UNITED STATES
NUCLEAR REGULATORY COMMISSION
ADVISORY COMMITTEE ON REACTOR SAFEGUARDS
WASHINGTON, DC 20555 - 0001

January 20, 2004

MEMORANDUM TO: ACRS MEMBERS

FROM: Med El-Zeftawy, Senior Staff Engineer MMZ
ACRS

SUBJECT: CERTIFICATION OF THE MINUTES FOR THE MEETING OF THE ACRS SUBCOMMITTEE ON HUMAN FACTORS, DECEMBER 2, 2003 ROCKVILLE, MARYLAND

The minutes of the subject meeting, issued on December 16, 2003, have been certified as the official record of the proceedings for that meeting. A copy of the certified minutes is attached.

Attachment: As stated

cc: ACRS Members
R. Savio

cc via e-mail: ACRS Members
J. Larkins
R. Savio
H. Larson
S. Duraiswamy
ACRS Staff Engineers

MEMORANDUM TO: Med El-Zeftawy, Senior Staff Engineer
ACRS

FROM: Stephen L. Rosen, Chairman
Human Factors Subcommittee

SUBJECT: CERTIFICATION OF THE MINUTES FOR THE MEETING OF THE ACRS SUBCOMMITTEE ON HUMAN FACTORS, DECEMBER 2, 2003–ROCKVILLE, MARYLAND

I do hereby certify that, to the best of my knowledge and belief, the minutes of the subject meeting on December 2, 2003, are an accurate record of the proceeding for that meeting.

S. Rosen 1/15/04

Stephen L. Rosen Date
Subcommittee Chairman

UNITED STATES
NUCLEAR REGULATORY COMMISSION
ADVISORY COMMITTEE ON REACTOR SAFEGUARDS
WASHINGTON, D.C. 20555-0001

December 16, 2003

MEMORANDUM TO: Stephen L. Rosen, Chairman
Human Factors Subcommittee

FROM: Med El-Zeftawy, Senior Staff Engineer
ACRS

SUBJECT: WORKING COPY OF THE MINUTES FOR THE MEETING OF THE ACRS SUBCOMMITTEE ON HUMAN FACTORS, DECEMBER 2, 2003–ROCKVILLE, MARYLAND

A working copy of the minutes for the subject meeting is attached for your review. Please review and comment on them at your earliest convenience. Copies are being provided to each ACRS Member who attended the meeting for information and/or review.

Attachment:
As Stated

cc: ACRS members
J. Larkins
S. Bahadur
H. Larson

CERTIFIED BY: S. Rosen
On: January 15, 2004

ISSUED Date: December 16, 2004

ADVISORY COMMITTEE ON REACTOR SAFEGUARDS
HUMAN FACTORS SUBCOMMITTEE
MEETING MINUTES–DECEMBER 2, 2003
ROCKVILLE, MARYLAND

INTRODUCTION

The ACRS Subcommittee on Human Factors met on December 2, 2003, at 11545 Rockville Pike, Rockville, Maryland, in Room T-2B3The purpose of this meeting was to review and discuss with the NRC staff the recent updates to Chapter 18.0 of the Standard Review Plan, Human Factors Engineering, and the documents referenced in that chapter.

The Subcommittee received a request from one member of the public to make an oral statement during the meeting concerning the subject matter. The entire meeting was open to public attendance Med El-Zeftawy was the cognizant ACRS staff engineer and the Designated Federal Official for this meeting. The meeting was convened at 1:00 p.m. and adjourned at 4:30 p.m.

ATTENDEES

ACRS

S. Rosen, Subcommittee Chairman
T. Kress, Member
D. Powers, Member
J. Sieber, Member
M. El-Zeftawy, ACRS Staff Engineer

NRC

J. Bongarra, NRR
S. Cooper, RES
R. Eckenrode, NRR
J. Flack, RES
J. Ibarra, RES
J. Kramer, RES
M. Keefe, RES
P. Lewis, RES
G. Parry, NRR
J. Persensky, RES

OTHER

R. Fuld, public member
J. Higgins, BNL
J. O'Hara, BNL

A complete list of attendees is in the ACRS Office file and will be made available upon request The presentation slides and handouts used during the meeting are attached to the office copy of these minutes.

OPENING REMARKS BY THE SUBCOMMITTEE CHAIRMAN

Mr. Stephen Rosen, Human Factors Subcommittee Chairman, convened the meeting at 1:00 p.m. He stated that the purpose of this meeting is to hear a briefing by and hold discussions with the NRC staff regarding the recent updates of draft NUREG-0800, Standard Review Plan (SRP) Chapter 18.0, Human Factors Engineering. The Subcommittee will gather information, analyze relevant issues and facts, and formulate proposed positions and actions, as appropriate, for deliberation by the full Committee.

Mr. Rosen indicated that the subject of this meeting is of great importance to the agency and to the public at large, especially in the context of the current discussions on fire safety and manual actions as to whether they would be credited or not. This matter will also be discussed during the December 3-5, 2003 full Committee meeting. Mr. Rosen stated that the Subcommittee has received a request for time to make an oral statement from a member of the public, and the Subcommittee will honor such request.

NRC Staff Presentation

Mr. Julius Persensky, NRC Office of Nuclear Regulatory Research, stated that Chapter 18.0 of the SRP provides the framework for the conduct of human factors engineering reviews for nuclear power plants. For more detailed guidance concerning the review process, the SRP references NUREG-0711, "Human Factors Engineering Program Review Model." For review criteria that are specifically tailored to the review of plant modifications and license amendment requests involving credited operator actions, the SRP references NUREG-1764, "Guidance for the Review of Changes to Operator Actions." For guidance concerning human-system interfaces, the SRP and NUREG-0711 reference NUREG-0700, "Human-System Interface Design Review Guidelines."

Mr. James Bongarra, NRC Office of Nuclear Reactor Regulation, indicated that Chapter 18.0, NUREG-0700, and NUREG-0711 are fundamental human factors review documents and they have been revised to support reviews of advanced reactors and digital updates to existing control rooms. NUREG-1764 is a new document that provides means to use risk information to determine the appropriate level of human factors review.

SRP Chapter 18.0 provides a high level framework for all human factors engineering (HFE) reviews. It includes three technical review areas: review of new plants; review of control room modifications; and review of changes to human actions. The Chapter also describes a process for evaluating designs, design processes, design reviews, and operator actions submitted by applicants and licensees for the NRC review. The staff identified 12 areas of review that are needed for successful integration of human characteristics and capabilities into plant design. Such areas include HFE program management, operating experience, human reliability analysis, staffing and qualifications, human factors verification and validation, human-system interface design, training, and monitoring.

