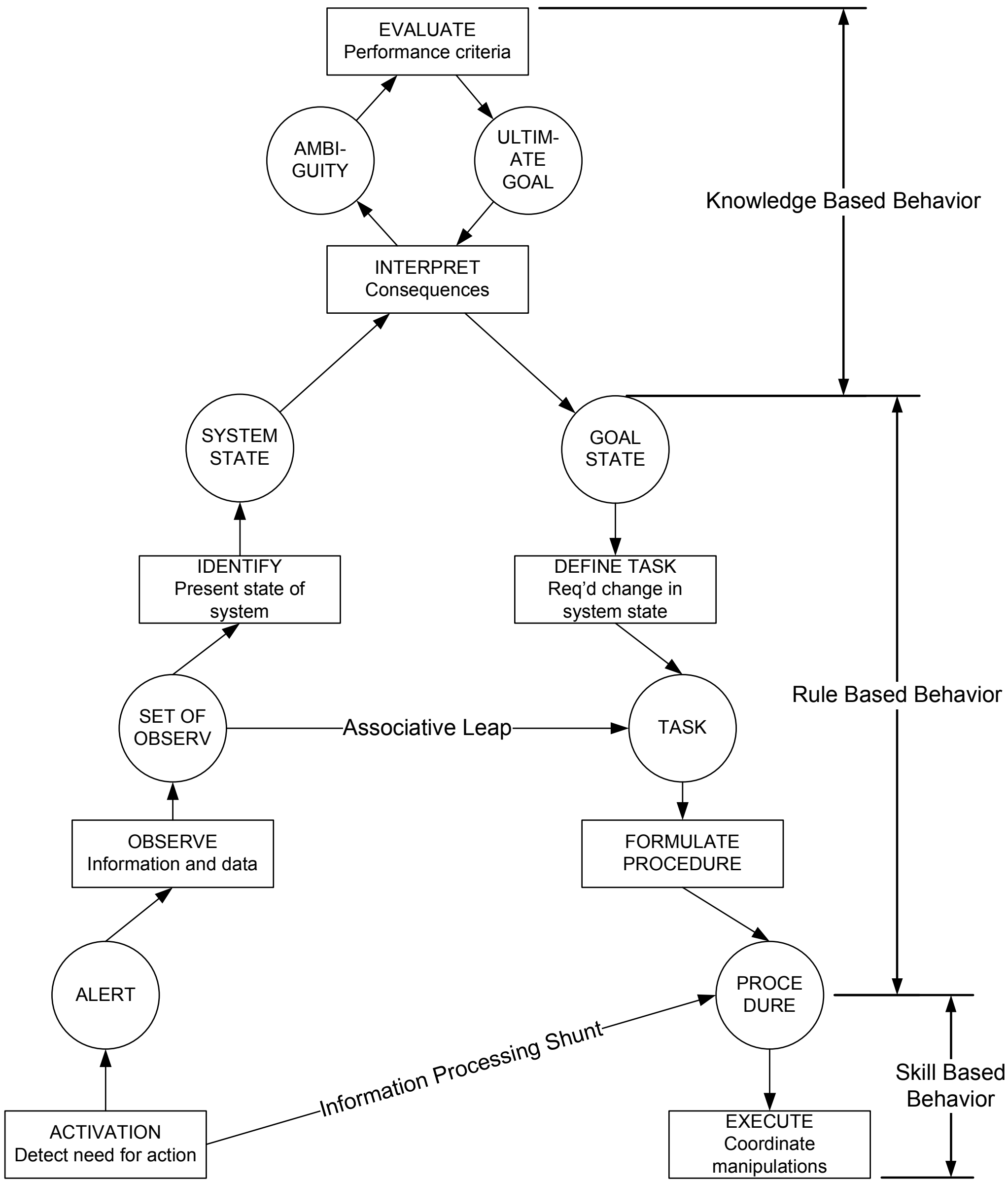


Figure 11 - ConTA Decision Ladder, showing levels of skill, rule, and knowledge based behavior

For high-fidelity simulations covering a wide range of a system's work domain, designers might find ConTA useful in evaluating simulator completeness, and in validating the complexity of participants' behavior. Comparing control task trajectories and average completion times between operators and participants might be one approach.

ConTA might also contribute to designing microworlds and scaled worlds when correspondence with a range of systems is desired. As discussed in Section 3.5, when performing a field study observing workers' activities in a natural system, analysts can expect to find activities that are fundamental to the natural system's control, and workaround activities required as a result of the natural system's existing work support tools (Vicente, 1999). Conducting a ConTA is intended to aid an analyst in distinguishing between the two. Designers may reduce complexity by neglecting workaround activities when designing a microworld or scaled world. Depending on the research questions, hopefully this will entail only an acceptable reduction in psychological fidelity. Some examples of workarounds could be industrial process control interfaces that require operators to enter authorization codes before making parameter changes, or forest firefighting situations where staff must manually maintain representations of the current state of firefighting units.

When designing more limited scope simulators for specific research purposes, ConTA may be useful in determining the work domain elements that are investigated or acted upon in the completion of a control task. This could allow the principled creation of laboratory tasks such as the MCF simulator discussed in Section 4.1.

Regardless of the fidelity of a simulator, ConTA could also be used to encourage or discourage certain classes of participant responses. The steps of a decision ladder are associated with layers of abstraction in the work domain, and categories of worker competencies (Vicente, 1999). For example, Activation and Execution, the initial and final information processing steps of a control task, form the base of the decision ladder. These describe actors' interactions with the work domain at the physical form level, which must be carried out using skill based behavior (SBB). At the top of the decision ladder, abstract considerations of ambiguity and ultimate goals are associated with consideration of the functional purposes of the system and use of knowledge-based behavior (KBB). The intervening steps of the decision ladder involve primarily rule-based behaviors (RBB).

Encouraging KBB

If a research question requires expert participants to exhibit KBB, a ConTA can be used to determine commonly occurring RBB shunts and leaps across the decision ladder. Then, to discourage such expert behavior, the simulator and scenarios could be designed so that 'what usually works' is discouraged and participants are more likely to consider ambiguity and conflicting goals. Some design interventions to increase ambiguity are:

- Decreasing the diagnosticity of simulated sensors, thus requiring participants to interpret the system state. Note that is not the same as omitting sensors, as that would be expected to preclude effective KBB (Jamieson & Vicente, 2005).
- Allowing an analytical interpretation of system state by designing high engineering fidelity in Abstract Functions such as first principles physical laws, and Value and Priority balances.
- Developing a wide range of scenarios with faults that are as novel as possible.
- Withholding procedure reference material from novice subjects
- Refraining from including automated control sequence features.

To induce conflict between purposes, scenario design may be the most effective intervention, if:

- The functional and engineering fidelity of the simulation implements tradeoffs, means-ends relations between Abstract Functions and multiple Functional Purposes
- Simulator performance measures reflect all abstract functions, so that participants are not tempted to neglect one functional purpose over another
- Scenarios and engineering fidelity allow sufficient time for slow, effortful KBB

Discouraging KBB

If research interests require KBB to be discouraged, or if a microworld is to be easily learned by novices, converse design decisions can be made.

Ambiguity about the simulated system's state can be reduced by incorporating super-sensors or an ecological interface can be designed (Burns & Hajdukiewicz, 2004). Designing super-sensors for simulations is usually easier than for natural systems, as the state of the simulated system is precisely knowable. The MCF task's use of the CSGL super-sensor discussed in Section 4.1 can be considered an exemplar of this approach.

Conflicting purposes can be reduced by

- Reducing the simulation scope to eliminate functional purposes (possibly through means-ends pruning described in Section 5.2)
- Generating scenarios and cover stories that de-emphasize Functional Purposes.
- Possibly introducing time pressure

The last suggestion in particular can be expected to distort human behavior, likely altering participants' cognitive strategies and introducing stress.

Encouraging SBB

Skill based behaviors are slowly learned and success in transferring expertise from a natural system to a simulation depends strongly on the engineering and physical fidelity of a simulator. A thorough ConTA should identify SBBs that can be supported in a simulation of the natural system. Because of their tight coupling to a wide range of cues and dynamics of the natural system, SBBs are generally difficult to perfectly support in a simulator.

Discouraging SBB

Discouraging SBB is much easier to do, at the cost of reducing correspondence with the natural system. One approach is to eliminate continuous spatiotemporal interactions from interfaces, for example by replacing physical continuous controls such as a joystick or trackball with more 'clunky' controls. These could be discrete pushbutton-type, such as used in the MCF task from Section 4.1, or more drastically, use of propositional forms such as command-line text entries or alphanumeric displays, as used in the original version of DESSY, discussed in Section 4.2. As an example, the TRACON scaled world (Section 4.4), uses simulated radar displays that discourage SBB by presenting data in slowly updated step intervals (Kilgore & St-Cyr, 2006).

5.4 Strategies Analysis

The information processing steps of a Control Task in Decision-Making Terms, such as identifying the state of the system, can often be achieved using different cognitive procedures, or strategies. Strategies analysis (StrA) of an information processing stage should identify a variety of information processing strategies that can be used to accomplish the step, as well as strengths and weaknesses of each strategy (Vicente, 1999).

Since experts adaptively switch between strategies, and because of the difficulty of eliciting internal mental processes from expert workers, applications of StrA in the literature are rare (Burns, Momtahan, & Enomoto, 2006). While an exhaustive StrA of the system is unlikely to be practical for most simulator design schedules and budgets, if a microworld or scaled world has been designed to focus on certain control tasks, it may be more apparent where analytical effort can be targeted.

Identify Strategies

The first potential benefit of a StrA to the representative design of simulators is identifying the range of strategies used in a natural system so as to build simulators to support them. If high correspondence is desired and expert participants are to be used, a simulator should support as many of the strategies observed in the natural system as possible. For example, the substitution of virtual aircraft tracking strips in TRACON likely eliminated some valid aircraft management strategies associated with their physical counterparts in the natural system.

Conversely, analysts should also attempt to determine if physical or engineering fidelity design choices will encourage the use of strategies that would be inappropriate in the natural system (Lintern & Naikar, 2000; Moray, 2004). Examples of this could be simplified problem diagnosis strategies in response to limited fault possibilities (as discussed in Section 3.1.2), or introduction of new sensori-motor control strategies that rely on defects of a simulator's computer display technology such as coarse pixilation.

Promote or Discourage Strategies

If strategies can be supported by a simulation, researchers and trainers may be interested in encouraging participants to switch strategies. Design of support tools to make strategies less effortful has been suggested as one way to accomplish this (Vicente, 1999). In a simulator, researchers can also encourage strategies by changing some of the work domain or task parameters that influence strategy choice. Examples include:

- Changing the frequency or type of faults, as certain strategies may be more fault-tolerant than others (Vicente, 1999)
- Altering the time or resource cost of making observations and actions, as these may influence shifts between 'technician' and 'engineering' approaches to problem solving (Rasmussen, 1974)
- Changing work domain complexity, such as by altering the number of aircraft in an air traffic control simulator such as TRACON (Sperandio 1978, qtd. in Vicente, 1999)

Scenario Sorting

Finally, as mentioned in Section 3.6 strategies can be used to generate equivalent difficulty microworld scenarios, or categorize more complex scenarios (Loft et al., 2004; Vicente, 1995, 1999).

5.5 Social Organization & Cooperation Analysis

All complex socio-technical systems involve some form of social organization or cooperation, whether between humans or between humans and computer actors. Often, the size of social networks involved in such systems makes duplicating social organizations impractical, even in high-fidelity simulators. A Social Organization & Cooperation Analysis (SOA) may help in adapting a natural system to a simulator's budgetary and tractability constraints.

It must be noted that the tools commonly associated with CWA address only a small range of social complexities, and that simulator designers should also consult the wider organizational theory literature. For example, while social organization can play a large part in participant engagement, CWA's analysis tools do not address motivational issues.

Social Organization in Work Domain Terms

WDA has been used to represent zones of responsibility assigned to workers, such as comparing the roles of members of hospital operating room (Hajdukiewicz, Vicente, Doyle, Milgram, & Burns, 2001). Such an analysis may be useful for simulator designers, either to match responsibilities for correspondence, or to modify them for tractability and research purposes. For example, in a simulation of an Unmanned Aerial Vehicle (UAV) task, overlaps between responsibilities of team members were eliminated through design of simulator workstations' information content. This then forced participants to cooperate and communicate, providing data to researchers (N. J. Cooke et al., 2004).

Social Organization in Control Task Terms

Designers can use ConTA representations to consider division of task responsibilities, for example between operators and computerized control systems (Rasmussen & Goodstein, 1987). Such an approach may also be useful for simulator design, especially if tractability requires large teams' members to be simulated by computer actors. A decision ladder representation of work distribution may be useful in designing plausible and effective computer actors. An application of this approach could be to extending the interactions possible from computer aircraft pilot actors in the TRACON simulator.

Communication Pattern Constraints

While the former two forms of SOA focus on the division of work, the frequency and content of communications between workers may be of equal interest to simulation designers. Natural systems differ in their management styles: some may have hierarchical organizational structures that require workers to report to and follow orders from management, others more distributed organizational structures that encourage communication between workers at the same level of authority (Vicente, 1999).

Simulation designers may be interested in manipulating management styles through simulator design choices. One approach shown by C3FIRE researchers in Section 4.2.2 involved imposing work structures by restricting communications between simulator team participants (Granlund et al., 2004).

Communication Method Constraints

A final application of SOA to simulator design could be to analyze the methods workers use for communication in the natural system (e.g. face-to-face, radio, or email) (Rasmussen et al., 1994). For design of high-fidelity simulators, such analysis may be most useful to ensure that all relevant communication methods are simulated.

For microworld and scaled world design, analysts may wish to design for higher tractability in communications data collection. Verbal protocol analysis, for example, is labor intensive. As shown by the C3FIRE microworld and TRACON scaled world in Section 4.4.2, replacing spoken communications with either typed messaging or other computer interfaces can greatly increase tractability and provide data for development of performance measures. Designers should be mindful that the added effort this could impose on communications could alter strategies preferred by participants.

5.6 Worker Competencies Analysis

While previous phases of CWA address simulator fidelity with respect to properties of the natural system such as equipment, purposes, and organization, a Worker Competencies Analysis (WCA) focuses on comparing the capabilities of workers and participants and the performance demands placed on them.

In psychological research, it is expected that researchers ensure participants are representative of the population to which results should generalize[15]. To determine the validity of such choices, researchers must argue that experimental participants are similar in relevant abilities to the target population, or that the experiment either corrects such differences or makes them irrelevant. A wide range of comparison criteria has been developed, but we will discuss only one possible approach, based on Rasmussen's Skills, Rules, and Knowledge (SRK) taxonomy of human performance.

WCA in Control Task Terms

In order to compare the demands placed on workers and participants, designers may perform an SRK inventory of portions of the natural system of interest (Kilgore & St-Cyr, 2006). This involves selecting a subset of control tasks from previous analysis, and determining how each information processing step can be carried out with respect to signals, signs or symbols (corresponding to skill, rule, and knowledge based behaviors respectively) (Rasmussen, 1983). Any information processing shunts or associative leaps identified in a ConTA should also be included.

With this inventory of human performance possibilities, analysts can consider for each combination of information processing step and human behavior type:

- The support provided by the natural and simulated systems
- The frequencies of use by workers and participants
- The frequency of *required* use by workers and participants (such as novel events that require KBB)
- The effectiveness of workers and participants (speed and/or reliability)

This range of questions, like the SRK inventory serve "as placeholders to motivate the generation of design concepts" (Kilgore & St-Cyr, 2006, p. 507). The first question would consider issues of simulator physical fidelity, the second and third issues of simulator scenario design, and the last participant training or selection requirements. The answers may also include a better understanding of the extents of a simulator's valid applicability. For example, such an analysis would prompt consideration of the skill-based expertise of aircraft pilots, the powerful associative rule-based expertise of experienced firefighters (Klein, 1993), or the freshly-acquired thermodynamics knowledge of 3rd-year engineering undergraduates. Large differences between participants' and workers' abilities may also indicate potential for loss of participant engagement due to frustration.

5.7 All CWA phases

The final simulator design technique that will be proposed is the only one not based on existing CWA design tools, but instead on a formative method employed during the

---

[15] Of course, undergraduate students are very often chosen for convenience's sake.

recent design of a hydroelectric system simulator intended to evaluate functional interfaces (Li et al., 2006).

Temporal Matching

The role of a natural system's temporal behavior is difficult to address in CWA. The phase of analysis intended to analyze physical processes and intentional constraints, WDA, is explicitly time-independent (Vicente, 1999), while later phases are more concerned with temporal coordination than absolute time scales.

This is not to suggest that a full modeling of system dynamics is necessary: such an effort is often impractical and more appropriate for a control engineer. However, the general time scale over which system and human processes act has been proposed as a useful measure by which to determine simulator scope for training (Goettl, 2006) and research purposes (Memisevic, Sanderson, Wong, Choudhury, & Li, 2007). Such researchers have proposed that reducing simulator complexity through selecting subranges of time scales in which actions are possible is a useful approach that can produce valid results.