Mr. Paul Lewis, NRC Office of Nuclear Regulatory Research, outlined specific changes in the above documents as follows:

- NUREG-0800-- SRP Chapter 18 was revised to : 1) make its format consistent with the format of NUREG-0711, 2) provide guidance for the review of HFE aspects of new plants, control room modifications, and modifications affecting risk-important human actions, and 3) provide guidance for a risk-informed, graded approach to HFE reviews of changes to human actions.

- NUREG-0711 was revised to: 1) make it applicable to all human factors reviews, not just advanced reactors, 2) make it single source of review procedures, and 3) to update the technical content of the individual elements to reflect the current state-of-the-art.

- NUREG-0700 was revised to: 1) address important human-system interaction (HSI) topics, such as controls and computer-based procedures, and 2) limit its content to HSI review guidelines and not the review process.

Ms. Susan Cooper, NRC Office of Nuclear Regulatory Research, stated that NUREG-1764 is a new document developed to: 1) provide a risk-informed screening method as a graded approach to human factors reviews that are commensurate with the risk importance of the human actions, and 2) consolidate in one document review guidance for changes to credited human actions.

Ms. Cooper indicated that the staff's review of license amendments and actions involving plant changes that affect important human actions (HAs) use a graded, risk-informed approach in conformance with Regulatory Guide 1.174, "An Approach to Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-specific Changes to the Licensing Basis."The staff's review uses a two-phase approach. The first phase is a screening analysis to determine the risk associated with the plant modification and its associated HAs using both quantitative and qualitative information. A graded risk-informed approach is used to determine the appropriate level of HFE review. This first phase has four steps: use of R.G 1.174 to determine the risk importance of the entire plant change or modification which involves human action; a quantification of the risk importance; a qualitative evaluation of HA; and an integrated assessment to determine the appropriate level of HFE review.

As a result of phase one, the proposed HAs are assigned to one of three risk levels (high, medium, low)In the second phase , HAs are reviewed using standard criteria in HFE to ensure that the proposed action can be reliably performed when required. Changes that involve more risk-significant HAs receive a detailed review, while those of moderate risk significance receive a less detailed review, and the HAs in the lowest risk region receive minimal HFE review.

Mr. Robert Fuld's comments

Mr. Fuld introduced himself as a private citizen and wished all a happy 50th anniversary of "Atoms for Peace." Mr. Fuld is currently certified as a Human Factors Professional by the Board of Certification in Professional Ergonomics. He has worked mainly in nuclear power since 1976, when he joined the Navy nuclear power program. He indicated that his statement is not regarding the technical contents, but it is rather a counterpoint to provide some balance.

Mr. Fuld stated that his concerns are related to Chapter 18.0 of the SRP, and the continued impact of the long-running NUREG-0711 initiative on its contents. His concerns are directed towards the human factors engineering program review model (PRM) used in NUREG-0711. He stated that the PRM generally promotes the interests of his profession to the detriment of the interests of the industry and the public goods, and the growing costs of these activities are often not matched by commensurate safety benefits. Mr. Fuld added that Chapter 18.0 SRP is being invited to incorporate, and thus validate, the essential rhetoric of NUREG0711This will bring NUREG-0711 a step closer to insinuating itself into the Federal Regulations. Mr. Fuld's statement is attached and it is part of the official record for the meeting. Mr. Rosen, Subcommittee Chairman, and members of the Subcommittee thanked and welcomed Mr. Fuld's comments.

General comments and observations from the Subcommittee members

- Mr. Rosen indicated that the update to Chapter 18.0 of the SRP seems to properly incorporate the needed changes.

- Mr. Rosen noted that approximately 50% or more of operating events are due to human factors and organizational weakness. He emphasized that the reliability of an organization is very important to the operation of nuclear power plants.

- Mr. Sieber cautioned that if the NRC staff would overly and fully standardize the human factors review process that could result in restricting the design process.

- Dr. Kress noted that future human reliability analysis (HRA) research program activities should include new knowledge for emerging needs such as latent conditions, advanced reactors, and external events.

- Mr. Rosen indicated that in the HRA arena, expert opinion is important and urged the NRC staff to seek such information.

- Mr. Rosen asked the staff to review the ACRS letter of November 13, 1995 regarding NUREG-0700, Revision 1, which expressed two concerns: 1) that the detailed HSI design review guidance in Part 2 may discourage the approval of other, equally acceptable alternatives, and 2) that the guidelines in Part 2 will become de facto regulations.

- Mr. Rosen urged the staff to review the ACRS letter dated September 24, 2002, and respond to such issues as : study human performance during severe accidents, evaluate if the ROP detects human performance degradation, generate guidance for the use of inspection and review tools, consider need for simulator devoted to research, investigate latent errors and how to treat in PRA, perform critical review of HRA models, and study team and individual performance in the context of plant organization.

SUBCOMMITTEE ACTION

This matter will be discussed further on December 4, 2003 during the ACRS meeting. The Committee expects to write a letter on this matter.

Documents provided to the Subcommittee prior to December 2, 2003:

- U.S. Nuclear Regulatory Commission Standard Review Plan, NUREG-0800, Chapter 18.0, "Human Factors Engineering," Draft Revision 2, December 2003.
- U.S. Nuclear Regulatory Commission, NUREG-0700, Revision 2, "Human-System Interface Design Review Guidelines," May 2002.
- U.S. Nuclear Regulatory Commission, NUREG-0711, Revision 2, "Human Factors Engineering Program Review Model."
- U.S. Nuclear Regulatory Commission, NUREG-1764, "Guidance for the Review of Changes to Human Actions," Final Report.

****

NOTE: Additional details of this meeting can be obtained from a transcript of this meeting available for downloading or viewing on the Internet at "http://www.nrc.gov/ACRSACNW" or can be purchased from Neal R. Gross and Co. (Court Reporters and Transcribers) 1323 Rhode Island Ave., NW, Washington, DC 20005 (202) 234-4433.

Dr. Maija Kukla, Program Director, Materials Research Science and Engineering Centers, Division of Materials Research, Room 1065, National Science Foundation, 4201 Wilson Boulevard, Arlington, VA 22230, Telephone (703) 292–4940.

Purpose of Meeting: To provide advice and recommendations concerning progress of Materials Research Science and Engineering Center.