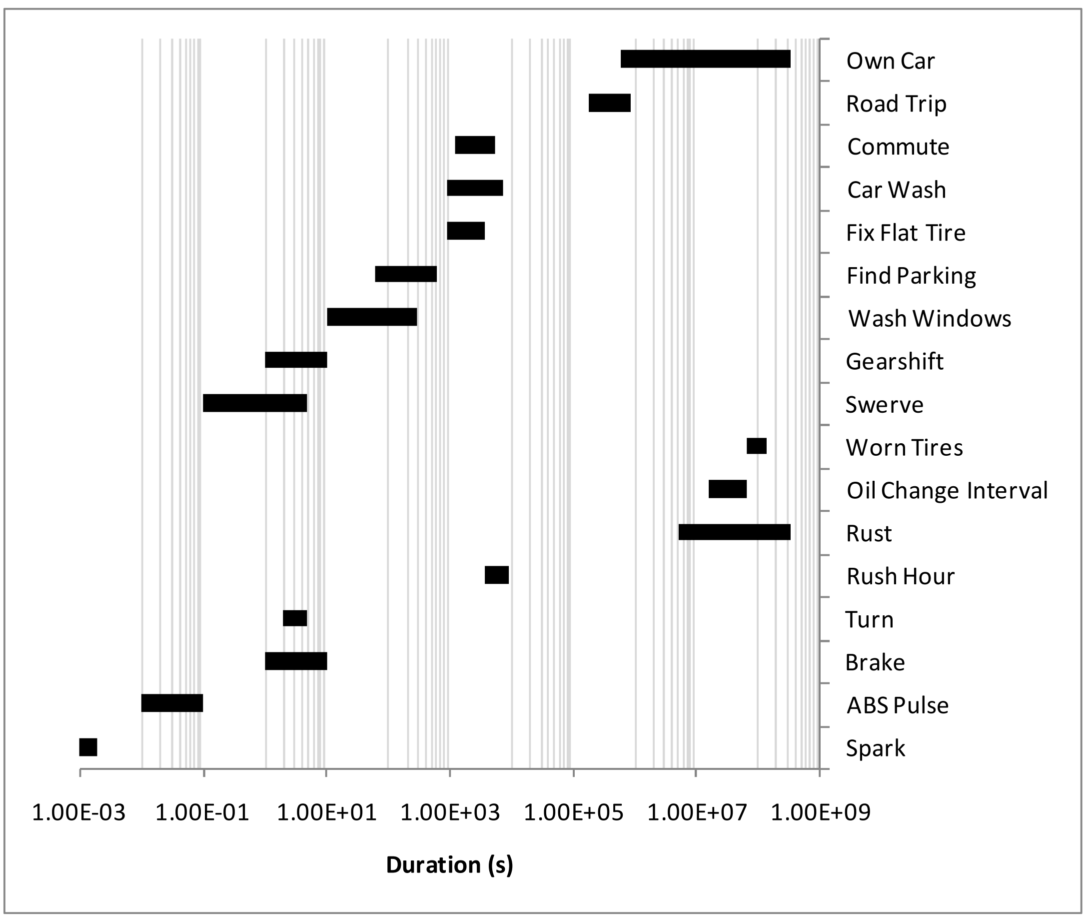


Figure 12 - Temporal Matching Chart for automobile Work Domain (bottom) and Control Tasks (top)

The method we propose is an extension of that previously used (Memisevic et al., 2007), to plot a temporal matching chart in which the time scales of elements from the following CWA stages are compared:

- System dynamics of elements at all appropriate abstraction levels
- Control task execution time or loop frequency
- Strategy processing time
- Social-Organizational communication time scales
- Worker ability ranges, such as for manual control

The comparison can be plotted using common time units on a logarithmic chart, as in the example Figure 12. In keeping with CWA's formative approach, a range of *possible* time ranges for each item should be estimated rather than purporting to determine specific values as in a timeline task analysis (Vicente, 1999). The temporal scale for these processes can range from millisecond-long electrical disturbances to very slow economic effects spanning many years (Memisevic et al., 2007). If simulator designers consider a time range window that contains the phenomena of experimental interest, such a tool may help in determining which processes can plausibly be included in a simulation or scenario (Li et al., 2006). For example, processes which occur over time scales longer than an allowable experimental session may be difficult to incorporate. Likewise, social communications such as contacting head office for approval may take place over a time scale of hours, and may be difficult to investigate in the same simulator as skill-based responses to colleagues' shouted warnings.

Choosing a range of time scales to simulate can simplify the development of acceptable engineering fidelity simulations, with an associated increase in tractability and reduction in cost. However, researchers must consider how to represent the effects of elements at time scales not included in a simulation. Tracing means-ends links may provide insight into elements of the work domain that can be substituted, or scenario designs that may plausibly explain their absence. Simulator cover stories may also need to be edited.

5.8 Selection of methods

The suggested applicability of CWA phases to simulator design choices is summarized by shaded cells in Table 6. This includes both previous work described in Chapter 3, and case studies discussed in Chapter 4.

While simulator scope and fidelity choices can be informed primarily by WDA, later analysis phases are useful for determining the need for physical and engineering fidelity. Performance measures draw primarily from the abstract function level of a WDA, and can incorporate elements at more concrete abstraction layers if they are diagnostic. Control Task completion times, strategy preferences, and inter-actor communications are also potentially useful as performance measures. Developing cover stories to explain microworld functionality to participants can draw from several phases: WDA functional purposes, to explain overall objectives; generalized functions to supply a familiar process-based language; CtrlTA in Work Domain Terms to explain the general situation of a scenario; CtrlTA in Decision-Making terms to give task instructions; and Temporal Matching to indicate neglected elements that should be explained away. Generation of scenarios is the design choice that draws from the widest range of CWA. While scenarios must incorporate an initial state of the simulation and a set of goals, depending on the research question they can also incorporate disturbances or faults, which may cause shifting of strategies, or require changes in social organization. Scenarios can be assembled based on temporal matching analysis.

Table 6 - Summary of potential CWA applications

	WDA							CtrlTA					
Simulator Design Choices	FP	AF	GF	PFn	PFo	Part-Whole	Causal	WD Terms	DM Terms	StrA	SOA	WCA	TM
Simulator Scope						■	■	■	■				■
Physical Fidelity				■	■				■	■		■	
Engineering Fidelity		■	■						■	■			■
Psychological means-ends Fidelity	■	■	■	■	■								
Functional Fidelity		■	■	■		■							
Performance Measures		■	■	■	■				■	■	■		
Socio-Organizational modification											■		
Constraining Behavior									■		■		
Developing Cover Stories	■		■					■	■				■
Participant Selection												■	
Participant Engagement	■										■	■	
Scenario Generation	■	■	■	■	■	■	■	■	■	■	■	■	■

Several caveats for the CWA applications proposed in this chapter must be noted. First, CWA is by no means a comprehensive modeling framework. Rather, it is intended to be a theoretically consistent framework to link analysis tools and constructs (Vicente, 1995, p. 246). Thus, the incorporation of design guidance from other methods is encouraged. Models of human performance in particular can be substituted in Worker Competencies analysis.

Second, consistent with the scope discussed in Section 2.3, the proposals focus on simulator and scenario design, not operating or data collection procedures, nor knowledge elicitation methods. Methodological suggestions for conducting CWA phases have been presented elsewhere (Naikar, Hopcroft, & Moylan, 2005).

Third, while these proposed design methods reference CWA representation tools (such as the ADS or Decision Ladder) to guide design, analysts should always be cautious to not confuse *models* of a system produced by CWA with the system itself. Or as quoted by Vicente, "don't eat the menu" (Golomb, 1968), and be sure to justify design choices in the context of the natural system.

Finally, the proposed design methods have not been validated, nor is it clear how they could be, considering the variety of research questions for which simulators are used. While CWA tools will help reduce the degrees of freedom in simulator design, they cannot uniquely specify a 'best' simulator for a given application. The most straightforward way to gauge the usefulness of these methods may be to apply them to the design of a simulator. We present such an example next, in Chapter 6.

Chapter 6 Application to HEPHAISTOS

To demonstrate some of the design heuristics introduced in the previous chapter, we will apply them to the HEPHAISTOS infrastructure introduced in Chapter 1. HEPHAISTOS requires a suitable simulation to be a useful research tool. Through industry partnerships, the CEL has acquired a petrochemical plant unit simulation. However, as the simulation was originally developed for operator procedural training, its suitability for cognitive engineering research topics has not been evaluated.

This chapter will describe the first steps in such an evaluation, beginning with a description of possible research goals, proceeding to an analysis of the complexities inherent in the natural petrochemical plant system, and concluding with some preliminary recommendations that can be used when modifying the simulation.

6.1 Simulator Research Purpose

The HEPHAISTOS simulator is shared between the University of Toronto and Lambton College, and will be used for different purposes. The CEL is primarily interested in generalizable research into cognitive engineering theory, while Lambton College is interested in more applied research into specific industries. The simulator will likely also be used to familiarize instrumentation & control student participants with the simulation to attempt to reduce learning effects in research studies. This means that while the main use of the simulator will be for research, training considerations may still be valuable.

Research Topics

Current research subjects that may be investigated with HEPHAISTOS include:

1) Trust in Automation:
   a) Can CWA integrate Purpose, Process, and Performance information about automation? (Lee & See, 2004)
   b) How can Purpose, Process and Performance be manipulated experimentally?
   c) Does team or shift work change trust in automation?
2) Development of CWA and associated design techniques such as EID:
   a) Can later steps of CWA be incorporated into EID, such as StrA or SOA?
   b) Can EID increase Situational Awareness? (Skraaning Jr. et al., 2007)
   c) How robust are EID interfaces to distortion or reliability of instrument data? (Kilgore & St-Cyr, 2006; Reising & Sanderson, 2002a)
   d) Which operator training methods work best with EID interfaces?
3) Process industry specific:
   a) General research in alarm design
   b) General research in procedure design

The simulator design process must consider the fidelity that a simulator will require to produce valid results for the above research questions.

Implications for Simulator Type

In this case, simulator experimental results and training are intended to generalize beyond the specific natural system that has been simulated, so correspondence need not be too high. This, combined with the restricted budgets and high student turn-over of academic institutions, suggest that a high-fidelity simulator is not necessary and might present financial and operational difficulty. However, these research topics mandate that the simulation have complex automation, teams of participants, and other elements typical of the process control industry that are too complex for a single microworld. Thus, the following analysis will focus on design suggestions for development of a scaled world simulation.

6.2 Natural System, Scope, and Limitations

The simulation that the CEL has acquired is based on a Diesel Hydro-Treater (DHT) petrochemical plant unit. As part of a refinery, a DHT removes sulfur compounds from diesel oil to meet low-smog emissions standards. Regulations on sulfur levels have been reduced by orders of magnitude in recent years, with the latest standard, Ultra Low Sulfur Diesel (ULSD) requiring less than five parts per million (ppm) sulfur.

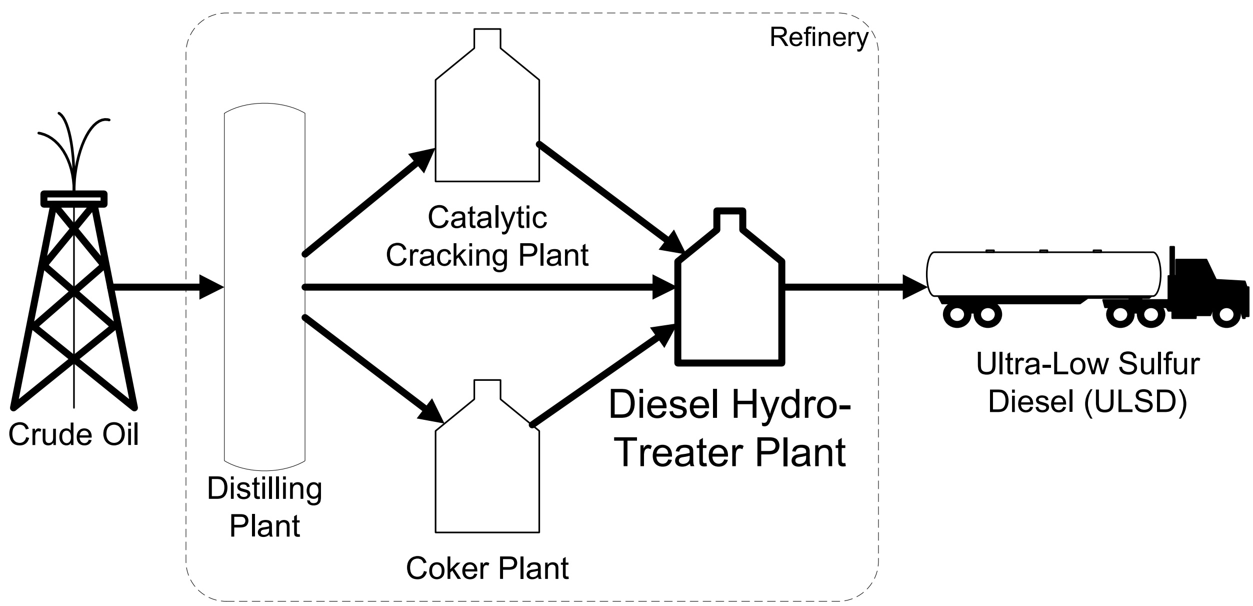


Figure 13 - Diesel Hydro-Treater System in refinery context

Previous work has included four applications of CWA to petrochemical applications. Two of these (Burns, Garrison, & Dinadis, 2003; C. A. Miller & Vicente, 1998a) have analyzed a natural system, while others were based on a process simulation (Jamieson, 2003; Jamieson & Vicente, 2001). The scarcity of precedents means that CWA modeling conventions for natural systems in the petrochemical domain have not yet been fully established, and analysts must consider sometimes conflicting rules of thumb.

Documentation

As discussed in Section 5.1, selecting a system boundary should be guided by areas of weak interaction in the natural system itself. The information sources used to perform this and further stages of analysis included process descriptions and terminology common to all DHT processes (Harwell, Thakkar, Polcar, Palmer, & Desai, 2007; Leffler, 2002), as well as some specific to the existing DHT simulation's natural system, including:

- Process Flow Diagrams
- Process & Instrumentation Diagrams (P&IDs)
- Critical and High-Priority alarm setpoints and recommended procedure overviews
- Simplified process descriptions

Documentation for the existing simulation was not available, and time constraints did not permit the reverse-engineering of its design rationale. Additionally, no site visits or interviews with workers were possible, and the only content review was by design consultants moderately familiar with the natural system.

Scope of Analysis

Due to the restrictions on natural system documentation, and time available to perform the analysis, a narrower analysis was chosen than is optimal. Ideally, a full CWA would be performed, including:

- WDA of the DHT, possibly including other closely coupled refinery plants (C. A. Miller & Vicente, 1998a)
- ConTA in Work Domain Terms to determine typical operating modes of the DHT
- ConTA in Decision-Making Terms of a number of Control Tasks, focusing on those involved in many of the plant's operating modes.
- StrA of Control Tasks, focusing on those involving interactions with the physical system, or manual control interventions
- SOA of DHT control room field workers' communication patterns and areas of responsibility
- Temporal Matching chart of Work Domain time delays and dynamics, typical durations of plant operating modes from ConTA in WD Terms, typical execution time for Control Tasks, and time required for various communication types.

Because of lack of access to system documentation or field observations, and time constraints beyond the control of the author, the extent of the analysis to date has been limited to a WDA of the DHT system in isolation. The extensive future work will be discussed in Section 6.5.

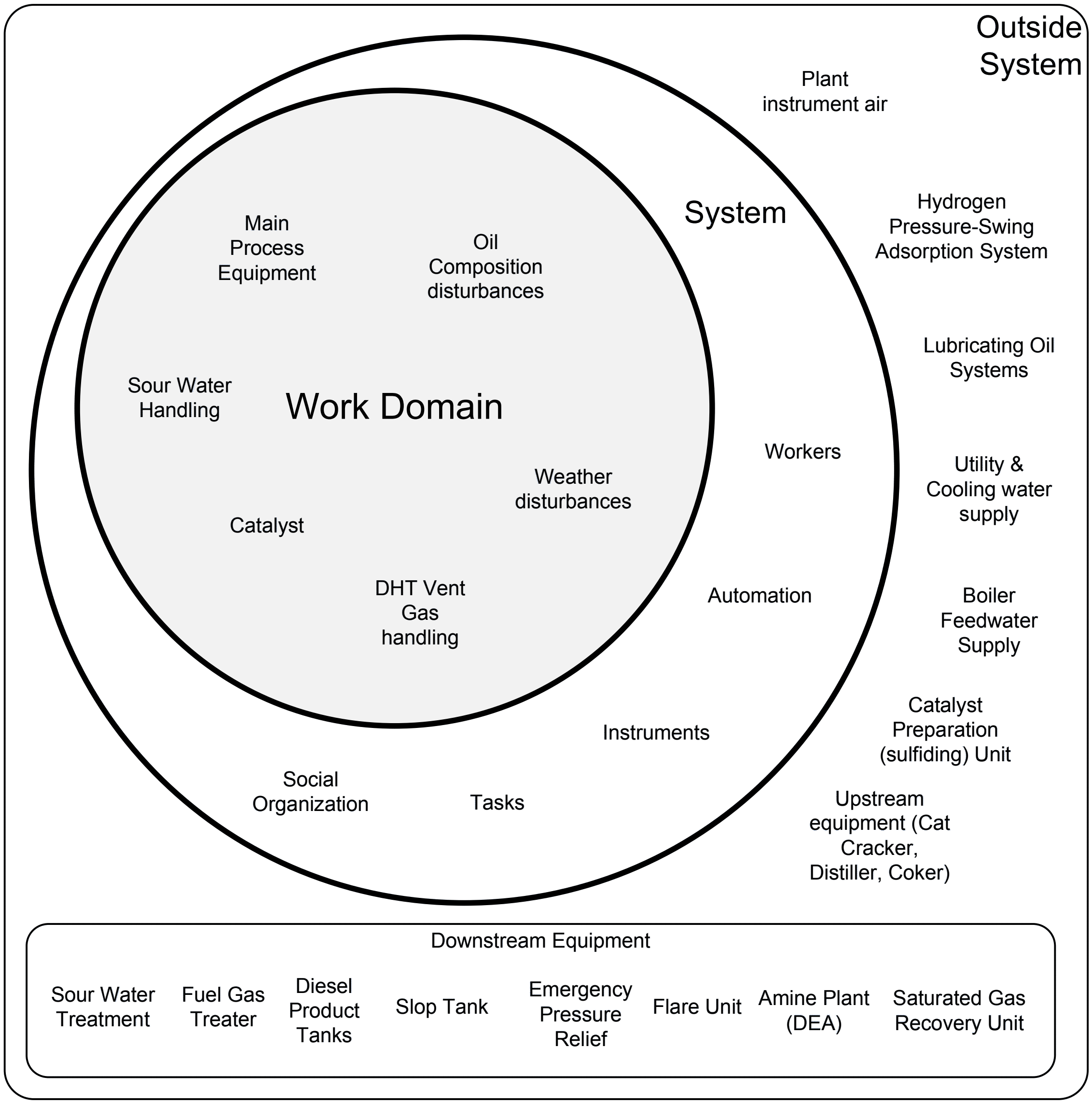


Figure 14 - Diesel Hydro-Treater System analysis boundary

The scope diagram in Figure 14 shows the elements that will be considered in the WDA. Items within Figure 14's outer circle such as workers, automation, and instrumentation are within the system boundary but are only included in later stages of analysis (Vicente 1999). Descriptions at the sides and bottom represent other refinery systems that have not been analyzed due to weaker interactions with the DHT, as well as lack of time and available documentation.

The DHT system will be explained in more detail as the WDA is presented, to illustrate how analysis can contribute to understanding of a natural system. Significant process interactions beyond the system boundary will be described on a case by case basis in the context of their influences on the DHT.

6.3 Work Domain Analysis

The WDA process resembled that described in the literature (Naikar et al., 2005) with the caveats discussed above in Section 6.2.

6.3.1 ADS

To construct an ADS framework for the DHT, analysts adopted the five abstraction levels typical of most process control systems (Burns & Hajdukiewicz, 2004), then determined how many levels of part-whole decomposition were needed to consistently describe system physical functions (Burns et al., 2003; Burns & Hajdukiewicz, 2004). Next, analysts determined which cells of the ADS to model. While this choice depends on the system and the research question, representations falling close to the System/Functional Purpose diagonal are usually most psychologically relevant (Vicente, 1999). Additional cells can later be modeled if deemed relevant.

Table 7 – Generic Abstraction-Decomposition Space, showing DHT analysis scope in bold

Level of Decomposition

Level of Abstraction	System	Unit	Sub-unit	Equipment	Component	Sub-component
Functional Purpose (FP)	Reason for system's existence					
Abstract Function (AF)	Physical laws governing system	Energy flow, storage, Mass transfor-mations, etc.				
Generalized Function (GF)		Summary of unit in process terms	Pressurizat-ion, flow, warmth, chemical rxns	Specific pressuriz-ations, flows, cooling, etc.		
Physical Function (PFn)		Summary of unit functions	Specific Sub-unit functions	Pump, vessel, control valve functionality	Standby equipment, check valve functionality	Maintenance fittings, block valves, etc.
Physical Form (PFo)	Condition, location, & appearance of system equipment.					

An ADS framework for the DHT process is shown in Table 7, with cells within the analysis scope shaded. Some elements in the table have been filled in to show what types of language are used to describe the DHT at different layers of abstraction and decomposition. We will first present two 'slices' through the ADS, the first illustrating part-whole decomposition at the Physical Function level, the second means-ends abstraction at the Unit level.

6.3.2 Part-whole Decomposition

This section will describe the Physical Functions of the system in increasing detail, covering the three bold cells at the bottom of Table 7. This will be accompanied by explanations of the DHT system's physical functionality.

Physical Function at Unit Level

The unit-level aggregation of the system's physical functions is shown in Figure 15, including topographical/causal links. The icon at the bottom right of a diagram indicates the ADS cell it represents and can be used for navigation in conjunction with Table 7.

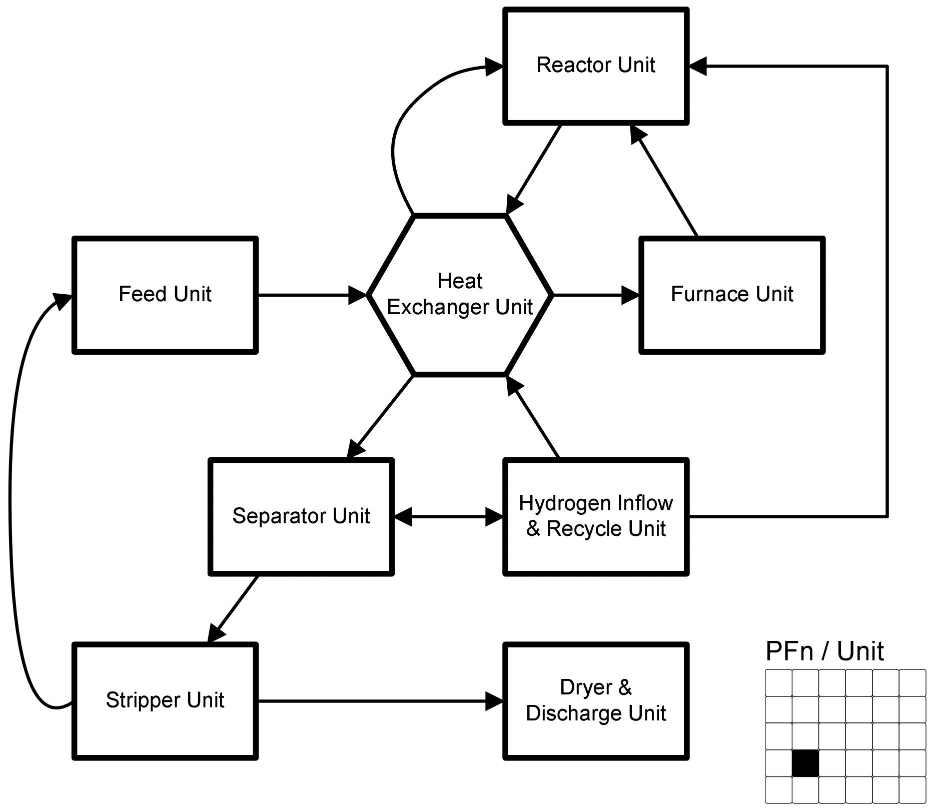


Figure 15 - DHT Physical Functions at Unit decomposition level

The feed unit, at the top left of Figure 15, introduces crude oil from several sources around the refinery, and combines it with recycled treated oil. Pumps move the oil through heat exchanger unit and the furnace unit, then into the reactor unit where the hot oil is combined with both pre-heated and cool hydrogen in the presence of a catalyst. The treated oil and waste gases continue through the other side of the heat exchanger unit, and to the separator unit that separates treated oil, gas, and water. Some of the gas is sent to the hydrogen inflow & recycle unit to be treated and reused in the reactor unit. The separated oil continues to the stripper unit, which removes further light gases from the oil. The diesel oil from the stripper unit is partially recycled to the feed unit, and the remainder sent to the dryer and discharge unit and to holding tanks.

Readers may note that this description of the DHT leaves out chemical, thermal and flow processes. This is deliberate, and can be contrasted with more abstract representations generated later in Section 6.3.3

Physical Function at Sub-Unit Level

The system units in Figure 15 can be broken into sub-units to give a more detailed representation of the DHT's functionality, shown in Figure 16.

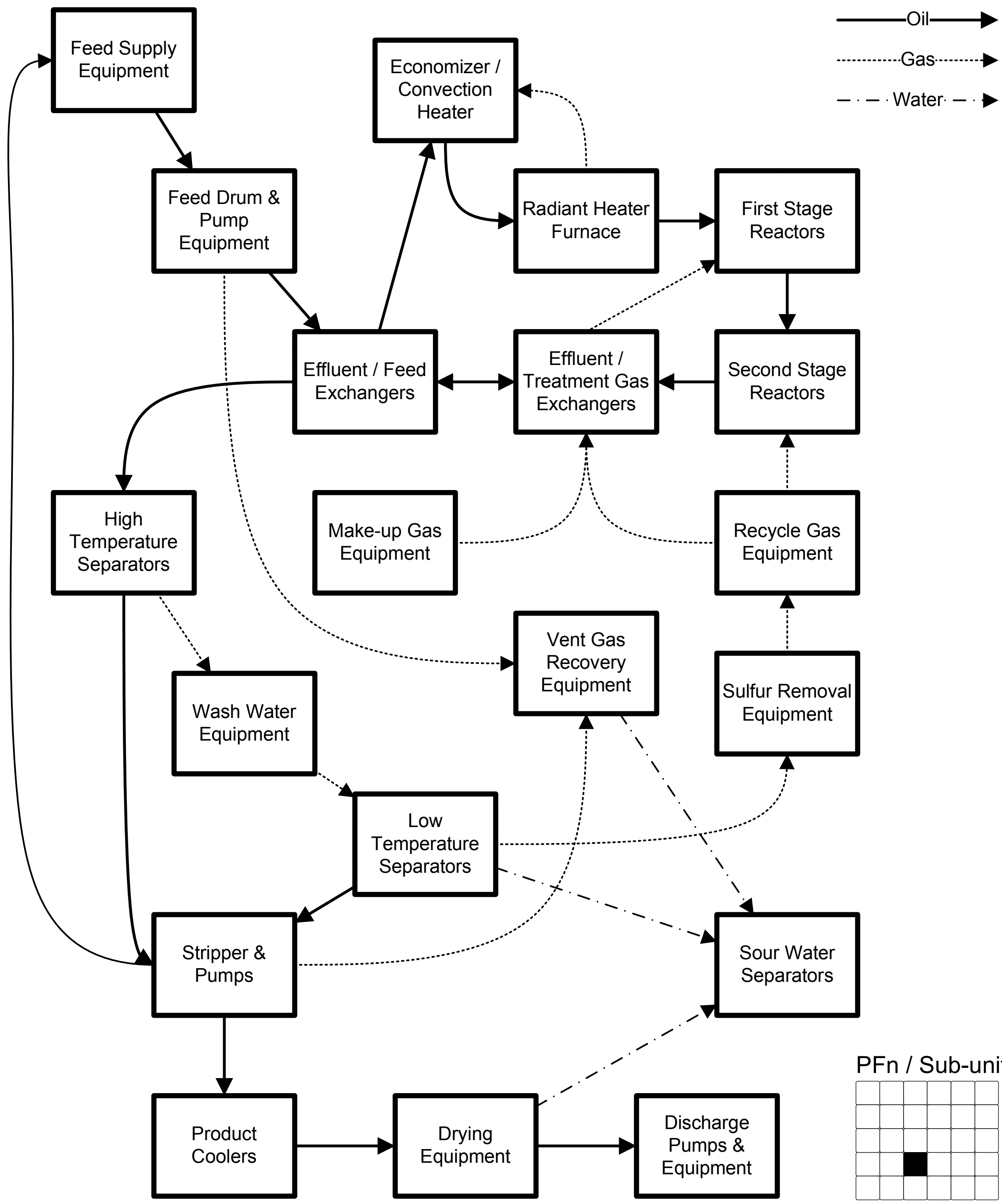


Figure 16 - DHT Physical Functions at Sub-unit Decomposition level

To improve legibility, Figure 16's causal links distinguish between connections involving oil, gas, and wash water. As before, the process begins from the upper-left quadrant, and causal connections follow the same general path. Elaborations at this more detailed level include:

- The main heat exchanger unit is made up of heat exchangers that heat the feed oil and those that heat the treatment gas.
- The furnace unit is decomposed into a convection and radiant heater. The hot combustion gases from the radiant furnace are ducted into the convection heater section.
- Both the first and second stage reactor sub-units are shown. The first stage reactors receive heated gas from the effluent/treated gas heat exchangers supplied by both make-up (fresh) and recycle gas units, while the second stage reactors receive cold recycle gas only.
- The diagram now shows the separator water washing equipment installed on the high temperature separators' gas outlet.
- Sub-units that handle oil, vent gas, and sour (sulfur-laden) water are shown separately, instead of being aggregated as in Figure 15.

Physical Function at Equipment Level

When process sub-units are further decomposed into equipment, they begin to be representable using standard engineering drawing vocabulary. Valves' and pumps' functionality, for example, have icons that will be familiar to industrial engineers. As plant units are broken apart into their components, the complexity of the representation required grows: full diagrams for the Equipment decomposition level are too large and have been detailed in Appendix A.

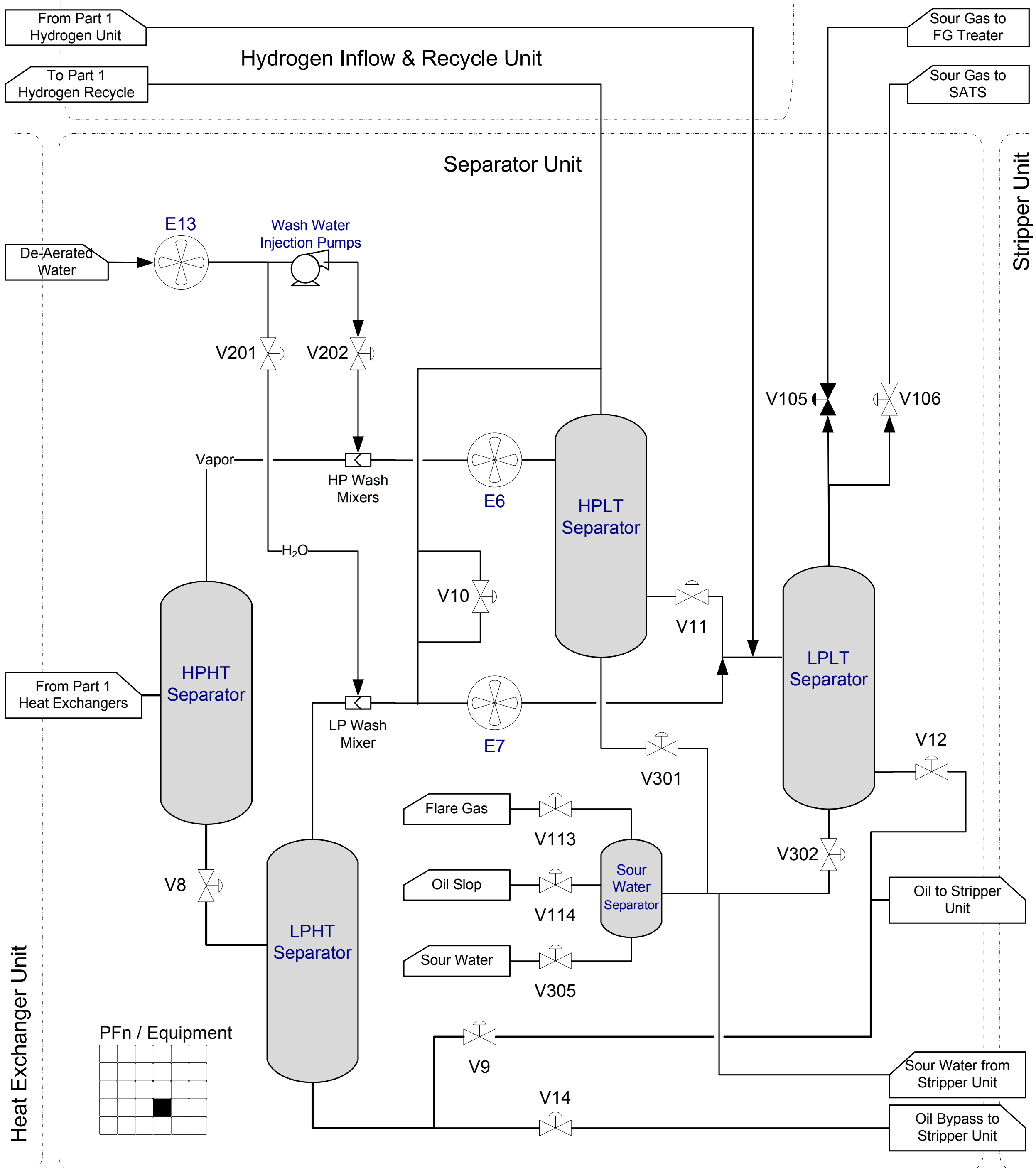


Figure 17 - DHT Separator Unit Physical Functions at Equipment level, with neighboring Unit boundaries

An excerpt of the separator unit is shown in Figure 17. The Equipment part-whole level resembles a process flow diagram (PFD) typical in industry, but with important distinctions. Like a PFD, the Equipment diagram does not show backup or standby equipment that duplicates the same function. Instead, they are combined into one icon to reduce the complexity of this representation. Part-whole decomposition applies this philosophy to every element; Figure 17's V202, for example, is made up of four control valves, splitting the wash water flow to four mixers and E6's four parallel air coolers.

Another distinction from a PFD representation is that the lines connecting equipment do not all have arrowheads to indicate the direction of normal process flow. This is deliberate: while check valves have not been included in the Equipment decomposition layer, the constraints they place on direction of flow are represented by arrowheads on Figure 17's causal links. The reason for this rigor is that during process upsets, the direction of flow in system pipes may stop or reverse unless prevented by check or control valves. Explicitly representing these possibilities helps analysts keep an open mind about constraints on system functionality.

Physical Function at Component level

A component decomposition level diagram is not included, due to the analysis scope. Its representation would span many pages. Table 7 shows some of the functionality that would be included at a component level of decomposition, and component-level decompositions are included in Appendix B. Continuing to a sub-component decomposition level produces representations resembling industry standard Process & Instrumentation Diagrams (P&ID). The sub-component level includes all manual valves, bypass lines, and pipe features, as well as tank components such as internal trays, weirs, and access ports. Since a P&ID level description of the process functionality spans 40 pages it has not been included, but an excerpt from of a full decomposition is provided in Figure 18.

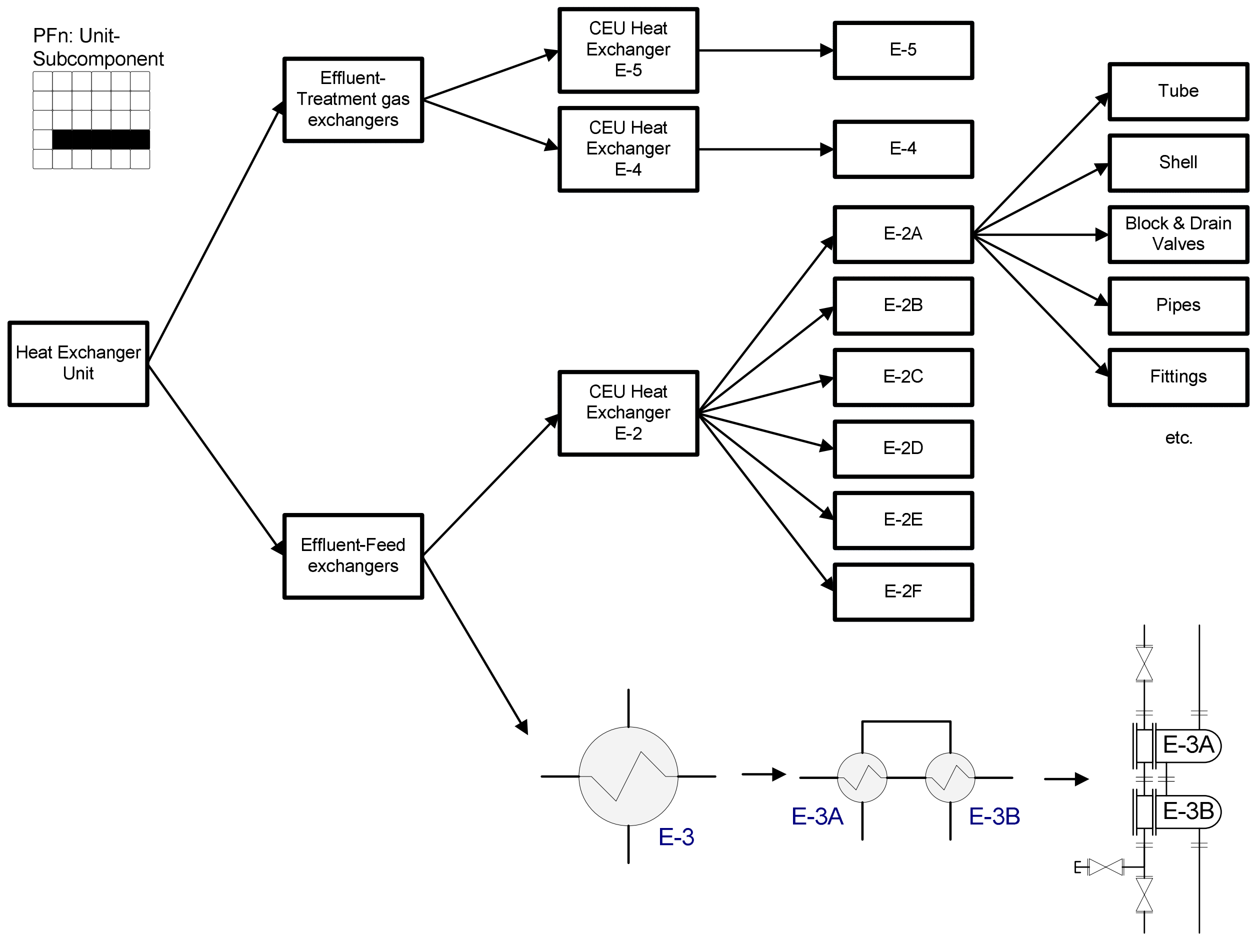


Figure 18 – DHT Heat Exchanger Physical Function showing Unit through Sub-component decomposition

On the left side of Figure 18 is the DHT's feed heat exchanger unit, as also shown in Figure 15. This is decomposed into heat exchangers that transfer between the reactor products and the feed oil, and others that heat the treatment gas, as shown in Figure 16. At the equipment decomposition level, the heat exchangers are further decomposed into groups of adjacent heat exchangers that share common process flows. These are labeled, such as E-3 (Exchanger 3), and can be represented with engineering icons as in Figure 17. With more detail at the component level, individual heat exchangers and their sequence of connections are modeled. For E-3, Figure 18 shows that the two components are connected in series. Finally, at the sub-component level, E-3A & B's manual isolation valves, cleaning ports, and internal functionality are represented. Additional component and sub-component decompositions of some of the DHT's Physical Functions are included in Appendix B.