Agenda:

December 4, 2003—Open for Director's overview of Materials Research Science and Engineering Center and presentations.

December 5, 2003—Closed to review and evaluate progress of Materials Research Science and Engineering Center.

Reason for Closing: The work being reviewed may include information of a proprietary or confidential nature, including technical information; financial data, such as salaries and personal information concerning individuals associated with the proposals. These matters are exempt under 5 U.S.C. 552b(c), (4) and (6) of the Government in the Sunshine Act.

Dated: November 18, 2003.

Susanne Bolton,

Committee Management Officer.

[FR Doc. 03–29151 Filed 11–18–03; 12:15 pm]

BILLING CODE 7555–01–M

NUCLEAR REGULATORY COMMISSION

[Docket No. 50–443]

FPL Energy Seabrook, LLC; Notice of Withdrawal of Application for Amendment to Facility Operating License

The U.S. Nuclear Regulatory Commission (the Commission) has granted the request of FPL Energy Seabrook, LLC (licensee) to withdraw its March 22, 2002, application for proposed amendment to Facility Operating License No. NPF–86 for the Seabrook Station, Unit No. 1, located in Rockingham County, New Hampshire.

The proposed amendment would have revised Technical Specification (TS) 3/4.9.13, "Spent Fuel Storage," and associated TS figures and index.

The Commission had previously issued a Notice of Consideration of Issuance of Amendment published in the **Federal Register** on May 14, 2002, (67 FR 34489). However, by letter dated September 15, 2003, the licensee withdrew the proposed change.

For further details with respect to this action, see the application for amendment dated March 22, 2002, and the licensee's letter dated September 15, 2003, which withdrew the application for license amendment. Documents may be examined, and/or copied for a fee, at the NRC's Public Document Room (PDR), located at One White Flint North, Public File Area O1 F21, 11555 Rockville Pike (first floor), Rockville, Maryland. Publicly available records will be accessible electronically from the Agencywide Documents Access and Management Systems (ADAMS) Public Electronic Reading Room on the internet at the NRC Web site, *http://www.nrc.gov/reading-rm/adams/html.* Persons who do not have access to ADAMS or who encounter problems in accessing the documents located in ADAMS, should contact the NRC PDR Reference staff by telephone at 1–800–397–4209, or 301–415–4737 or by e-mail to *pdr@nrc.gov.*

Dated at Rockville, Maryland, this 13th day of November, 2003.

For the Nuclear Regulatory Commission.

Victor Nerses,

Senior Project Manager, Section 2, Project Directorate I, Division of Licensing Project Management, Office of Nuclear Reactor Regulation.

[FR Doc. 03–29020 Filed 11–19–03; 8:45 am]

BILLING CODE 7590–01–P

NUCLEAR REGULATORY COMMISSION

Advisory Committee on Reactor Safeguards; Subcommittee Meeting on Planning and Procedures; Notice of Meeting

The ACRS Subcommittee on Planning and Procedures will hold a meeting on December 3, 2003, Room T–2B1, 11545 Rockville Pike, Rockville, Maryland.

The entire meeting will be open to public attendance, with the exception of a portion that may be closed pursuant to 5 U.S.C. 552b(c) (2) and (6) to discuss organizational and personnel matters that relate solely to internal personnel rules and practices of ACRS, and information the release of which would constitute a clearly unwarranted invasion of personal privacy.

The agenda for the subject meeting shall be as follows:

Wednesday, December 3, 2003–11:45 a.m.–1:15 p.m.

The Subcommittee will discuss proposed ACRS activities and related matters. The Subcommittee will gather information, analyze relevant issues and facts, and formulate proposed positions and actions, as appropriate, for deliberation by the full Committee.

Members of the public desiring to provide oral statements and/or written comments should notify the Designated Federal Official, Mr. Sam Duraiswamy (telephone: 301–415–7364) between 7:30 a.m. and 4:15 p.m. (ET) five days prior to the meeting, if possible, so that appropriate arrangements can be made. Electronic recordings will be permitted only during those portions of the meeting that are open to the public.

Further information regarding this meeting can be obtained by contacting the Designated Federal Official between 7:30 a.m. and 4:15 p.m. (ET). Persons planning to attend this meeting are urged to contact the above named individual at least two working days prior to the meeting to be advised of any potential changes in the agenda.

Dated: November 13, 2003.

Sher Bahadur,

Associate Director for Technical Support, ACRS/ACNW.

[FR Doc. 03–29017 Filed 11–19–03; 8:45 am]

BILLING CODE 7590–01–P

NUCLEAR REGULATORY COMMISSION

Advisory Committee on Reactor Safeguards; Meeting of the Subcommittee on Human Factors; Notice of Meeting

The ACRS Subcommittee on Human Factors will hold a meeting on December 2, 2003, Room T–2B3, 11545 Rockville Pike, Rockville, Maryland.

The entire meeting will be open to public attendance.

The agenda for the subject meeting shall be as follows:

Tuesday, December 2, 2003—1 p.m. until 5 p.m.

The purpose of this meeting is to review the proposed revisions to Standard Review Plan Chapter 18, "Human Factors Engineering." The Subcommittee will hear presentations by and hold discussions with representatives of the NRC staff, its consultants, and other interested persons regarding this matter. The Subcommittee will gather information, analyze relevant issues and facts, and formulate proposed positions and actions, as appropriate, for deliberation by the full Committee.

Members of the public desiring to provide oral statements and/or written comments should notify the Designated Federal Official, Dr. Medhat M. El-Zeftawy (telephone 301–415–6889), five days prior to the meeting, if possible, so that appropriate arrangements can be made. Electronic recordings will be permitted.

Further information regarding this meeting can be obtained by contacting the Designated Federal Official between 7:30 a.m. and 4:15 p.m. (ET). Persons

planning to attend this meeting are urged to contact the above named individual at least two working days prior to the meeting to be advised of any potential changes to the agenda.

Dated: November 12, 2003.

Sher Bahadur,

Associate Director for Technical Support, ACRS/ACNW.

[FR Doc. 03–29018 Filed 11–19–03; 8:45 am]

BILLING CODE 7590–01–P

NUCLEAR REGULATORY COMMISSION

Advisory Committee on Reactor Safeguards; Meeting of the Subcommittee on Plant License Renewal; Notice of Meeting

The ACRS Subcommittee on Plant License Renewal will hold a meeting on December 3, 2003, Room T–2B3, 11545 Rockville Pike, Rockville, Maryland.