By aggregating individual system elements into functionally relevant wholes, a part-whole decomposition mimics heuristics humans use to manage large problem spaces and complexity (see Section 3.2.3). Before discussing the possible implications of this analysis for simulator design, we will present examples of modeling using the vertical axis of the ADS: means-ends abstraction.

6.3.3 Means-Ends Abstraction

To describe the processes at work in the DHT, the physical function descriptions above can be abstracted by considering the purposes that each element serves in the system. For example, pumps typically exist to provide flow, or a heat exchanger to provide heating, cooling, or both. Such descriptions are more independent of the specifics of system equipment and generalize to a wider range of systems, making them of interest to simulation designers. We will present a vertical section through the ADS, beginning at the Abstract Function level in unit decomposition.

Abstract Function at Unit level

The Abstract Function language describes the system in terms of universal laws that govern system behavior. For process control systems, it is common to distinguish between mass and energy and describe each using conservation of mass and the first law of thermodynamics respectively. These constraints can allow deep analytical insight into the system's functioning, as is required by system designers and workers when troubleshooting an unfamiliar problem.

Just as the physical function equipment of Figure 17 could be represented using engineering iconography, the Abstract Function level can represented using visual shorthand adapted from the Multilevel Flow Modeling framework (Lind (1994), qtd. in C. A. Miller & Vicente, 1998a). We will use only a small excerpt of the extensive MFM language, summarized in Figure 19.

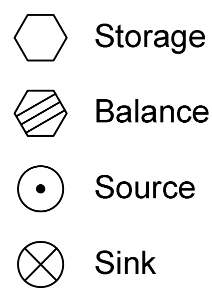


Figure 19 - Multilevel Flow Modeling Legend for Abstract Function diagrams

Those familiar with MFM notation will note that two of the MFM icons, transport and barrier, are not used. This deviates somewhat from previous WDA practice (Jamieson, 2003; C. A. Miller & Vicente, 1998a), but allows for a more compact representation structure.[16] Diagrams for mass and energy are shown in Figure 20 and Figure 21, respectively. In the Abstract Function-Mass diagram on Figure 20, transport elements are inherent in the linking arrows, and passive barriers (such as pipes or tanks) are omitted if they do not affect the system under normal (leak-free) conditions. Beginning

[16] In previous work, for example, heat exchangers are represented in both mass and energy diagrams as a barrier and balance respectively. This captures the functionality of a heat exchanger (transferring energy and separating mass), but can result in cumbersome diagrams. Likewise formal MFM requires transport functions to be defined for most links, doubling the size of the diagram for large interconnected systems.

from the top left, a source and storage represent the oil feed inflow, and the surge tank. Since the main heat exchangers and feed heater do not influence the mass of the feed oil, they are omitted from this diagram.

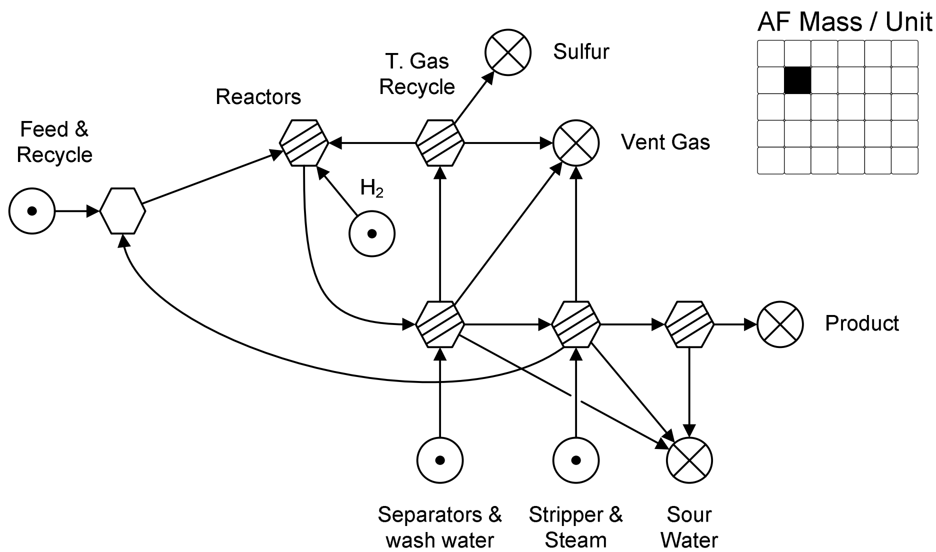


Figure 20 - DHT Abstract Function at Unit level. Conservation of Mass relations shown.

Continuing, the reactors at the top left of Figure 20 are shown as a balance of mass. Oil feed, hydrogen and recycle gas flows undergo chemical reactions, but their total mass is conserved. After leaving the reactor, the mass flow is separated in several stages, some to treatment gas recycling, and the rest through separators and the stripper. Mass leaves the system through sinks for sulfur gas, vent gas, sour water, and treated product. This diagram shows that, when in steady state, inflows by weight of oil, hydrogen, water, and steam must equal outflows of sulfur compound, vent gases, sour water, and finished diesel product. If this constraint is broken then either some storage is occurring, or there is a leak in the DHT plant. In petrochemical systems, especially those involving sulfurous gases, any leaks are extremely hazardous. Sub-unit representations of mass flows might distinguish between mass sources and balances of chemical compounds (such as sulfur) that relate to the system's functional purpose.

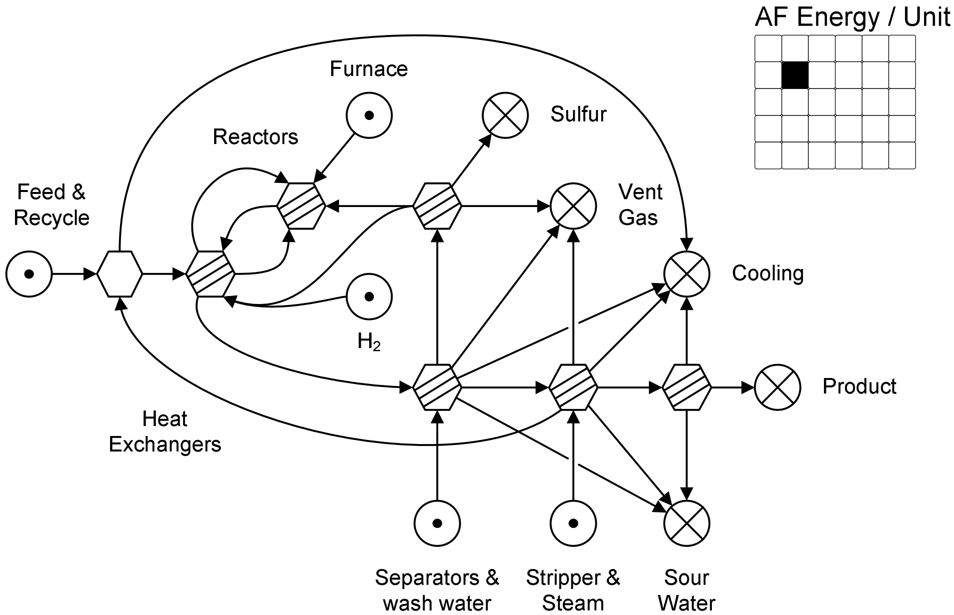


Figure 21 - DHT Abstract Function at Unit level. Conservation of Energy relations shown.

Because fluid flows have both mass and energy, Figure 20's mass elements also appear on the energy diagram in Figure 21. While the feed source is unchanged, the feed storage has an additional flow, losing energy by way of an air cooling energy sink.[17] Similarly, the feed heat exchangers' energy balance has been added between the feed storage and the reactor. The feed furnace is modeled as a simple energy source, adding heat energy to the reactor. With the exception of connections to the air cooling energy sink, the remainder of the energy flows are parallel to Figure 20's mass flows.

Energy flows within the system can help to explain some of the plant's behaviors. For example, the reactor balance shows that there are four energy inputs to the reactors (hot oil feed, furnace energy, hot treatment gas and cool treatment gas), but only one outlet, the energy carried with the reactor outflow's mass. This suggests that cooling the reactor may be an operational challenge, as mass and energy outflow from the reactor may introduce oscillations that will upset downstream balances. Figure 21 also illustrates how energy removal at each balance is associated with removing progressively more gas and water from the treated product. While small energy leaks such as an un-insulated pipe are not as potentially dangerous as mass leaks, over time they can contribute to inefficiencies within the plant.

Physical Function to Abstract Function, at Unit level

The links within the diagrams shown so far have represented topological or causal relations inherent to the system. More psychologically relevant for workers are opportunities to apply physical functions to manipulate the system's abstract

[17] At finer part-whole decomposition, a transport icon would be included to represent cooling equipment's control over energy flow.

properties, captured in means-ends relationships *across* these diagrams. To explore the DHT's means-ends relations we will first show a vertical slice through the ADS at unit-level decomposition. This will include elements from the Physical Function (Figure 15) and Abstract Function (Figure 20 & Figure 21) levels, with a new Generalized Function representation in between.

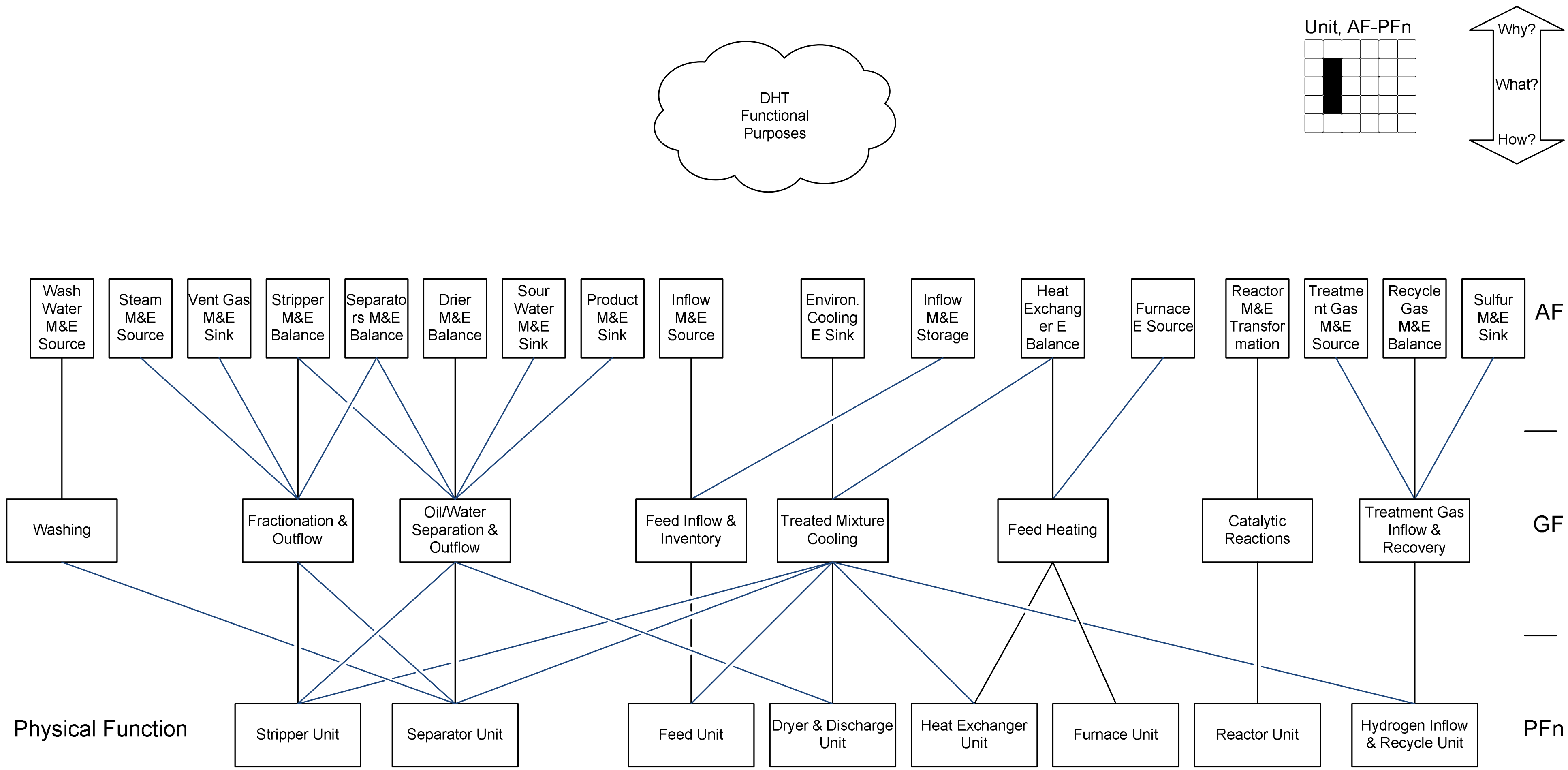


Figure 22 - DHT Unit level showing means-ends relations from Abstract Function to Physical Function levels

As before, the language used in the upper two layers of Figure 19 differs from the physical function language used in Section 6.3.2. For the Abstract Function layer, textual descriptions of mass and energy replace the MFM notation. At the Generalized Function level, the DHT is described in terms of processes, including flow, heating, separation, and reactions. Adjacent levels are linked upward with ‘why’ relations, and downward with ‘how’ relations. For example, the reason why the dryer and discharge unit is included in the system is to perform oil / water separation and disposal. The reason why the system performs oil and water separation is to produce ULSD product, and remove sulfurous ‘sour’ water from the system.

The means-ends links between levels reveal some features of the equipment units. As shown at the Physical Function to Generalized Function transition of Figure 19, many physical function units incorporate a cooling function, but only two are capable of heating the feed oil. Similarly, while the stripper, separator, and dryer units can separate oil and water, only the separators and stripper can separate, or ‘fractionate’ oil and lighter gases. More complexities should become apparent at finer levels of detail.

Physical Function to Generalized Function, at Sub-unit level

While the process language introduced for the Generalized Function layer in Figure 19 is useful in describing the DHT, the unit aggregation renders it vague. Sub-unit decomposition, shown in Figure 16 at the Physical Function layer, seems to provide a good compromise between explanatory power and representation complexity. Below in Figure 20, we show a transition between the Physical and Generalized Function for the sub-unit level.

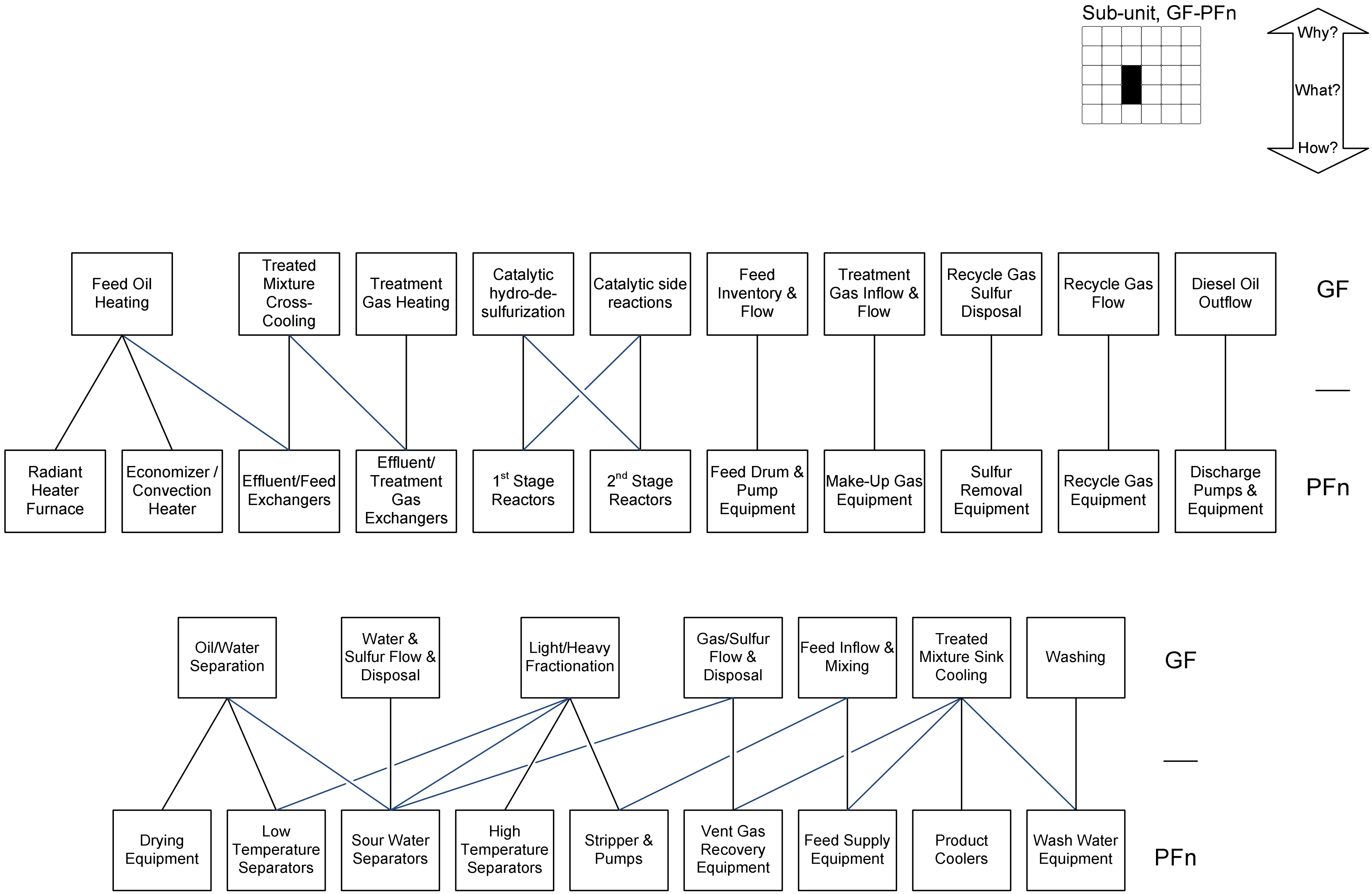


Figure 23 – DHT Sub-unit level showing means-ends relations between Generalized Function and Physical Function levels (split for clarity)

Interconnecting means-ends relations have grouped some functions in Figure 20. Heating, cooling, and removal of heat, gas, and water from the diesel fuel can be accomplished by several different equipment functions, while other processes are associated with only one physical function each. Un-grouped functions and generalized functions with only one means can highlight potential functional vulnerabilities in the system. For example, if the sour water separators fail, water cannot be removed from the system, and the process must be halted. If the high temperature separators malfunction, however, the process may be capable of continuing at reduced capacity since the low temperature separators, sour water separators, and stripper are all capable of fractionating light gases from heavy oils. While not definitive, such clues can prompt analysts to systematically examine unusual situations that may require infrequently used action possibilities.

Generalized Function at Sub-unit level

Finally, a representation of Figure 20’s Generalized Function layer at the sub-unit decomposition will be presented. This fills the gap between the analysis presented for the Abstract Function at unit level (Figure 20, Figure 21) and the Physical Function at equipment level (Figure 17, Appendix A).

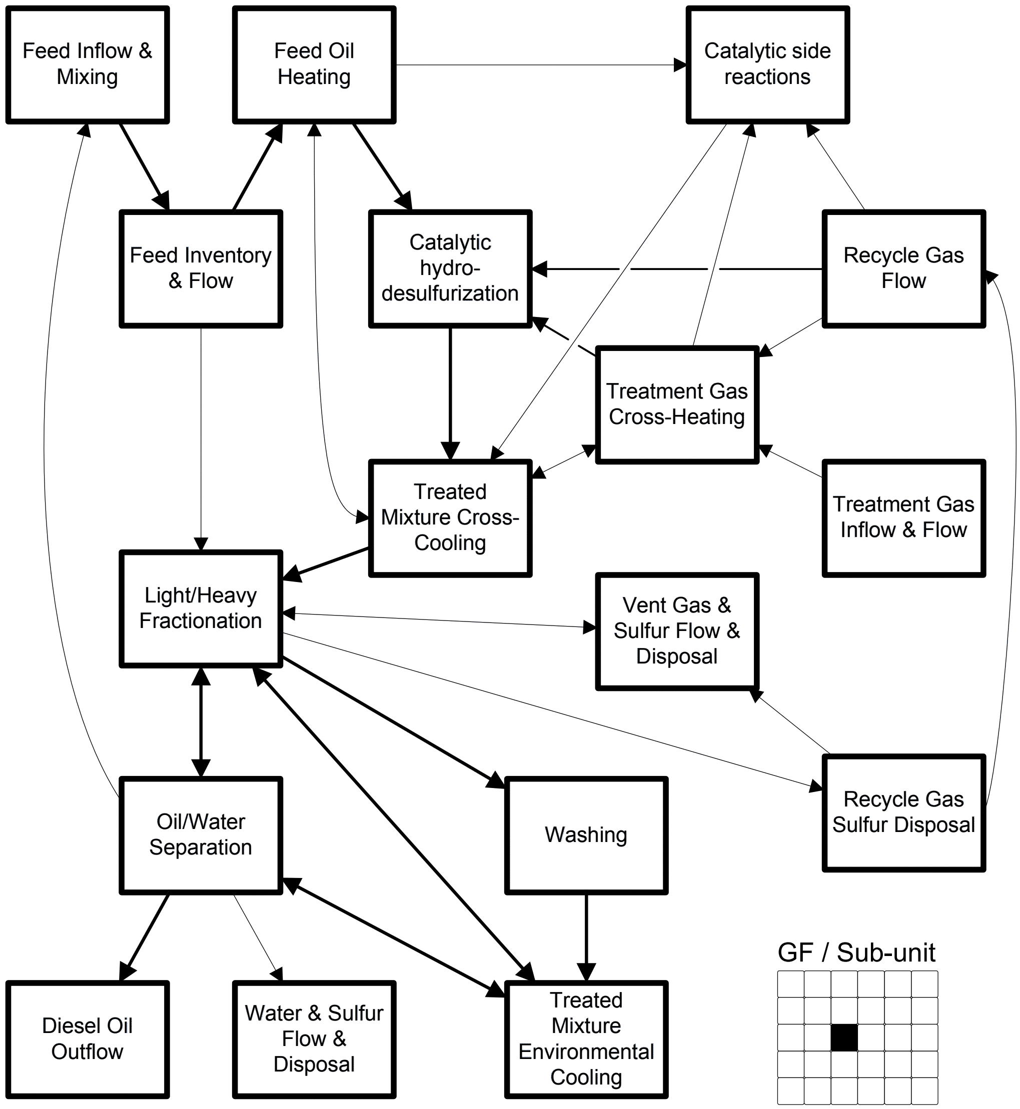


Figure 24 - DHT Generalized Function at sub-unit level

The language used in Figure 24 complements that used to introduce the system in Section 6.3.2. The largest difference from the Physical Function subunit description of Figure 16 is the network of causal paths surrounding Light/Heavy Fractionation in the lower-left quadrant. This corresponds to the separation, washing, and cooling group in the lower rung of Figure 20.

6.3.4 Detailed ADS

Functional Purposes are the final cells in the ADS that this analysis will cover. As in other petrochemical plant units, the overall purpose of the DHT is to "contribute to the financial viability of the installation as a whole" (Jamieson & Vicente, 2001, p. 1063). This functional purpose can be decomposed into sources of profitability, and specific gains and losses (Memisevic et al., 2005). Such elements can include saleable products, and costs such as energy, feed materials, maintenance, and salaries. Meeting regulatory standards on ULSD product quality is a financial decision as well: higher sulfur concentrations are permissible in non-motor vehicle fuel and in international markets. Arguably, elements such as safety and adherence to environmental regulations can also be considered as components of profitability. These components of the functional purpose were not modeled in detail, as their dependence on DHT abstract functions were not clear from the limited documentation available. A summary of the analysis is provided below in Table 8, as an expanded version of Table 7.

Table 8 – Diesel Hydrotreater Abstraction-Decomposition Space

Level of Abstraction

Level of Decomposition

	System	Unit	Sub-unit	Equipment	Component	Sub-component
Functional Purpose (FP)	Contribute to the profitability of the refinery as a whole.	Revenue from standards-compliant ULSD product, safety costs, regulatory penalties, etc…				
Abstract Function (AF)	Whole-system 'black box' energy accounting	Unit Mass & Energy sources, sinks, storage, and balances.	Sulfur & oil mass & energy sources, sulfur mass sinks, subunit energy balances…			
Generalized Function (GF)		Inflow, feed heating, catalytic reactions, gas recovery, cooling, separating…	Feed inventory & flow, hydro-desulfurization, side catalytic reactions, gas heating…	Hydro-de-sulfurization, de-nitrification, cracking, etc. temps, pressures, heavy oil flow...	Pressures, flows, temperatures for components, chemical reactions in components, etc..	Individual reactor bed temperatures, pressure gradient…
Physical Function (PFn)		Feed, heat exchanger, furnace, reactor, separator, stripper, dryer, hydrogen units…	High & Low temp separators, 1st stage, 2nd stage reactors, make-up & recycle gas, dryer…	Feed Pumps, V102, V301, E-3, E-6, Make-up gas compressor 1 & 2, economizer, salt dryer,surge drum	Standby sour water pumps, E-3A-F, Manual & check valves, reactor housing, reactor catalyst	Stripper trays, reactor catalyst in bed 1, bed 2… Pump seals, motors, bearings, exchanger shells..
Physical Form (PFo)	Aerial photograph of facility	Smell of air in equipment halls	Rumbling sound from recycle gas equipment	Appearance of flare tower 1-A	Rusty spot on standby pump mount	Crack in pump sealing ring

6.4 Simulator implications

Below, we discuss implications for simulator design choices, and support them where possible with findings from the WDA. The choice of which simulator design choices to address was based on the proposed extent of WDA applicability from Table 6.

6.4.1 Functional Fidelity

The first application of this analysis will be to guiding choices concerning the functional fidelity of a scaled world. To generate a scaled world simulation, some of the DHT's complexity must be simplified. Reducing functional fidelity through Part-whole amalgamation may be a useful approach.

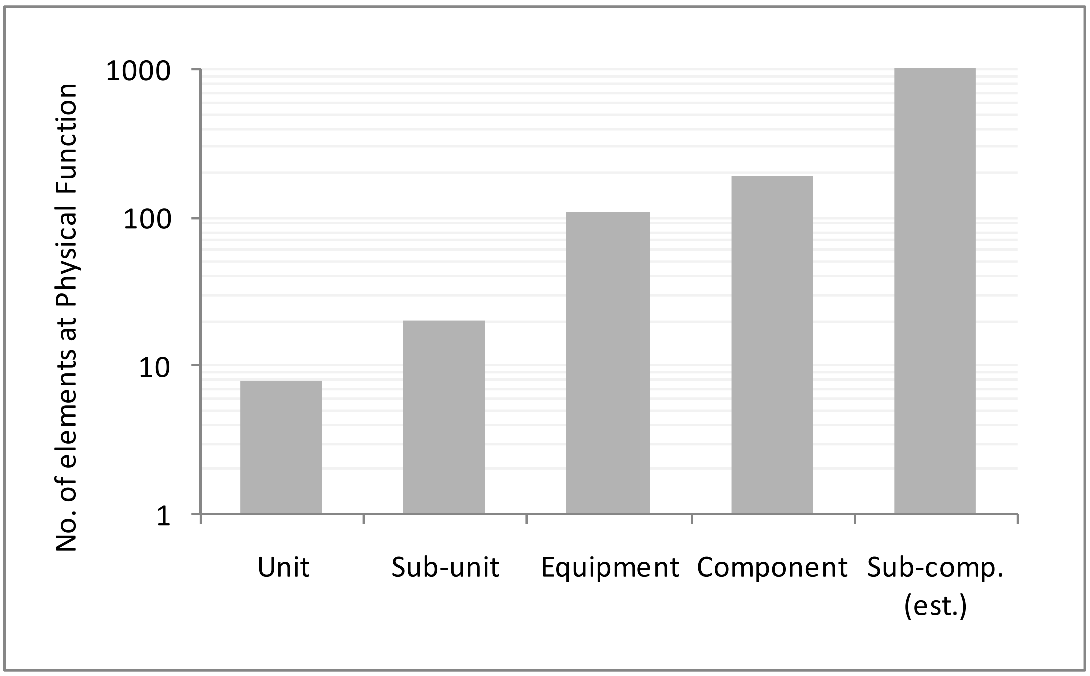


Figure 25 - Number of elements at the Physical Function level of each decomposition dimension

The increase in the number of physical functions at finer decomposition levels is illustrated in Figure 25, based on the diagrams presented in Section 6.3.2. When subunit functions are decomposed into control valves and process vessel equipment, the number of functions quintuples to over 100. With the addition of non-control valves, heat exchangers' parts, and major process vessel features, the component level representation contains almost 200 functions. Subcomponent level decomposition has been estimated in Figure 25 at roughly 1000. While the exact number of elements at each decomposition level depends in part on analysts' choices, the change between layers can be in orders of magnitude.

If trying to identify aggregated functional units in a natural system, their boundaries can seem vague. When used to specify a simulator, however, functional entities are more concrete: they are the smallest units at which action is possible or from which information can be obtained (Vicente 1999). For example, a microworld such as CAMS includes roughly nine functions[18], comparable to the DHT's Subunit level. While a real life support system's CO_2 scrubber unit would have many subfunctions, in CAMS it is simulated as an indivisible functional unit that can either be functioning, or ineffective (Sauer et al., 2000). Illusions of finer decomposition can be attempted by describing functional faults in terms of components, but participants will likely find such distinctions meaningless if they have no ability to diagnose or influence such components.

One strategy to determine appropriate functional fidelity would be to consider "how much can participants handle". CWA would suggest that this question depends on intervening characteristics of the natural system that would be considered in later stages of analysis, such as:

- Number and complexity of control tasks, related to functional and engineering fidelity respectively
- Available strategies, determined in part by the interface provided
- Social organization, including the division of work between teams of participants and automation.
- Worker competencies, such as level of participant expertise

Thus, the answer is "it depends", and further analysis of the DHT can help justify such a decision. However, a rough estimate can be made based on existing analysis by comparisons to other research simulators. Comparators include two non-automated, single-participant process control microworlds: DURESS II, with eleven physical functions (Vicente, 1999) and Pasteurizer, with twenty five (Reising & Sanderson, 2002b). Comparisons should consider that the HEPHAISTOS simulation will differ in two relevant respects: First, it is intended to be used by teams of two participants who will be of above-average competency (but not domain experts, either). Second, the HEPHAISTOS research focus suggests that extensive automation will be included. Thus, a *very* rough estimate could begin with Pasteurizer's twenty five functions, double it for two participants, double again for participant expertise, and again for division of tasks to automation. This extremely unscientific estimate corresponds to roughly the same number of functions as in the DHT's Component level of decomposition.

To preserve means-ends psychological fidelity, adjacent levels of abstraction should be simulated at equivalent levels of part-whole decomposition. For example, if the DHT simulation were to use an Equipment level of decomposition, it would contain physical functions diagrammed in Figure 17. For these elements' usefulness to be represented, the simulation would also need to incorporate corresponding generalized functions

[18] Air dehumidifier, air cooler, crew members, CO_2 scrubber, air vent, O_2 valve, backup O_2 valve, N_2 valve, mixing valve.

from Table 8 such as control valve influences on pressurizations, flows, and temperatures, as well as processes such as evaporation, condensation, and multiple catalytic reactions.

6.4.2 Simulator Scope

Instead of amalgamating functions to simplify a simulator, scope could be topographically reduced to focus on a portion of the DHT. A drastic example of this approach is the MCF laboratory task of Section 4.1, which was reduced to only four physical functions[19]. While many such choices have already been made up-front by the analysts by excluding DHT components from the system boundary (see Figure 14), further reductions can be performed in a more principled manner.

One potentially suitable dividing line can be identified from the PFn, GF, and AF causal/topographical diagrams shown in Figure 16, Figure 20, Figure 21 and Figure 24. A boundary drawn between the treated product heat exchangers and the separators divides the system roughly in half, while minimizing the number of causal links that cross the boundary.

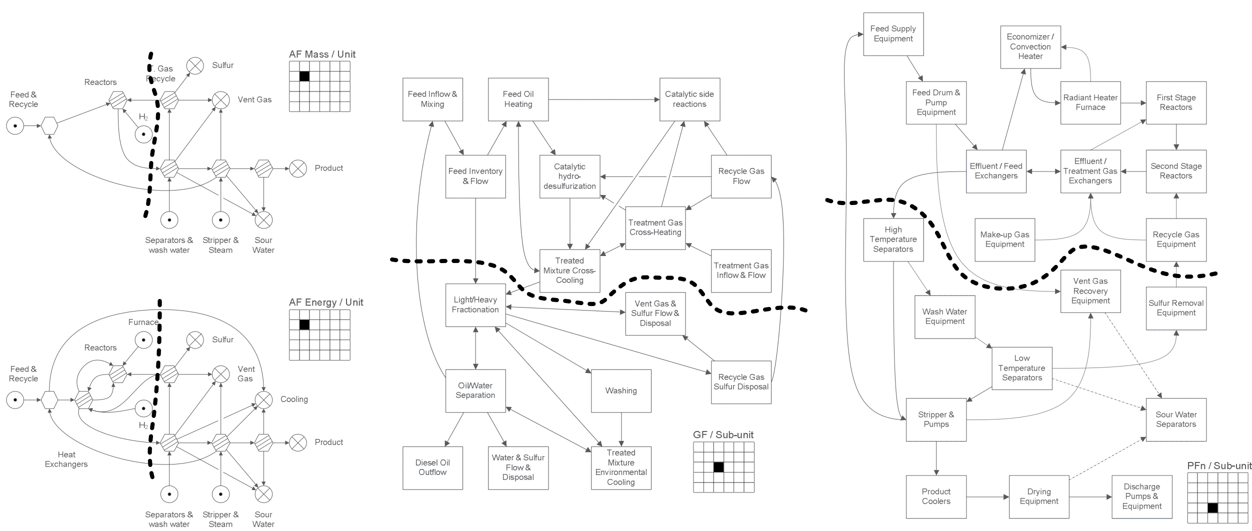


Figure 26 - Proposed simulation scope division line, at multiple layers of abstraction

Researchers could split the simulation into two parts as shown in Figure 26, either to allow single-participant trials, or if the whole simulation is too difficult for teams of participants to manage.

6.4.3 Engineering Fidelity (and dynamics)

Since the research applications of this simulator will be focused on participant KBB such as problem-solving and supervisory control, abstract and generalized functions such as those listed in Table 8 should be implemented with at least moderate engineering fidelity. Simple first-order lag responses are not sufficient.

---

[19] Boiler feed valve, steam outflow valve, boiler, and heater

If control challenges posed by engineering dynamics are to be compared between simulators, or a scope reduction's implications considered, the abstract and generalized function diagrams may be useful. For example, as discussed in Section 6.3.3 and shown in the right of Figure 26, the DHT system's energy and mass flows have qualitatively different properties in the sections upstream of the separators and downstream of the heat exchangers. The upstream section is characterized by linear mass flows and recirculating energy flows, reminiscent of control challenges previously explored in microworld research such as Pasteurizer (Reising & Sanderson, 2000), while the downstream half is associated with progressive energy and mass removals to sink. Catalytic reactions at the generalized function level were not modeled beyond the sub-unit level in this analysis, although they are also expected to greatly contribute to system complexity.

6.4.4 Physical Fidelity

Because the DHT is a computer-mediated natural system, it may sometimes be appropriate to include existing interfaces in a WDA, as discussed in Section 5.2. In this case it was not, since the HEPHAISTOS research program will likely involve the experimental variation of interfaces.

The DHT's physical form is relevant to simulator design, but without site visits, specific relevant features are difficult to identify. Examples of physical form are shown in Table 8. Some functional failures such as a hydrogen sulfide gas leak are detectable at the physical form level by humans' sense of smell[20]. Another physical form representation that might be especially diagnostic is flame from flare towers, and possibly its size and color.

The level of physical fidelity constrains the design of scenarios: if scenarios include faults that will likely be detected or diagnosed in the natural system through field observations, such as a minor sulfide gas leak, then in order to be representative physical form must be considered for inclusion. One way to do this would be video monitors using simulated stock footage, or communications with a simulated on-site team member.