The entire meeting will be open to public attendance.

The agenda for the subject meeting shall be as follows:

Wednesday, December 3, 2003—8 a.m.—11:30 a.m.

The purpose of this meeting is to discuss the Virgil C. Summer Nuclear Station license renewal application and the NRC staff's draft Safety Evaluation Report. The Subcommittee will hear presentations by and hold discussions with representatives of the NRC staff, South Carolina Electric and Gas Company, and other interested persons regarding this matter. The Subcommittee will gather information, analyze relevant issues and facts, and formulate proposed positions and actions, as appropriate, for deliberation by the full Committee.

Members of the public desiring to provide oral statements and/or written comments should notify the Designated Federal Official, Mr. Marvin D. Sykes (telephone 301–415–8716), five days prior to the meeting, if possible, so that appropriate arrangements can be made. Electronic recordings will be permitted.

Further information regarding this meeting can be obtained by contacting the Designated Federal Official between 7:30 a.m. and 4:15 p.m. (ET). Persons planning to attend this meeting are urged to contact the above named individual at least two working days prior to the meeting to be advised of any potential changes to the agenda.

Dated: November 13, 2003.

Sher Bahadur,

Associate Director for Technical Support, ACRS/ACNW.

[FR Doc. 03–29019 Filed 11–19–03; 8:45 am]

BILLING CODE 7590–01–P

SECURITIES AND EXCHANGE COMMISSION

[Release No. 34–48787; File No. SR–BSE–2003–17]

Self-Regulatory Organizations; Notice of Filing of Proposed Rule Change by the Boston Stock Exchange, Inc. Establishing Fees for the Proposed Boston Options Exchange Facility

November 14, 2003

Pursuant to section 19(b)(1) of the Securities Exchange Act of 1934 ("Act"),[1] and Rule 19b–4 thereunder,[2] notice is hereby given that on November 14, 2003 the Boston Stock Exchange, Inc. ("BSE" or "Exchange") filed with the Securities and Exchange Commission ("Commission") the proposed rule change as described in Items I, II, and III below, which Items have been prepared by the Exchange. The Commission is publishing this notice to solicit comments on the proposed rule change from interested persons.

I. Self-Regulatory Organization's Statement of the Terms of Substance of the Proposed Rule Change

The BSE seeks to enact fees for the proposed Boston Options Exchange ("BOX") facility. Proposed new language is *italicized.*

* * * * *

Fee Schedule

Sec. 1 Trading Fees for Public Customer Accounts

None.

Sec. 2 Trading Fees Broker Dealer Proprietary Accounts

a. *$0.20 per contract traded;*

—or—

b. $ 0.40 per contract traded against an order the Trading Host filters to prevent trading through the NBBO, pursuant to the procedures set forth in Chapter V, Section 16(b) of the BOX Rules.

c. Plus, where applicable, any surcharge for options on ETFs that are passed through by BOX. The applicable surcharges are as follows:

(1) $ 0.10 per contract for options on the ETF Nasdaq 100 ("QQQs").

Sec. 3 Market Maker Trading Fees

a. Per contract trade execution fee:

1. $ 0.20 per contract traded in assigned classes;

—or—

2. $ 0.20 per contract traded in unassigned classes;

-or-

3. $ 0.40 per contract traded against an order the Trading Host filters to prevent trading through the NBBO, pursuant to the procedures set forth in Chapter V, Section 16(b) of the BOX Rules.

4. Plus, where applicable, any surcharge for options on ETFs that are passed through by BOX. For a list of applicable ETF surcharges, see Section 2(c), above.

b. Minimum Activity Charge ("MAC")

The "notional MAC" per options class (see table below) is the building block for the determination of the BOX Market Maker's monthly total MAC which is payable at the end of each month if the per contract fee of $ 0.20 per contract traded, when multiplied by the Market Maker's actual trade executions for the month, does not result in a total trading fee payable to BOX at least equal to the monthly total MAC.

The MAC is totaled across all classes assigned to a Market Maker so that volume for one class is fungible against other classes for that Market Maker. As a result, although the volume on a given class needed to reach an implicit cost of $0.20 a contract may not be achieved, this can be compensated by volume in excess of the MAC on another class.

1. MAC "Levels."

The table below provides the MAC for each of the six "categories" of options classes listed by BOX. The category for each class is determined by its total trading volume across all U.S. options exchanges as determined by OCC data. The classifications will be adjusted at least twice annually (in January and July, based on the average daily volume for the preceding six month period).

Class category	OCC average daily volume (# of contracts)	MAC per market marker per appointment per month
A	<100,000	$15,000

[1] 15 U.S.C. 78s(b)(1).

[2] 17 CFR 240.19b–4.

ADVISORY COMMITTEE ON REACTOR SAFEGUARDS

SUBCOMMITTEE MEETING ON HUMAN FACTORS

DECEMBER 2, 2003
Date

NRC STAFF SIGN IN FOR ACRS MEETING

PLEASE PRINT

NAME	NRC ORGANIZATION
Richard Eckenrode	NRR/DIPM/IROB
Joel Kramer	RES/DSARC/REAHFB
Molly Keefe	RES/DSARE/REAHFB
Gareth Parry	NRR/DSSA
John Flack	RES/DSARE/REAHFB
Jose Ibarra	RES/DSARE/REAHFB
J Persensky	RES/DSARE/REAHFB
Paul Lewis	RES/DSARE/REAHFB
Jim Bongarra	NRR/DIPM/IROS

ADVISORY COMMITTEE ON REACTOR SAFEGUARDS

SUBCOMMITTEE MEETING ON HUMAN FACTORS

DECEMBER 2, 2003
Date

PLEASE PRINT

NAME	AFFILIATION
Jim Higgins	BNL
John O'Hara	BNL

UNITED STATES
NUCLEAR REGULATORY COMMISSION
ADVISORY COMMITTEE ON REACTOR SAFEGUARDS
WASHINGTON, D. C. 20555

January 14, 1994

The Honorable Ivan Selin
Chairman
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555

Dear Chairman Selin:

SUBJECT: FINAL REPORT ON THE USE OF THE DESIGN ACCEPTANCE CRITERIA PROCESS IN THE CERTIFICATION OF THE GENERAL ELECTRIC NUCLEAR ENERGY ADVANCED BOILING WATER REACTOR DESIGN

During the 405th meeting of the Advisory Committee on Reactor Safeguards, January 6-7, 1994, we completed our review of the Design Acceptance Criteria (DAC) to be included in the Certified Design Material (CDM) for the General Electric Nuclear Energy (GENE) Advanced Boiling Water Reactor (ABWR). The four subject areas addressed by DAC are Human Factors Engineering, Radiation Protection, Piping Design, and Instrumentation and Control.

Our Ad Hoc Subcommittee on DAC, in a joint meeting on November 2, 1993, with the Computers in Nuclear Power Plant Operations Subcommittee, reviewed Chapter 7, "Instrumentation and Control Systems," of the GENE Standard Safety Analysis Report (SSAR), the NRC staff Final Safety Evaluation Report (FSER) for this Chapter, and the related DAC. This DAC was further discussed during our November 4-6, 1993 meeting. Our ABWR Subcommittee, during its meeting of November 17, 1993, reviewed the human factors aspects of Chapter 13, "Conduct of Operations," and Chapter 18, "Human Factors Engineering," of the GENE SSAR, the NRC staff FSER for these Chapters and the related DAC for Human Factors Engineering. The DACs on Radiation Protection and Piping Design were discussed during our December 9-11, 1993 meeting. In each of these meetings, we had the benefit of discussions with representatives of the NRC staff and GENE. We also had the benefit of the documents referenced.

In addition to the meetings described above, both ACRS and its Ad Hoc Subcommittee on DAC (which was established to review the DAC process as requested by the Commission in its April 1, 1992 Staff Requirements Memorandum) met on a number of occasions to consider the overall DAC process as it was evolving. We provided two interim reports during this period. With this report, we believe that the Ad Hoc Subcommittee on DAC has now completed its assignment.

BACKGROUND

Since our last report, considerable effort has been expended by the NRC staff, GENE, NUMARC, and interested industry participants in the development of the Tier 1 CDM for the ABWR. As described in the GENE CDM submittal of December 7, 1993, the Tier 1 CDM relevant to the four subject areas that use the DAC process is contained in Section 3.0 "Additional CDM." This section consists of those aspects of the certified design that do not lend themselves to the system-by-system coverage provided in Section 2.0 of the CDM for individual plant systems. Each of the four DAC CDM sections consists of a Design Description and associated Inspections, Tests, Analyses, and Acceptance Criteria (ITAAC). Certain elements of these ITAAC are designated as DAC because they describe the design process to be used in implementing the design commitments stated in the Design Description. This is in contrast to the general case in which ITAAC will be used to confirm that the as-built plant systems have the design characteristics stated in the Design Description. Both the CDM and the associated Tier 2 material constitute the complete set of requirements for the certified design.

RECOMMENDATIONS AND COMMENTS

With respect to the material in Section 3.0 "Additional CDM" covering the four subject areas historically referred to as DAC, we are generally satisfied that it provides a reasonable basis for the staff final safety determination needed to support Final Design Approval. Our comments on each of these CDM are as follows:

Section 3.1 - Human Factors Engineering (HFE)

This section imposes Tier 1 requirements on the Combined Operating License (COL) holder with respect to the implementation of the human-system interface (HSI) for certified design. All six elements of ITAAC associated with this CDM have been designated as DAC by the staff and GENE.

Our review of HSI covered Chapter 18 of the FSER and the "HFE Program Review Model and Acceptance Criteria for Evolutionary Reactors," both dated December 1993. The latter document provides the technical basis for the staff review of the HFE design process proposed for certification. It also specifies the acceptance criteria by which the staff will evaluate the HFE program elements proposed by an applicant. We commend the staff for the development of this document. It provides much needed guidance to applicants on the staff expectations with regard to HFE for evolutionary reactors.

The HSI scope is limited to the main control room and the remote shutdown system. We commented, in our report of June 16, 1992, that the scope of the DAC then under development should be expanded

to include "... transmission switchyard work stations, because of the importance of offsite power to the safety of nuclear power plant operations" and "... incorporation of human factors principles in the design of local panels where instrumentation and controls important to safety are located." Although not included in this section of the CDM, we believe that these issues have been appropriately addressed elsewhere in the CDM.

Section 3.2 - Radiation Protection

This section imposes Tier 1 requirements on the COL holder with respect to the design of radiological shielding and ventilation systems. The scope of this section includes the design of these features for the Reactor Building, Turbine Building, Control Building, Service Building, and Radwaste Building. All six elements of ITAAC associated with this section have been designated as DAC by the staff and GENE.

The Design Description requires that the plant shielding design permit operators to perform required safety functions in "vital areas" of the plant under "accident conditions." The definition of "vital areas" in the Design Description differs from that in 10 CFR 73.2. We believe that other terminology should be used in this Design Description to avoid confusion with the definition used by the nuclear power plant security community.

ITAAC 3 of Table 3.2a contains the design commitment that "the plant shielding design shall permit plant personnel to perform required safety functions ... under accident conditions," and defines the accident radiation source term to be used for the shielding design. We agree that this source term is appropriate for this purpose.

Acceptance Criteria 1.a, b, and c of Table 3.2b distinguish, for purposes of ventilation system design, among "normally occupied rooms," "rooms that require infrequent access," and "rooms that seldom require access." The distinction between 1.b and 1.c is not obvious and should be more sharply drawn.

Section 3.3 - Piping Design

This section imposes Tier 1 requirements on the COL holder with respect to: (1) the design of nuclear safety-related piping systems and certain non-nuclear safety-related piping systems; (2) the analysis of the dynamic effects associated with postulated high energy pipe breaks on structures, systems, and components that are required to be functional during and following a safe shutdown earthquake; and (3) the reconciliation analysis of the as-built piping against the piping design. All three elements of this ITAAC have been designated as DAC by the staff and GENE.

The scope of this section is spelled out in the Design Description. There are, however, a number of additional aspects of piping design and analysis important to nuclear power plant safety which are not covered by this section. These have been discussed in detail with the staff and GENE on a number of occasions. We have been told that these piping design and analysis issues will be included elsewhere in the CDM. We will continue to follow this matter until we are satisfied that these issues have been properly addressed.

Section 3.4 - Instrumentation and Control

This section imposes Tier 1 requirements on the COL holder with respect to: (1) the configuration of safety-related digital instrumentation and control (I&C) equipment encompassed by the Safety System Logic and Control (SSLC); (2) the hardware and software development process used in the design, testing, and installation of I&C equipment; and (3) the diverse features included in I&C system design to provide backup support for postulated worst-case common-mode failures of SSLC. ITAAC 7 through 11 have been designated as DAC by the staff and GENE.