Another factor contributing to the importance of physical form is the degree to which workers and participants are familiar with the DHT's appearance. If DHT control room workers have spent enough time as field operators, they may be able to integrate functional and physical knowledge to solve problems. For example, if workers know that two malfunctioning components from different functional groups are physically located side by side, they might hypothesize that something is wrong in their shared physical environment, e.g. a cooling fan has failed and the equipment space is

[20] Hydrogen Sulfide is associated with the smell of rotten eggs, and can be detected by the human nose at concentrations of 0.3 ppm. At higher concentrations, desensitization and very toxic effects can occur, so gas monitoring sensors are an essential part of refinery safety systems (Leffler, 2002).

overheating. It is difficult to evaluate the extent to which participants can become as familiar with the DHT's physical layout as expert workers.

6.4.5 Psychological means-ends Fidelity

Despite the narrow analysis scope, some opportunities for exploration of alternate means-ends relations can be seen in the DHT. The multiple means to light/heavy oil fractionation and oil/water separation shown in Figure 20 may be worth preserving in a simulation, if their overlapping functionality can be verified. Likewise, while the convective heater section, the radiant heater section, and the heat exchangers provide means to heat feed oil, the only means of heating treatment gas is through the heat exchangers.

Interestingly, the catalytic reactors have very self-contained means-ends links throughout Figure 19 and Figure 20. On its own, this representation is misleading. The reactors are very complicated in part *because* of the lack of means-ends: actors can control reactions only indirectly by manipulating temperatures, flows, and pressures *upstream* of the reactors. Another contributor to reactor coupling is that the recycle gas subunit, shown in Figure 21, has the only causal link that can decrease energy levels in the reactors. Decomposition of functional purposes may uncover more adaptive problem solving opportunities, such as shifting system operation in response to changing financial conditions, but pending further analysis, a DHT simulation's development may be more effectively guided by functional and engineering fidelity.

6.4.6 Performance Measures

As discussed in Section 5.2, the Abstract Function level is expected to provide a useful source of objective performance measures. Industry literature confirms this, suggesting that the treated product's sulfur mass concentration and the ratio of hydrogen mass consumption per mass of treated oil are key DHT performance metrics (Harwell et al., 2007). Because the DHT's outputs must be processed by other refinery plant units, the variability in mass sink flowrates such as sour water, sulfide, and vent gases may also be useful diagnostic measures of performance: large spikes in flow rates may cause difficulties in other plant equipment. While outside the system boundary for this analysis, use of the refinery flare system as an emergency mass or energy sink would also be both a loss of money and harmful to the environment.

Generalized functions can also be considered for performance measures if they are diagnostic of abstract functions. One example related to sulfur mass removal is the control of catalytic chemical reactions, which has not yet been modeled in sufficient detail to be discussed further.

6.4.7 Participant Engagement

While the purpose of the DHT is to contribute to the profitability of the refinery, the degree to which this motivates workers is not known. Some organizations engage in profit-sharing or quarterly bonus managerial techniques to synchronize system and worker goals, but in a refinery setting professional self-worth and social hierarchy likely play a strong role. Recruiting process control trainees as participants should be helpful, as they can be expected to engage with the simulation based on their desire to develop professional competency. Monetary compensation may only need to be provided for participation, rather than used as a performance incentive.

6.4.8 Cover Stories

For the design of a scaled world with high functional fidelity, there is no sense in making up an alternative cover story for the DHT simulation. However, if simulator technical limitations limit engineering fidelity or introduce other idealizations, cover stories may be used to coordinate participants' understanding of them.

6.4.9 Scenario Generation

Since workers' adaptive response to unanticipated events is a core focus of the HEPHAISTOS research program, simulator design should include a wide range of failure possibilities, both familiar and unfamiliar to participants.

Scenarios may be inspired by starting with conflicts between functional purpose sub-units in Table 8. Shifts in the financial environment that require process changes are examples of such challenges, or scenarios that induce conflicts as to whether to vent excess mass and energy or attempt to sustain process operations. Unfamiliar challenges can be developed by analysis of the equipment physical function level: explicitly identifying how check valves constrain process flow direction, as shown in Figure 17, can allow researchers to induce abnormal system flow patterns in scenarios. Reverse flows are often surprising to workers that are accustomed to systems' normal behavior (Vicente & Ethier, 2000).

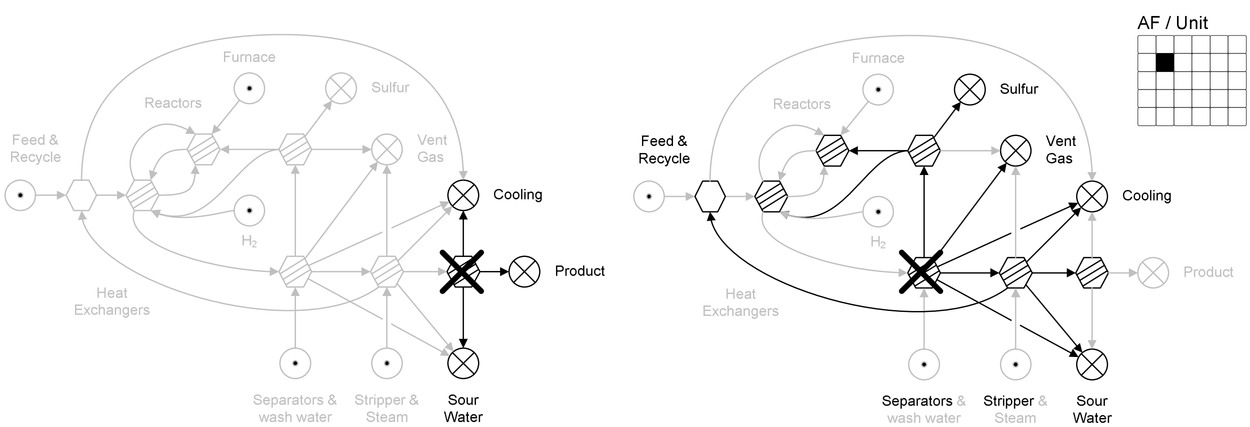


Figure 27 - Cascading effects of energy disturbance in downstream versus tightly coupled functional units

To categorize scenario difficulty, physical, generalized, and abstract function causal links may be useful. Comparing the location of faults' functional effects with respect to mass or energy sinks might be one approach, illustrated in Figure 27. For example, a malfunction in a unit that is close to a sink such as the drier or vent gas compressor may require only reduced feed flow rate and diversion of affected product. Scenarios that affect mass or energy feeds in highly interconnected areas of the system, such as the separators, may introduce cascading effects with wider control implications.

6.5 Discussion and Future work

The design suggestions outlined above provide a starting point for the design of a DHT simulation. They include:

- A preliminary proposal that the Component decomposition level be used as a baseline for definition of simulator functions.
- A low-interaction boundary along which to split the DHT simulation, if required.
- Preliminary modeling of energy and mass conservation relations to include in simulation dynamics
- Two areas where experts may make use of alternate means that should be preserved in a simulation implementation
- Five performance measures
- Two scenario difficulty classification measures, and a comprehensive model of system functions for scenario designers to consult

Using limited documentation, this analysis has produced a comprehensive list of simulation functions and causal models of system dynamics to guide software implementation, defined some data collection requirements, and provided guidance to developing and comparing scenarios. While the above suggestions fall short of being firm design decisions, simulator development involves many tradeoffs and only some of these have been identified in the WDA. One benefit of this systematic approach is that decisions can be made and easily revisited later if desired. For example, if later research questions suggested development of a microworld, the sub-unit level decomposition could be used to define a new simulator scope or functional fidelity without requiring re-analysis.

Future work should include a WDA of finer decomposition levels, and additional phases of CWA as described in Chapter 5. Accompanying this analysis, field studies of a DHT system or interviews with subject matter experts (SMEs) should be performed to inform and validate analysis. Some opportunities for future work are outlined below:

WDA

Completing a component level part-whole decomposition for Physical and Generalized functions is necessary to justify scaled world scope and functional fidelity decisions. Such detail will uncover more psychologically relevant means-ends connections than the existing sub-unit level analyses. Connections and elements in the Abstract Function level could be expanded to reflect the constraints that guide financial decisions to provide more insight into functional purposes. Financial elements could include the changing costs and values of electricity, fuel, hydrogen and other process inputs. These elements might also be linked to physical functions, if models can be found to describe the links between equipment operation, resulting wear and tear, and financial repercussions (R. Cooke & Paulsen, 1997).

A detailed analysis of the catalytic reactions and associated processes is definitely necessary. To complement this, the Abstract Function level may be extended to model features of the second law of thermodynamics such as entropy and exergy. While these do not encompass the same conservation constraints as mass and energy, entropy determines the reversibility of chemical reactions and can capture wasteful destructions of energy quality, such as friction associated with pressure drops. Because of the high pressures and flowrates involved in the DHT system, finer decomposition of Abstract Functions should also model energy inputs from electrically driven pumps and compressors.

CtrlTA

Existing automation must be analyzed, as its operation will be essential to the management of a scaled world DHT simulation. Since the proposed research program will involve manipulating automation performance and features, a thorough automation analysis is necessary.

Analysis of control tasks can be expected to suggest complementary performance measures to those discussed above. These could include number of control actions, problem-solving effectiveness, or diagnostic measures of use of leaps and shunts.

A CtrlTA in Work Domain Terms might also be useful in justifying simulator design scope and in generating scenarios. For example, operation of the DHT changes significantly during the lifespan of the catalyst. Fresh catalyst can make the system more subject to rapid chemical reactions and accompanying high temperature risks, while spent catalyst requires careful process operation to maintain product quality and extend catalyst life as long as possible.

StrA

Identifying plant operation strategies may help in making design choices mentioned in the Physical Fidelity section above. For example, if operators use monitoring strategies that involve physical form elements, their inclusion in a simulation will be more justified.

SOA

Further analysis to distinguish between automation and worker responsibilities, as well as that between field and control room workers will likely highlight opportunities for generalizable process control research. In some process control plants, for example, operators also have responsibility for physical manipulation of equipment in their area.

This can lead to strategic dilemmas between remaining at a computer console to refine hypotheses about a process upset, or running to the field in order to quickly activate potentially important manual valves.

Communications between DHT staff are presumably carried out using portable radios, which allow quick, flexible, rich communication but can present tractability problems for research due to the effort involved in transcribing and analyzing verbal protocols. Developing more tractable communication formats for simulation use would be very helpful in addressing team performance research questions and providing performance measures.

TM

Temporal Matching has not yet been applied as part of a CWA, so it may present analysts with theoretical challenges. Several temporal features of the DHT may benefit from its use, however. DHT process pressure limits found in available documentation (see Section 6.2) reflect both short term constraints, such as the elastic limits of process piping flanges, and other more long-term failure risks, such as hydrogen embrittlement of process vessels or corrosion of heat exchangers from catalytic reaction by-products. Comparing the time scales of these constraints should allow systematic consideration of which causal effects to include in a simulation, and suggest additional performance measures required to capture behaviors that trade off short-term benefits for long-term costs.

Such longer-term processes would be modeled as processes at the Generalized Function level, and would require additional study to determine psychologically relevant abstract functions to capture their effects. For any process to be included in a simulation, some form of mathematical model is required to meaningfully prompt and evaluate participants’ actions. Previous work in this area may provide a starting point (R. Cooke & Paulsen, 1997).

Chapter 7 Conclusion

As computer equipment continues to become more affordable, so does the opportunity for simulation of complex socio-technical systems for both training and research. The extent to which a simulator addresses training or research needs while minimizing costs depends on choices made during the design process.

The most popular methods for simulator design are based on task analysis or CTA approaches. Task analysis is an efficient way to capture procedures and strategies from natural systems that have been refined to address a wide range of system conditions. However, for cognitive engineering research, this type of approach has some weaknesses that have motivated development of alternative methods. Firstly, task analysis methods are poorly suited to model unanticipated events and associated adaptive problem solving (Vicente, 1999). This can result in simulations that focus on routine, procedural situations and encourage participants to simplify their problem solving behavior (Beare et al. 1984, qtd. in Moray, 2004). Second, task analysis does not distinguish between work and work-arounds, since task observations are intertwined with the displays, controls, automation, and team structure used in the natural system. Because of this, task analysis methods provide little analyst support in simplifying the observed tasks into more abstract microworlds, relying on sometimes unsubstantiated expert designer judgments (N. J. Cooke et al., 2004). Finally, a lack of consistent terminology makes comparing design efforts difficult and hampers development of graphical design tools or structured methods. Such weaknesses have been recognized and calls made for more formalized methods (Manser et al., 2007; Sauer et al., 2000)

This paper has proposed Cognitive Work Analysis as an alternative simulator design framework for computer-mediated complex sociotechnical systems. Previous work has suggested that CWA's strengths can address the current limitations of task-based approaches (Lintern & Naikar, 1998). For example, CWA's formative, constraint-based modeling language produces an analysis that is more flexible for describing and developing novel scenarios (Naikar & Sanderson, 1999). Secondly, CWA's systematic approach to description of the natural system provides structured set of terminology to consider and document scope and fidelity decisions, as well as to compare simulations. Finally, the analysis tools associated with CWA can help bridge the intuitive leaps necessary to transform analysis of the natural system into simulator specifications.

Disadvantages include the effort and time required to perform a CWA compared to more efficient task analysis methods. For example, the effortful analysis presented in Chapter 6 resulted in only six simulator design recommendations. The construction of CWA models likely demands more effort at the beginning of the simulator design process compared to CTA-based approaches. However, as proposed in Table 9, this may be partially offset by the simplification of the simulator implementation process. To implement a computer simulation model, functional specifications are often required

(Naikar & Sanderson, 1999), and a preliminary physics model (such as performed at the AF level) may also assist simulation engineers. For example, when using WDA as part of an interface design application, the contents of the ADS representation describe the *information requirements* that interfaces must contain if workers are to respond effectively to unforeseen events (Burns & Hajdukiewicz, 2004). In simulator design applications, once simulator functional fidelity and scope have been determined, designers can use the appropriate cells of the ADS as an *information specification*, cataloguing the functional elements that a simulation will contain. Means-ends and causal links could be used as validation metrics for simulation acceptance testing, to ensure that correspondence is not lost through implementation shortcuts.

Table 9 - Comparison of simulator design phases between CTA and CWA with associated work load

Simulator design phase (from Figure 4)	**CTA-based approach**	**CWA-based approach**
Analyze	Efficient: knowledge elicitation and cataloguing of tasks	Effortful: knowledge elicitation and development of CWA
Abstract & Idealize	Intuitive: Analyst expertise and design heuristics guide decisions	Methodical / Intuitive: Analysis model guides and documents decisions
Implement	Effortful: Simulator functional specifications need to be developed	Efficient: Existing analysis of functions and system structure provides simulator specifications

In conclusion, the advantages of a CWA approach may be less relevant to design of laboratory tasks. Because of their simplicity and researchers' desire for high experimental power, such simulations often repeat simple scenarios whose development may not benefit from CWA's range of analysis. Because of their low cost, laboratory tasks can usually be readily modified after-the-fact to correct any omissions or losses of correspondence. At the same time, benefits to high fidelity simulators may be undermined by the increased effort of analysis. High correspondence requirements mean that little abstraction or idealization will be performed and system processes, instrumentation and interfaces may be simply duplicated. However, CWA may contribute to high-fidelity simulators' ongoing operation, as noted in previous applications (Naikar, 2006).

The design of microworlds and scaled worlds may present the most challenge to a simulator design framework. Designers must reduce simulation complexity in a principled, defensible manner, while attempting to 'get it right the first time' (Vicente, 1999) to ensure tractable development costs. Hopefully CWA can contribute towards more widespread use of representative, affordable microworlds for research and training.

References

Ackerman, P. L. (1992). Predicting Individual Differences in Complex Skill Acquisition: Dynamics of Ability Determinants. *Journal of Applied Psychology, 77*(5), 598-614.

Backs, R. W., Navidzadeh, T., & Xu, X. (2000). *Cardiorespiratory Indeces of Mental Workload During Simulated Air Traffic Control*. Paper presented at the IEA 2000 / HFES 2000 Congress.

Bennett, K. B., Nagy, A. L., & Flach, J. M. (2006). Visual Displays. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (Vol. 3rd, pp. 1191-1219). Hoboken, NJ: Wiley.

Bennett, K. B., Toms, M. L., & Woods, D. D. (1993). Emergent Features and Graphical Elements: Desigining More Effective Configural Displays. *Human Factors, 35*(1), 71-97.

Berndtsson, J., & Normark, M. (1999). Coordinative functions of flight strips: Air traffic control work revisited. In *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work* (pp. 101-110).

Berutti, M. (2007). Understanding Simulation Fidelity. Retrieved April 4th, 2007, from http://www.mynah.com/pdf/Understanding%20Simulation%20Fidelity.pdf

Blevins, T. L., McMillan, G. K., Wojsznis, W. K., & Brown, M. W. (2003). *Advanced Control Unleased*. Research Triangle Park: Instrumentation, Systems, and Automation Society.

Bowers, C., Jentsch, F., Schifflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (2004). Engineered Experiences for Scaled Worlds: An Example from Aviation. In Anonymous (Ed.), *Scaled Worlds: Development, Validation, and Applications* (Vol. 1, pp. 67-74). Aldershot, England: Ashgate.

Brehmer, B., & Allard, R. (1991). Dynamic Decision Making: The Effects of Task Complexity and Feedback Delay. In J. Rasmussen, B. Brehmer & J. Leplat (Eds.), *Distributed Decision Making: Cognitive Models for Cooperative Work* (1 ed., Vol. 1). West Sussex: Wiley.

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior, 9*(2-3), 171-184.

Brehmer, B., Leplat, J., & Rasmussen, J. (1991). Use of Simulation in the Study of Complex Decision Making. In J. Rasmussen, B. Brehmer & J. Leplat (Eds.), *Distributed Decision Making: Cognitive Models for Cooperative Work* (1 ed., Vol. 1). West Sussex: Wiley.

Brehmer, B., Schifflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (2004). Some Reflections on Microworld Research. In S. G. Schifflett, L. R. Elliott, E. Salas & M. D. Coovert (Eds.), *Scaled Worlds: Development, Validation, and Applications* (Vol. 1, pp. 22-36). Aldershot, England: Ashgate.

Burns, C. M., Garrison, L., & Dinadis, N. (2003). From analysis to design: WDA for the petrochemical industry. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 258-262): HFES.

Burns, C. M., & Hajdukiewicz, J. R. (2004). *Ecological Interface Design*. Boca Raton, Florida: CRC Press.

Burns, C. M., Momtahan, K., & Enomoto, Y. (2006). *Supporting the strategies of cardiac nurse coordinators using cognitive work analysis*. Paper presented at the Human Factors and Ergonomics Society 50th Annual Meeting.

Chapanis, A. (1967). The relevance of laboratory studies to practical situations. *Ergonomics, 10*(5), 557-577.

Chipman, S. F., Schraagen, J. M., & Shalin, V. L. (2000). *Cognitive Task Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cooke, N. J., Schifflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (2004). Designing a Synthetic Task Environment. In S. G. Schifflett, L. R. Elliott, E. Salas & M. D. Coovert (Eds.), *Scaled Worlds: Development, Validation, and Applications* (Vol. 1, pp. 263-278). Aldershot, England: Ashgate.