We would have preferred that the staff had based its review and acceptance of this section, the related Section 2.0, and SSAR Chapter 7 on a documented review model and specific acceptance criteria, as was done in the case for the Human Factors Engineering section discussed above. The staff has not yet formulated an identifiable set of criteria which must be met by digital I&C systems. In the FSER, reference is made to a menagerie of NRC regulations and regulatory guides, to a set of industry standards, and to several NRC publications which provide the basis for the staff conclusions concerning the process being followed by GENE. However, an examination of these indicates that most were developed before any significant application of digital technology to reactor safety systems, that only a few are relevant to many of the staff concerns, and that several are obsolescent if not obsolete.

We continue to recommend that the staff produce, on an expedited basis, a soundly conceived Standard Review Plan for digital I&C systems for both ALWRs and operating plant backfits.

Sincerely,

J. Ernest Wilkins, Jr.

J. Ernest Wilkins, Jr.
Chairman

References:

1. GE Nuclear Energy, "ABWR Certified Design Material," Volumes 1 and 2, December 7, 1993

2. GE Nuclear Energy, "ABWR Standard Safety Analysis Report," September 1993
3. Staff Requirements Memorandum from Samuel J. Chilk, Secretary of the Commission, to David A. Ward, ACRS Chairman, dated April 1, 1992, Subject: Periodic Meeting with the Advisory Committee on Reactor Safeguards on March 5, 1992
4. NRC staff Final Safety Evaluation Report for the General Electric Nuclear Energy Advanced Boiling Water Reactor, December 1993
5. NRC staff Final Safety Evaluation Report for the General Electric Nuclear Energy Advanced Boiling Water Reactor, "HFE Program Review Model and Acceptance Criteria for Evolutionary Reactors" (Appendix 18A), December 1993
6. ACRS report dated June 16, 1992, from Paul Shewmon, ACRS Chairman, to Ivan Selin, NRC Chairman, Subject: Interim Report on the Use of Design Acceptance Criteria in the Certification of the GE Nuclear Energy Advanced Boiling Water Reactor Design
7. ACRS report dated October 16, 1992, from Paul Shewmon, ACRS Chairman, to Ivan Selin, NRC Chairman, Subject: Second Interim Report on the Use of the Design Acceptance Criteria Process in the Certification of the General Electric Nuclear Energy Advanced Boiling Water Reactor Design

[Notes by Jerry Wachtel during ACRS meeting of November 2-4, 1995]

- George Apostolakos

Risk based - performance based rq.

* Does Admin Procedure Act apply to Rev 1 ?

- outline for 1023

* How build a firewall to prevent the electronic doc from being changed by abusers.

* How much $ — Apostolakos

* Comment re dedicated views for mirror image is in Pt 2 — is it within scope? Jett says was part of all 0700 —

AERS write letter?

- George Ap – wants to write a IP in an AERS letter
- It "inconceivable" that we still got this far without considering the AISA –
- J Carroll has written the boilerplate
- George's 2 IPs are due by end of July –

UNITED STATES
NUCLEAR REGULATORY COMMISSION
ADVISORY COMMITTEE ON REACTOR SAFEGUARDS
WASHINGTON, D. C. 20555

November 13, 1995

Mr. James M. Taylor
Executive Director for Operations
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555-0001

Dear Mr. Taylor:

SUBJECT: NUREG-0700, REVISION 1, "HUMAN-SYSTEM INTERFACE DESIGN REVIEW GUIDELINE"

During the 426th meeting of the Advisory Committee on Reactor Safeguards, November 2-4, 1995, we heard presentations by and held discussions with the NRC staff concerning the subject Design Review Guideline. We also had the benefit of the document referenced.

An outgrowth of the Three Mile Island accident was an NRC requirement that all licensees and applicants for commercial nuclear power plant operating licenses conduct detailed control room design reviews, including reviews of remote shutdown panels, to identify and correct design deficiencies related to human factors. Extensive guidelines published as NUREG-0700, "Guidelines for Control Room Design Reviews," were prepared to support these reviews.

The introduction of computer-based, human-system interface (HSI) technology into nuclear power plants prompted the development of Revision 1 to NUREG-0700. The objective of this document is to provide guidance to the NRC staff for HSI reviews of design submittals or as part of an inspection or other type of regulatory review.

The staff has developed technically defensible principles in Part 1 and a set of guidelines for HSI design reviews in Part 2. However, we are concerned that the detailed HSI design review guidance in Part 2 may discourage the approval of other, equally acceptable alternatives. Furthermore, we are concerned that the guidelines in Part 2 will become de facto regulations.

We plan to continue our review of the overall human factors program.

Sincerely,

T. S. Kress
Chairman

Reference:
U. S. Nuclear Regulatory Commission, NUREG-0700, Revision 1, "Human-System Interface Design Review Guideline," dated January 1995

ACRS Concerns
(11/30/95 - Revision 1)

In their letter of November 13, 1995, the ACRS identified two concerns with respect to the interpretation of the guidance provided in NUREG-0700, Revision 1. The letter stated: "We are concerned that the detailed HSI design review guidance in Part 2 may discourage the approval of other, equally acceptable alternatives. Furthermore, we are concerned that the guidelines in Part 2 will become de facto regulations." The staff agrees that these are legitimate concerns but were addressed in the development and evaluation of the document. First with respect to the ACRS's concern over discouraging design alternatives, the way in which guidelines were stated and the design review procedures for handling discrepancies, neither constrain design alternatives nor require strict adherence to the guidance. Second, with respect to the ACRS's concern over de facto regulation, the design review process and the manner in which the guidance is to be published indicate that it is not a regulation. Further, this concern was explicitly raised when the document was offered for public comment, and no comments expressing concern were received. Thus we feel that we have anticipated these concerns and addressed them in the documents development. A more detailed explanation of the staff's position on these concerns is provided below.

The purpose of NUREG-0700, Rev. 1 is to provide a basis for the NRC staff to review the human-system interfaces (HSIs) of nuclear plants. Considerable effort was made to ensuring that the HFE review guidance in Part 2 is based on valid, technically sound principles, as is indicated in the ACRS letter. Thus the guidelines are not arbitrary and deviations from them should be taken seriously by staff reviewers. However, the staff recognizes the need for flexibility in that technology changes rapidly and our knowledge about their effects on performance is continuously increasing. Thus NUREG-0700, Rev 1 uses wording within guidelines that indicates the are not requirements and has a review process that specifically identifies that deviations may be acceptable and identifies criteria for making such an evaluation. These are discussed below.

Guideline Wording

The use of the overall document and the interpretation of these individual guidelines is carefully indicated in the document. For example, in Section 1.3, Use of This Document (p. 1-7), it is stated that:

> "The review guidance provided in Parts 1 and 2 of this document is expressed using either 'should' or 'may.' The word 'should' is used to denote a recommendation; and the word 'may' is used to denote permission, neither a requirement nor a recommendation".

The word "must" is not used in Rev. 1 because the guidelines do not constitute requirements (These uses of the terms "must," "should," and "may" were adopted from ANSI). In the introduction to the guidelines in Part 2 it indicates that:

> "Each guideline contains a statement of a HSI characteristic with which the reviewer may judge the HSI's acceptability. The criterion is not a requirement and characteristics discrepant from the review criterion may be judged acceptable as per the procedures in the review process."

Another consideration that went into guidelines working was their level of detail. The review guidance in Part 2 of NUREG-0700 were written at a fairly general level of detail to provide for flexible approaches to design yet provide staff reviewers with the necessary information to identify aspects of the design that may negatively impact human performance. For example, Guideline 1.3.1-4, Character Size for Text Readability, specifies that a minimum size of 16 minutes of arc (MOA). Designers are not constrained to make the size 16 MOA. The human performance concern is that at smaller sizes, letters become more difficult to read which increases both reading time and errors.

As another example, Guideline 1.1-1 Display Screen Partitioning for HSI Functions, states that

> "A standard display screen organization should be evident for the location of various HSI functions (such as a data display zone, control zone, and message zone) from one display to another."

The guideline does not constrain designers to use a specific layout. It only indicates to the reviewer that if the location of such functions changes from display to display , it takes operators longer to search for desired functions and increases the chance of errors due to missing important information (such as system messages) because he did not precisely know where to find it.

Treatment of Alternatives

The staff will use Rev. 1 guidelines as potential indicators of HFE issues. Deviations from the criteria can be found acceptable so long as reasonable justification is provided. On page 4-7, it indicates that the design is evaluated as either acceptable or discrepant with respect to the guidelines. However, deviations from the guidance can be acceptable:

"Discrepancies could be acceptable within the context of the fully-integrated design. A guideline deviation, if sufficient justification exists, may not constitute an HED. The technical basis for such a determination could include an analysis of recent literature or current practices, tradeoff studies, or results of design engineering evaluations and data."

This point is reiterated on pages B-1 and B-2 of Appendix B, Review of an Applicant's HFE Guideline Document, it states that:

"Guidelines that are not derived from generic HFE guidelines may be justified by the applicant's:

- analysis of recent literature,
- analysis of current practices and operational experience,
- tradeoff studies and analyses, and
- the results of design engineering experiments and evaluations."

Further the staff encourages industry to develop its own HFE guidance. In Part 1, Section 4.2, HFE Design Verification, it states that:

"The criteria for HFE design verification are HFE guidelines. The guidelines used in the review depend on whether the applicant has developed a design-specific HFE guideline document. The development of a design-specific HFE guideline is a good practice since it can help ensure consistency, standardization, and the applications of HFE in the design. When such a document is available, it should be reviewed by the staff to ensure its acceptability. Procedures for such a staff review are described in Appendix B. The HFE guidelines in Part 2 of this document are used to support the staff's review of an applicant's HFE guidance document.

When applicant developed HFE guidelines are not available, the guidelines provided in Part 2 of this document can be used for the HFE design verification."

In Appendix B, Review of an Applicant's HFE Guideline Document, it states that:

"There is an increasing trend in the nuclear industry for applicants to create their own HFE

guidelines that reflect a tailoring of generic HFE guidelines to their specific designs. As indicated in Part 1, Section 4.2, HFE Design Verification, such a document can be reviewed by the staff (using this Appendix) and when any concerns are resolved, the applicant's guidance can be used as part of the basis for HFE Design Verification (see Section 4.2, p. 4-5, Item 3 - Guideline Selection). While this document does not make a specific recommendation that an applicant develop an HFE guideline, such a practice is desirable for a number of reasons. Use of an HFE guideline can help ensure that the design reflects good HFE principles and practices, and can promote consistency and standardization of the HSIs. Further, staff review of such a document can be performed before the design is finalized which will help ensure that the final design is acceptable to the NRC.

In summary, the staff has made a conscious effort to provide reviewers with a technical basis to recognize and flag potential issues, but has built a review process that acknowledges that alternatives may be acceptable and provides a means whereby a reviewer can make such an assessment.

With regard to ACRS's concern that the guidance will become de facto regulation, we feel the discussion above should clearly convey the means with which the staff will use and interpret the guidance.

Further, the NRC from the commission, to office directors, inspectors, to CRGR has taken great pains over the last decade to discourage the treatment as regulation, anything that is not specifically defined as a regulation. As per NUREG/CR-0070, "Guide to Types of NRC Formal Documents and Their Uses," NUREGs are staff reports on regulatory, technical, and administrative issues. One category of such reports is review plans and guidance, to which Rev. 1 belongs. This interpretation is supported by: (1) the fact that CRGR has declined to review the document; and (2) in offering NUREG-0700, Revision 1 for Public comment, the staff was encouraged by CRGR to request comment on its contention that the document does not impose backfit requirements on industry. Of the comments received, no concern over this issue was raised.

The Fiction of Function Allocation, Revisited

An invited manuscript for "Dialogues on Function Allocation"
A special issue of the International Journal of Human-Computer Studies
February 2000

Robert B. Fuld
Westinghouse Nuclear Automation
2000 Day Hill Rd.
Windsor, CT 06095-0500
USA

robert.b.fuld@us.westinghouse.com

Ph.# 860/731-6168

The Fiction of Function Allocation, Revisited

Robert B. Fuld[1]

Westinghouse Electric Corporation, 2000 Day Hill Rd., Windsor, CT 06095-0500, USA. Email: robert.b.fuld@us.westinghouse.com

Summary

In the human factors engineering literature, the function allocation concept has been a source of debate for decades, particularly in terms of its practical utility for general design. The present article revisits some fundamental criticisms of the hypothesized function allocation process, reviews related experience in the US nuclear power industry, and draws parallels to the histories of modern philosophy and science.