Cooke, R., & Paulsen, J. (1997). Concepts for measuring maintenance performance and methods for analysing competing failure modes. *Reliability Engineering and System Safety, 55*(2), 135-141.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psycological Bulletin, 130*(6), 959-988.

Elliott, L. R., Dalrymple, M. A., Schiflett, S. G., & Miller, J. C. (2004). Scaling Scenarios: Development and Application to C4ISR Sustained Operations Research. In S. G. Schifflett, L. R. Elliott, E. Salas & M. D. Coovert (Eds.), *Scaled Worlds: Development, Validation, and Applications* (Vol. 1, pp. 119-133). Aldershot, England: Ashgate.

Farbos, B., Mollard, R., Cabon, P., & David, H. (2000). *Measurement of Fatigue and Adaptation in Large-Scale Real-Time ATC Simulation*. Paper presented at the IEA 2000 / HFES 2000 Congress.

FISG. (1998, 2007/07/28). Fidelity Implementation Study Group Glossary. from http://www.sisostds.org/index.php?tg=fileman&idx=list&id=9&gr=Y&path=Fidelity+Experimentation

Goettl, B. P. (2006). Training and acquisition of complex tasks. In W. Karwowski (Ed.), *International Encyclopedia of Ergonomics and Human Factors* (second ed., Vol. 1, pp. 218-221). Boca Raton, FL: Taylor & Francis Group, LLC.

Goldstein, W. M. (2006). Introduction to Brunswikian Theory and Method. In A. Kirlik (Ed.), *Adaptive Perspectives on Human-Technology Interaction* (pp. 10-24). New York: Oxford University Press.

Golomb, S. W. (1968, January). Mathematical Models - Uses and Limitations. *Astronautics & Aeronautics*, 57-59.

Granlund, R. (2003). Monitoring experiences from command and control research with the C3Fire microworld. *Cognition, Technology & Work, 5*, 183-190.

Granlund, R., Johannson, B., Schifflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (2004). Monitoring Distributed Collaboration in the C3Fire Microworld. In Anonymous (Ed.), *Scaled Worlds: Development, Validation, and Applications* (Vol. 1, pp. 37-48). Aldershot, England: Ashgate.

Gray, W. D. (2002). Simulated Task Environments: The Role of High-Fidelity Simulations, Scaled Worlds, Synthetic Environments, and Laboratory Tasks in Basic and Applied Cognitive Research. *Cognitive Science Quarterly, 2002*(2), 205-227.

Gugerty, L., Schifflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (2004). Using Cognitive Task Analysis to Design Multiple Synthetic Tasks for Uninhabited Aerial Vehicle Operation. In S. G. Schifflett, L. R. Elliott, E. Salas & M. D. Coovert (Eds.), *Scaled Worlds: Development, Validation, and Applications* (Vol. 1, pp. 240-262). Aldershot, England: Ashgate.

Gula, C. A. (1998). *FATS III Combat firing simulator validation study*. North York: Defence and Civil Institute of Environmental Medicine.

Hajdukiewicz, J. R., Vicente, K. J., Doyle, D. J., Milgram, P., & Burns, C. M. (2001). Modeling a medical environment: An ontology for integrated medical informatics design. *International Journal of Medical Informatics, 62*(1), 79-99.

Hammond, K. R. (1998). Ecological Validity: Then and Now.   Retrieved 2007/07/28, from http://www.brunswik.org/notes/essay2.html

Harwell, L., Thakkar, S., Polcar, S., Palmer, R. E., & Desai, P. H. (2007). Study outlines optimum ULSD hydrotreater design. *Oil & Gas Journal, 105*(32).

Hess, S. M., MacMillan, J., & Serfaty, D. (1999, 1999). *Development of a team-in-the-loop, synthetic simulation environment based on data from Cognitive Task Analysis.* Paper presented at the Human Factors and Ergonomics Society Annual Meeting, 43rd, Houston, TX.

Hinsz, V. B., Malone, C. P., Schifflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (2004). Memory for Attributes of Information Presented in a Synthetic Task Environment: An Illustration with AWACS Weapons Directors' Displays. In Anonymous (Ed.), *Scaled Worlds: Development, Validation, and Applications* (Vol. 1, pp. 220-239). Aldershot, England: Ashgate.

Hockey, G. R. J., Wastell, D. G., & Sauer, J. (1998). Effects of sleep deprivation and user interface on complex performance: a multilevel analysis of compensatory control. *Human Factors, 40*, 233-254.

Humphreys, P. (2004). Computer Simulations. In *Extending Ourselves* (pp. 105-135). Oxford: Oxford University Press.

IFE. (2006). The Halden Reactor Project.   Retrieved May 26, 2007, 2007, from http://www.ife.no/hrp/

Jamieson, G. A. (2003). *Ecological Interface Design for Petrochemical Process Control: Integrating Task- and System-Based Approaches*. University of Toronto, Toronto.

Jamieson, G. A., & Vicente, K. J. (2001). Ecological interface design for petrochemical applications: supporting operator adaptation, continuous learning, and distributed, collaborative work. *Computers and Chemical Engineering, 25*, 1055-1074.

Jamieson, G. A., & Vicente, K. J. (2005). Designing effective human-automation-plant interfaces: A control-theoretic perspective. *Human Factors, 47*(1), 12-34.

Jobidon, M.-E., Breton, R., Rousseau, R., & Tremblay, S. (2006). *Team Response to Workload Transition: the Role of Team Structure*. Paper presented at the Human Factors and Ergonomics Society 50th Annual Meeting.

Johannson, B., Persson, M., Granlund, R., & Mattsson, P. (2003). C3Fire in command and control research. *Cognition, Technology & Work*(5), 191-196.

Jones, M. R. (2005). Idealization and Abstraction: A Framework. In M. R. Jones & N. Cartwright (Eds.), *Idealization XII: Correcting the Model. Idealization and Abstraction in the Sciences* (pp. 173-217). Amsterdam/New York: Rodopi.

Kilgore, R., & St-Cyr, O. (2006). *SRK Inventory: A tool for structuring and capturing a worker competencies analysis.* Paper presented at the Human Factors and Ergonomics Society 50th Annual Meeting.

Kilgore, R., St-Cyr, O., & Jamieson, G. A. (in press). From Work Domains to Worker Competencies: A Five-Phase CWA for Air Traffic Control.

Kirlik, A., Miller, R. A., & Jagacinski, R. J. (1993). Supervisory Control in a Dynamic and Uncertain Environment: A Process Model of Skilled Human-Environment Interaction. *IEEE Transactions on Systems, Man and Cybernetics, 23*(4), 929-953.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds.), *Decision Making in Action: Models and Methods* (pp. 138-147). Norwood, NJ: Ablex.

Kozlowski, S. W. J., DeShon, R. P., Schifflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (2004). A Psychological Fidelity Approach to Simulation-Based Training: Theory, Research, and Principles. In Anonymous (Ed.), *Scaled Worlds: Development, Validation, and Applications* (Vol. 1, pp. 75-99). Aldershot, England: Ashgate.

Lee, J. D. (2001). *Matching Automotive Simulator Capabilities to the Emerging Associated with In-Vehicle Information Systems*. Paper presented at the Human Computer Interaction International.

Lee, J. D., & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human factors, 46*(1), 50-80.

Leffler, W. L. (2002). *Petroleum refining in nontechnical language* (Third ed.). Tulsa: PennWell Corporation.

Li, X., Sanderson, P., & Memisevic, R. (2006). Evaluating functional displays for hydropower system: model-based guidance of scenario design. *Cognition, Technology & Work, 8*(4), 269-282.

Lindgren, I., & Smith, K. (2006). *National Patterns of Teamwork during an Emergency Management Simulation*. Paper presented at the Human Factors and Ergonomics Society 50th Annual Meeting.

Lintern, G. (2006). Foundational issues for work domain analysis. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*.

Lintern, G., & Naikar, N. (1998). *Cognitive Work Analysis for Training Design.* Paper presented at the Proceedings of the Australasian Computer Human Interaction Conference.

Lintern, G., & Naikar, N. (2000). The Use of Work Domain Analysis for the Design of Training Systems. *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennnium', San Diego, CA*, 198-201.

Loft, S., Hill, A., Neal, A., Humphreys, M., & Yeo, G. (2004). ATC-lab: An air traffic control simulator for the laboratory. *Behavior Research Methods, Instruments, & Computers, 36*(2), 331-338.

Manser, T., Dieckmann, P., Wehner, T., & Rall, M. (2007). Comparison of anaesthetists' activity patterns in the operating room and during simulation. *Ergonomics, 50*(2), 246-260.

Maran, N. J., & Glavin, R. J. (2003). Low- to high-fidelity simulation – a continuum of medical education? *Medical Education, 37*(Suppl. 1), 22-28.

Memisevic, R., Sanderson, P., Choudhury, S., & Wong, B. L. W. (2005). *Work domain analysis and ecological interface design for hydropower system monitoring and control.* Paper presented at the 2005 IEEE International Conference on Systems, Man and Cybernetics.

Memisevic, R., Sanderson, P., Wong, W. B. L., Choudhury, S., & Li, X. (2007). Investigating human-system interaction with an integrated hydropower and market system simulator. *IEEE Transactions on Power Systems, 22*(2), 762-769.

Miller, C. A., & Vicente, K. J. (1998a). *Abstraction Decomposition Space Analysis for NOVA's E1 Acetylene Hydrogenation Reactor* (No. CEL98-09). Toronto: Cognitive Engineering Laboratory.

Miller, C. A., & Vicente, K. J. (1998b). *Towards an integration of task- and work domain analysis techniques for Human-Computer Interface design.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society.

Miller, R. B. (1953). *Psychological considerations in the design of training* *equipment.* (No. WADC-TR-54-563, AD 71202). Wright Patterson Air Force Base, OH: Wright Air Development Center.

Moray, N. (2004). Ou sont les neiges d'antan? In D. A. Vincenzi, M. Mouloua & P. A. Hancock (Eds.), *Human Performance, Situation Awareness and Automation: Current Research and Trends*. Mahwah: LEA.

Moray, N., & Haudegond, S. (1998). An absence of vigileance decrement in a complex dynamic task. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*.

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There Is More to Monitoring a Nuclear Power Plant than Meets the Eye. *Human Factors, 42*(1), 36-55.

Naikar, N. (2005). *Theoretical concepts for Work Domain Analysis, the first phase of Cognitive Work Analysis.* Paper presented at the Human Factors and Ergonomics Society 49th Annual Meeting, Orlando, FL.

Naikar, N. (2006). Beyond interface design: Further applications of cognitive work analysis. *International journal of industrial ergonomics, 36*(5), 423-438.

Naikar, N., Hopcroft, R., & Moylan, A. (2005). *Work Domain Analysis: Theoretical Concepts and Methodology* (Report No. DSTOTR1665; DODAAR013299; ADA449707). Victoria (Australia). Air Operations Div; United States: Defence Science and Technology Organisation.

Naikar, N., Pearce, B., Drumm, D., & Sanderson, P. M. (2003). Designing teams for first-of-a-kind, complex systems using the initial phases of cognitive work analysis: Case study. *Human Factors, 45*(2), 202-217.

Naikar, N., & Sanderson, P. M. (1999). Work domain analysis for training-system definition and acquisition. *International Journal of Aviation Psychology, 9*(3), 271-290.

Nunes, A., & Matthews, M. L. (2002). *The Effect of Predictive Aid Usage on Controller Strategies & Mental Demand under Direct Routing*. Paper presented at the The Human Factors and Ergonomics Society 46th Annual Meeting.

O'Brien, K. S., & O'Hare, D. (2007). Situational awareness ability and cognitive skills training in a complex real-world task. *Ergonomics, 50*(7), 1064-1091.

Rasmussen, J. (1974). *The human data processor as a system component: Bits and pieces of a model* (No. 1722

). Roskilde, Denmark: Research Establishment Risø.

Rasmussen, J. (1983). Skills, rules, knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13*, 257-266.

Rasmussen, J. (1986). *Information processing and human-machine interaction : an approach to cognitive engineering*. New York: North-Holland.

Rasmussen, J., & Goodstein, L. P. (1987). Decision support in supervisory control of high-risk industrial systems. *Automatica, 23*, 663-671.

Rasmussen, J., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering*. New York: Wiley.

Rasmussen, J., Pejtersen, A. M., & Schmidt, K. (1990). *Taxonomy for cognitive work analysis*: Riso National Laboratory.

Reising, D. V. C., & Sanderson, P. M. (2000). *Testing the impact of instrumentation location and reliability on ecological interface design: Control performance*. Paper presented at the The XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennnium'. from http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0000092616&partnerID=40&rel=R5.6.0

Reising, D. V. C., & Sanderson, P. M. (2002a). Work domain analysis and sensors I: principles and simple example. *International Journal of Human-Computer Studies, 56*(6), 569-596.

Reising, D. V. C., & Sanderson, P. M. (2002b). Work domain analysis and sensors II: Pasteurizer II case study. *International Journal of Human Computer Studies, 56*(6), 597-637.

Roza, Z. C., Gross, D. C., & Harmon, S. Y. (2000). *Report Out of the Fidelity Experimentation ISG* (No. 00S-SIW-151): Simulation Interoperability and Standards Organization.

Sanderson, P. M., & Naikar, N. (2000). Temporal coordination control task analysis for analysing human-system integration. *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennnium', San Diego, CA*, 206-209.

Sauer, J., Hockey, G. R. J., & Wastell, D. G. (1999). Maintenance of complex performance during a 135-day spaceflight simulation. *Space, Aviation and Environmental Medicine, 70*(3), 236-244.

Sauer, J., Wastell, D. G., & Hockey, G. R. J. (2000). Conceptual framework for designing micro-worlds for complex work domains: a case study of the Cabin Air Management System. *Computers in Human Behavior, 16*(1), 45-58.

Sheridan, T. B. (2006). Supervisory Control. In *Handbook of human factors and ergonomics* (3rd ed., pp. 1025-1052). Hoboken, NJ: Wiley.

Skraaning jr., G. (2003). *Experimental Control versus Realism: Methodological Solutions for Simulator Studies in Complex Operating Environments.* Norwegian University of Science and Technology, Trondheim.

Skraaning Jr., G., Lau, N., Welch, R., Nihlwing, C., Andresen, G., Brevig, L. h., et al. (2007). *The Ecological Interface Design Experiment (2005) HWR-833*. Halden, Norway: OECD Halden Reactor Project

Snook, S. A. (2000). *Friendly Fire*. Princeton: Princeton University Press.

Stanton, N. (1996). Simulators: A review of research and practice. In N. Stanton (Ed.), *Human Factors in Nuclear Safety*: Taylor & Francis.

van den Bosch, K., Riemersma, J. B. J., Schifflett, S. G., Elliott, L. R., Salas, E., & Coovert, M. D. (2004). Reflections on Scenario-Based Training in Tactical Command. In Anonymous (Ed.), *Scaled Worlds: Development, Validation, and Applications* (Vol. 1, pp. 1-21). Aldershot, England: Ashgate.

Vicente, K. J. (1995). Cognitive work analysis: implications for microworld research on human-machine interaction. *Proceedings of the 1995 IEEE International Conference on Systems, Man and Cybernetics. Part 2 (of 5), 4*, 3432-3436.

Vicente, K. J. (1999). *Cognitive work analysis : toward safe, productive, and healthy computer-based work*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Vicente, K. J., & Ethier, C. R. (2000). Why fluid dynamics matters for display design in process control: Commentary on Bennett and Malek. *Human Factors, 42*(3), 451-454.

Vicente, K. J., & Rasmussen, J. (1990). The ecology of human-machine systems II: Mediating 'direct perception' in complex work domains. *Ecological Psychology, 2*, 207-250.

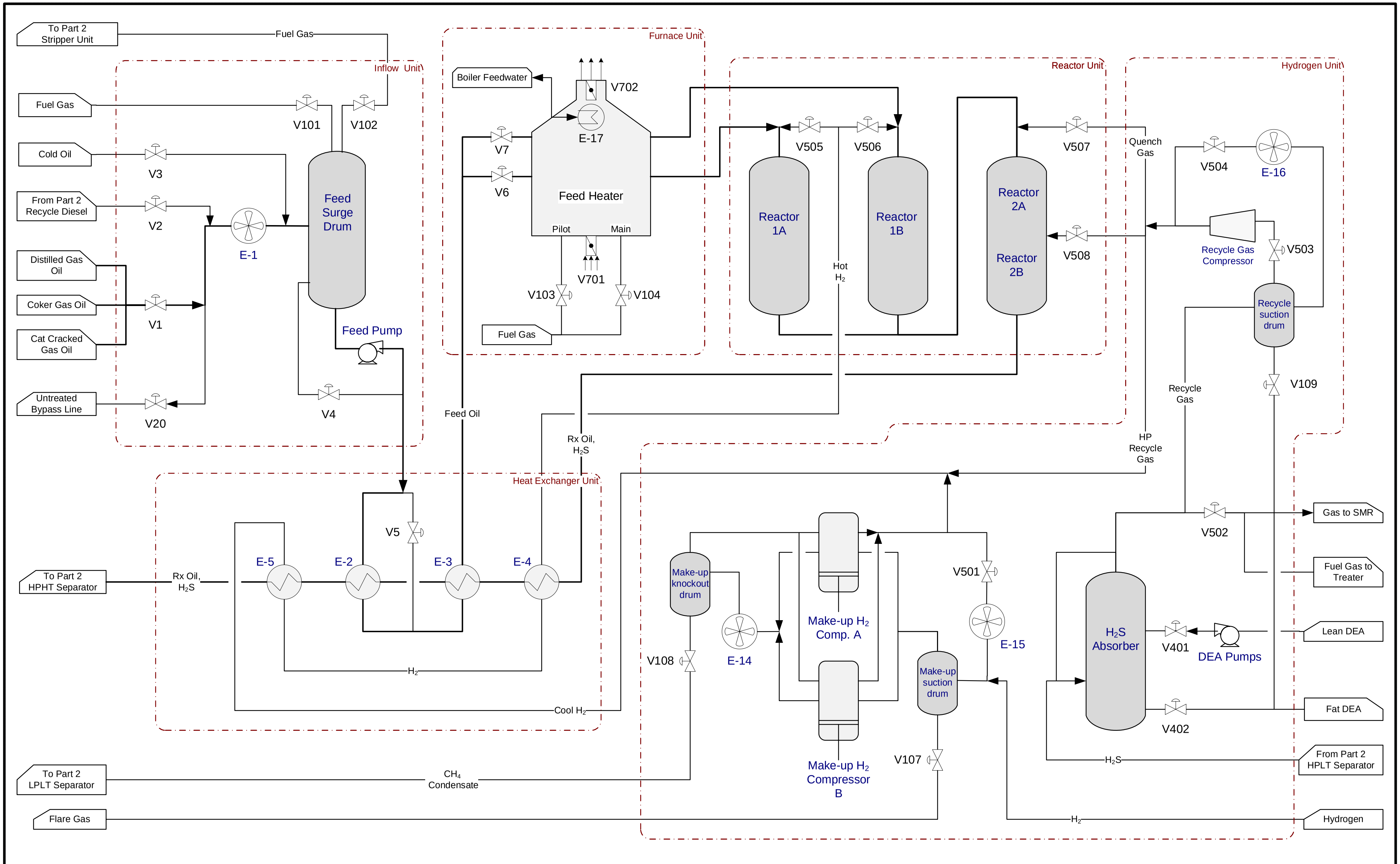
Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man and Cybernetics, 22*(4), 589-606.

Wesson. (2007). Wesson Interactive. Retrieved Sep.4, 2007, from http://www.wesson.com/

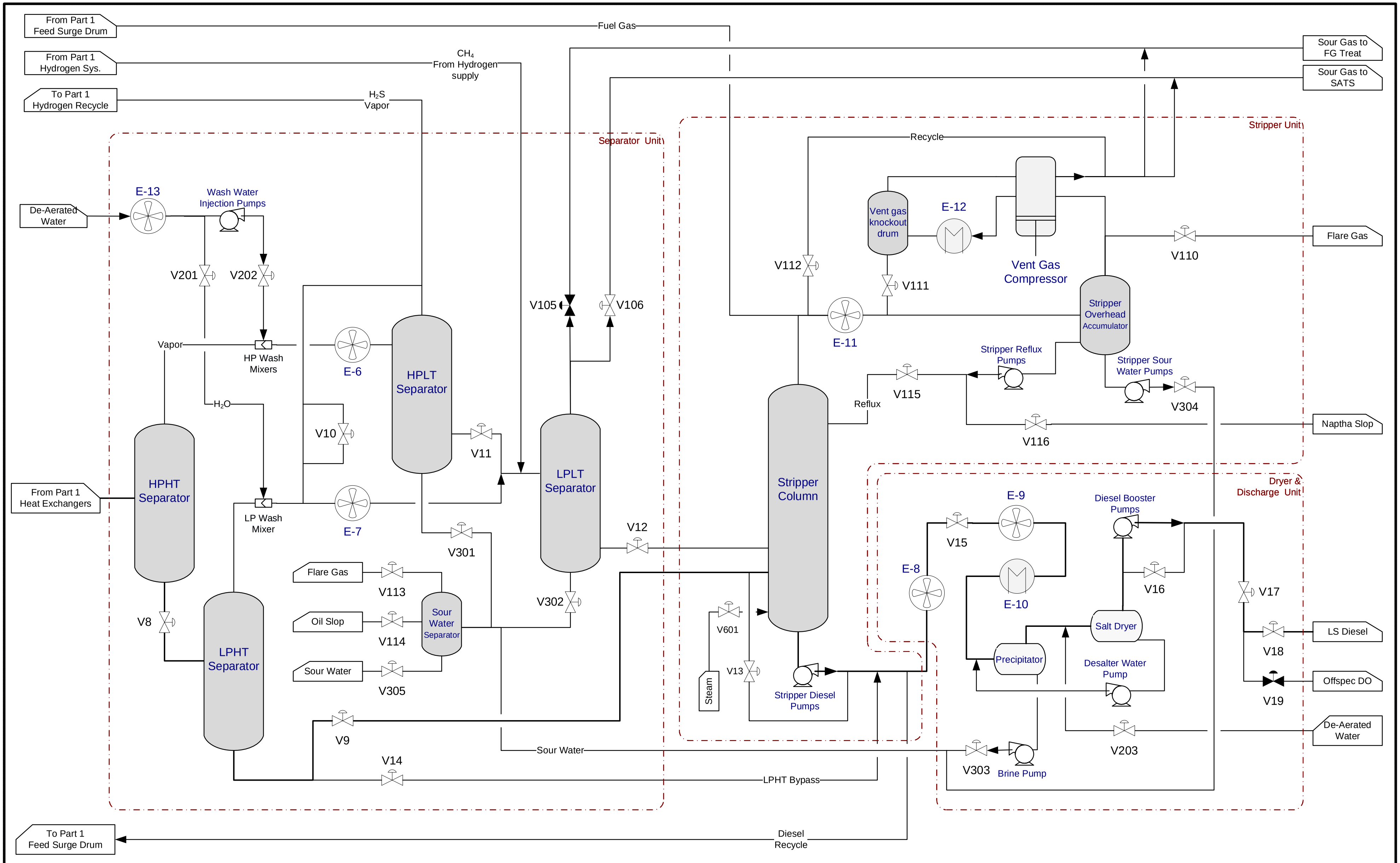
Zingale, C. M., Gromelski, S., Ahmed, S. B., & Stein, E. S. (1993). *Influence of individual experience and flight strips on air traffic controller memory / situational awareness* (No. DOT/FAA/CT-TN93/31). Atlantic City International Airport: Federal Aviation Administration Technical Center.

Zuboff, S. (1988). *In the age of the smart machine : the future of work and power*. New York: Basic Books.

Appendix A DHT Physical Functions at Equipment Level

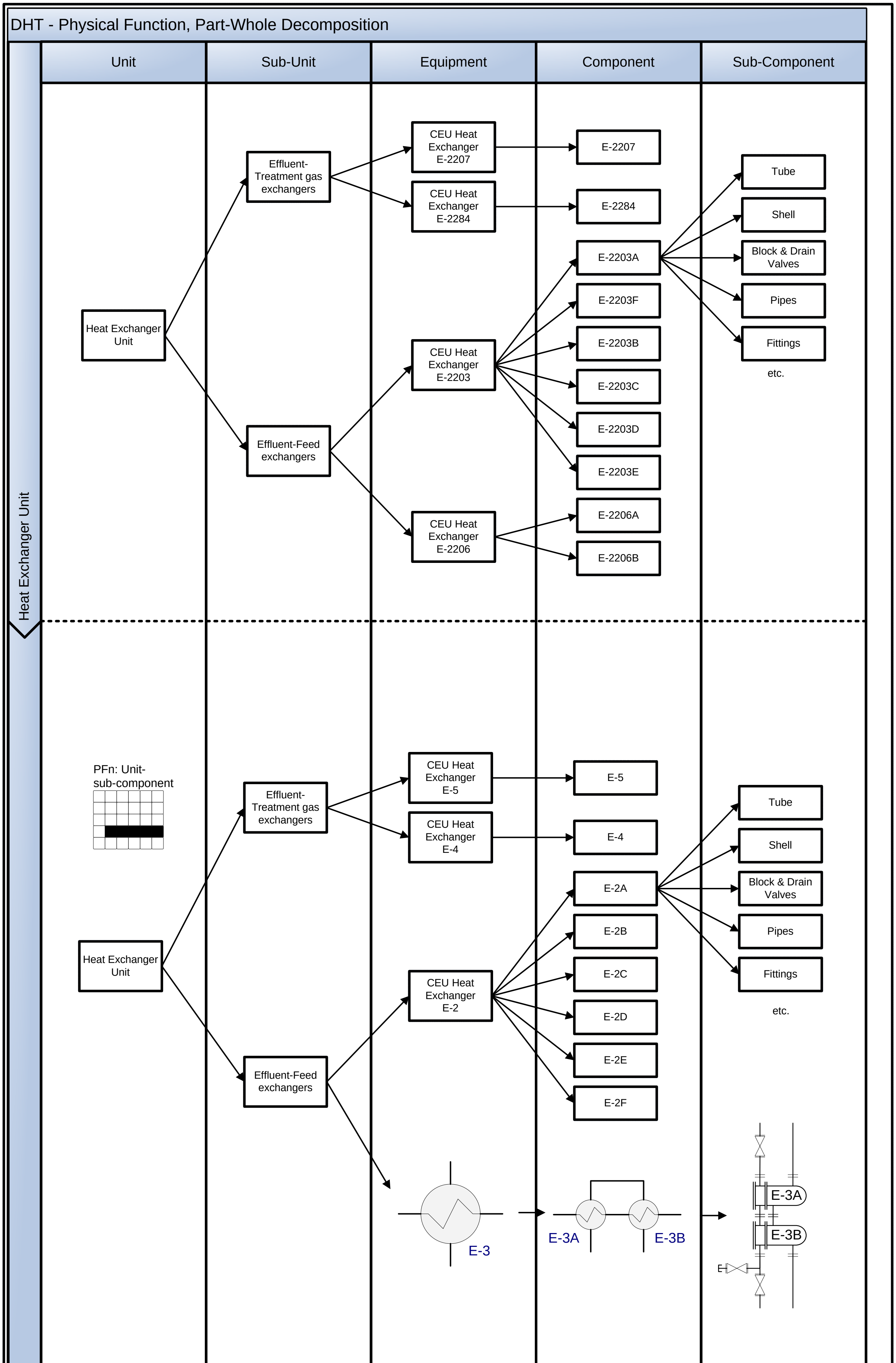


NODE: E-PFn-C TITLE: Equipment decomposition Physical Function, Causal Links - PART 1 NO.: 2007/08/28

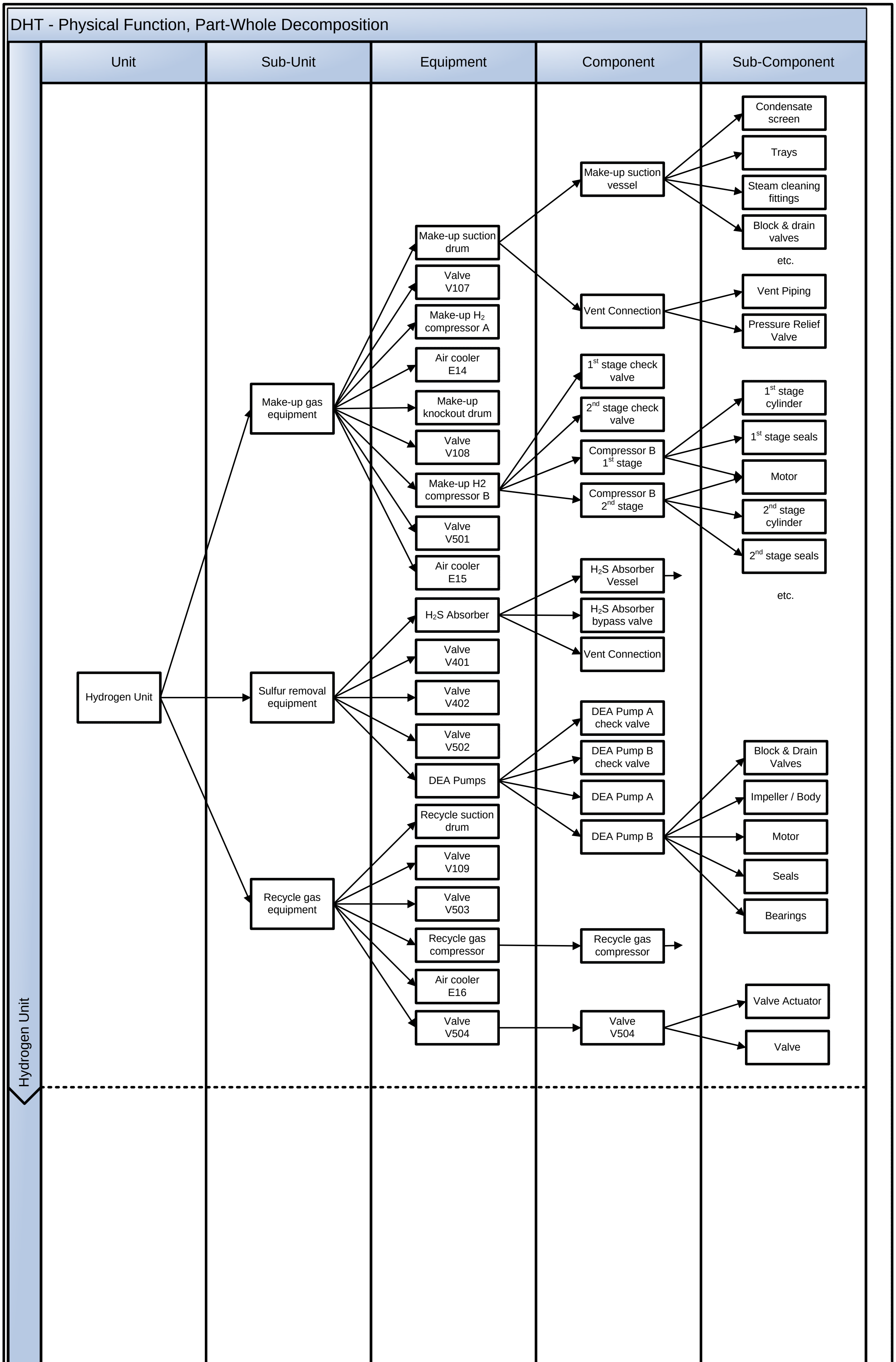


NODE: E-PFn-C | TITLE: Equipment decomposition Physical Function, Causal Links - PART 2 | NO.: 2007/08/28

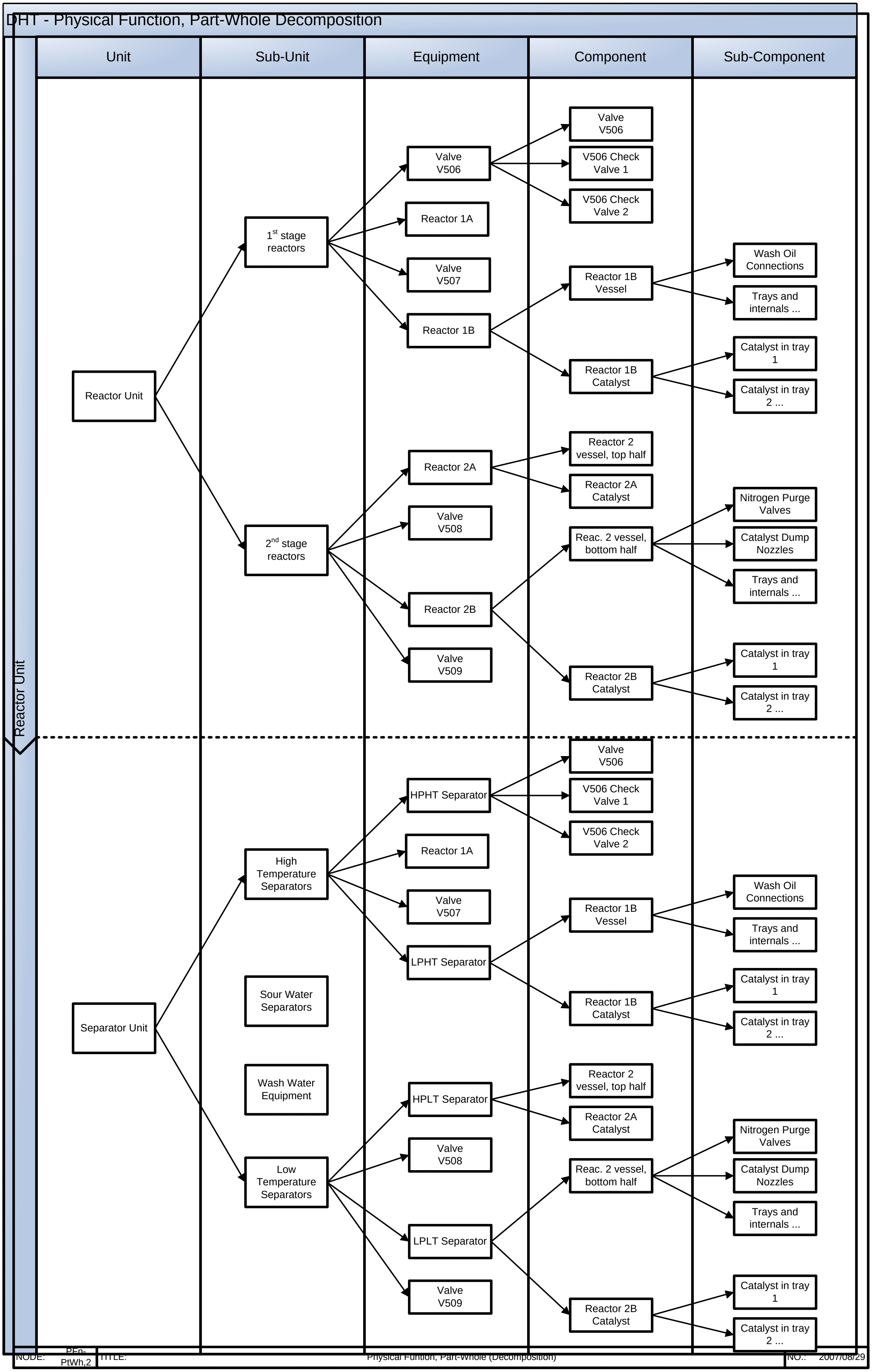
Appendix B Physical Function, Part-Whole Decomposition



NODE:	PFn-PtWh,1	TITLE: Physical Funtion, Part-Whole (Decomposition)	NO.:	2007/08/29



NODE: PFn-PtWh,2 | TITLE: Physical Funtion, Part-Whole (Decomposition) | NO.: 2007/08/29



DHT - Physical Function, Part-Whole Decomposition
Unit
Sub-Unit
Equipment
Component
Sub-Component
Reactor Unit
Reactor Unit
1st stage reactors
2nd stage reactors
Valve V506
Reactor 1A
Valve V507
Reactor 1B
Valve V506
V506 Check Valve 1
V506 Check Valve 2
Reactor 1B Vessel
Reactor 1B Catalyst
Wash Oil Connections
Trays and internals ...
Catalyst in tray 1
Catalyst in tray 2 ...
Reactor 2A
Valve V508
Reactor 2B
Valve V509
Reactor 2 vessel, top half
Reactor 2A Catalyst
Reac. 2 vessel, bottom half
Reactor 2B Catalyst
Nitrogen Purge Valves
Catalyst Dump Nozzles
Trays and internals ...
Catalyst in tray 1
Catalyst in tray 2 ...
Separator Unit
High Temperature Separators
Sour Water Separators
Wash Water Equipment
Low Temperature Separators
HPHT Separator
Reactor 1A
Valve V507
LPHT Separator
Valve V506
V506 Check Valve 1
V506 Check Valve 2
Reactor 1B Vessel
Reactor 1B Catalyst
Wash Oil Connections
Trays and internals ...
Catalyst in tray 1
Catalyst in tray 2 ...
HPLT Separator
Valve V508
LPLT Separator
Valve V509
Reactor 2 vessel, top half
Reactor 2A Catalyst
Reac. 2 vessel, bottom half
Reactor 2B Catalyst
Nitrogen Purge Valves
Catalyst Dump Nozzles
Trays and internals ...
Catalyst in tray 1
Catalyst in tray 2 ...
NODE: PFn-PtWh,2
TITLE: Physical Funtion, Part-Whole (Decomposition)
NO.: 2007/08/29