1. Introduction

Function allocation refers to the division of activities between humans and machines in a system. The concept was born in a research report on air traffic control edited by psychologist Paul Fitts (1951), along with the now eponymous list contrasting the abilities of humans and machines. Coincidentally, 1951 was the same year that electricity was first generated from nuclear power, under the direction of physicist Walter Zinn at the EBR-1 reactor in Idaho.

Who would have guessed that the two scientists' efforts were destined for an intimate rendezvous? Nuclear power would become a leading generator, not only of electricity but also of funds to search for optimal allocation methods. Yet, while nearly five decades of study suggest to some that, in practice, the function allocation process has been a failed hypothesis, others proclaim it to be proven and necessary for safe design. How is this chronic gulf possible?

The present article seeks to answer this question. To do so, the validity of Function Allocation (FA) as a design *process* will be challenged, both from the principled standpoint of its conceptual bases and from the practical standpoint of nuclear industry experience.

Apart perhaps from aviation, no industry has given FA a better chance to show its technical worth. Nuclear power thus yields important and often well-documented historical evidence that supports this author's earlier position (Fuld, 1993a). The present article extends this position in order to challenge the conventional wisdom of the human factors engineering community, and to support all industries faced with requirements of this type. Readers without

---

[1] The original version of this article appears in Proceedings of the First International Conference on Allocation of Functions, ALLFN'97, October 1-3, 1997, Galway, Ireland. It presents the opinions of the author, which are not necessarily those of his employer, the Westinghouse Nuclear Automation unit of the Westinghouse Electric Corporation, or its successors or assignees.

interest in the nuclear industry *per se* should nonetheless suffer these details with the prospect that the more general FA process will be uniquely revealed by the effort.

2. Ambiguities of Allocation

> "It is...easy to be certain. One has only to be sufficiently vague."
>
> Charles Sanders Pierce

In essence, FA transforms the man-machine system *concept* into a man-machine system *decomposition*. It weds the notorious man-machine dualism to the consideration of design options, while at the same time eschewing design details. Since "there is no such thing as an unmanned system" (Price, 1990), FA is thus both an ultimate and unconstrained explanation for design that wanders freely (if not always accountably) between different meanings:

- a hypothetical or actual *problem* (ill-considered FA)
- a design *process* to solve the problem (FA process or method)
- a design *product* or *resolution* of the problem (FA schemes)
- a human-initiated *change* to system mode (operator FA)
- an automatic mode change *feature* (dynamic/adaptive FA)

This pragmatical ambiguity reflects the disparate uses to which the FA concept-as-tool has been put. For example, it is easy to defend the cavalier claim that, "The general principles of allocation of functions have been well established by research which dates back some 40 years" (IAEA, 1992), *if* by that one means that a literature exists propounding the FA abstraction. Far less defensible is the claim that those principles have produced much practical benefit, considering the litany of caveats also established by that same literature. Indeed, the examples below reflect the added ambiguity of past and present, so that the more things change, the more they remain the same (including that more research is always needed):

- there is no adequate FA methodology (Swain & Wohl, 1961)
- there is no solid evidence to suggest an ideal allocation (Hopkins, et.al., 1982)
- equipment engineers lack a "traditional" FA process (Pulliam, et.al., 1983)
- Fitts Lists have had little impact on design practice (Price, 1985)
- Fitts Lists are the traditional FA process (Kantowitz & Sorkin, 1986)
- there is no commonly accepted FA methodology (Sheridan, 1996)

Allocation proponents need ambiguity to stress at once the importance and the inadequacy of the FA process, without loss of faith or admitted contradiction. But, critics find the formal allocation concept caught between a tautological rock and a self-contradictory hard place. On the one hand, since any artifact can be described in terms of functions, the notion of an implicit FA process is just an explanatory foil for a proposed explicit one (Fuld, 1993a). On the other hand, explicit or formal allocation (literally "placement," from the Latin, *locus*) suggests both the separation and comparability of man and machine, rather than their integration and complementarity (Jordan,

1963). It is thus ironic that the formal allocation process has been reified in the name of integration, i.e., by the Systems Approach to design.

3. Ethereal *A Priori*

> "It is a case of the stock rationalist trick of treating the *name* of a concrete phenomenal reality as an independent prior entity, and placing it behind the reality as its explanation."
>
> William James

Not unlike the Human Factors profession itself, the Systems Approach to Design (SAD) emerged as a defense industry development to support design ***innovation*** in the late '50s. Presenting a "top-down" hierarchy of requirements, SAD situated FA as a formal step within a larger design process. Since then, SAD and its contents have been offered as convenient, if not conventional, wisdom for other industries. For example, after ruing the non-impact of Fitts Lists on general design practice, Price turns with more optimism to the defense industry:

> "Some progress was achieved in the '60s and '70s as evidenced by the requirement for defining and allocating system functions in a Department of Defense specification, now MIL-H-46855B, Human Engineering Requirements for Military Systems, Equipment and Facilities..." (Price, 1985)

But, recalling the aforementioned methodological caveats, one might ask whether such allocation requirements were premature, rather than progressive. Indeed, only three years earlier, a literature survey for the US Nuclear Regulatory Commission (NRC) noted a basic lack of FA progress, *in terms of the same requirements*:

> "In spite of DOD regulations which specifically require allocation of functions as a step in the design cycle, no case was found in the literature in which the allocation of control functions had been determined in an orderly manner on a system-wide basis. This was true notwithstanding the fact that several methodological models had been developed and were available..." (Price, Maisano & Van Cott, 1982)

Since methods and requirements were already in place, this "failure to allocate tasks appropriately" was attributed to existing traditions and a lack of established procedures. The simpler notion that FA was a faulty concept (Jordan, 1963) was evidently not reconsidered. Instead, based on the procedural model of Price and Tabachnick (1968), the nuclear industry was provided with the elaborate and distinctly Cartesian procedure of NUREG/CR-3331 (Pulliam, et.al., 1983). In turn, that procedure has gone practically unused, suggesting perhaps that engineering traditions are a most stalwart barrier, but suggesting as well that certain comments on the method should be considered, at least, in the context of nuclear power plant design:

- the method is impractical (IAEA, 1992)
- the method is inappropriate (since it is innovative, not conservative; Bailey, 1982; MIL-H-46855A, 1972; Sec. 6.1)
- the method is unproven and provides no definitive results (Fuld, 1993a)

It has also been proposed that FA principles, while offering little support for the hypothesis phase of design, might better support the subsequent test phase (ibid.) If so, this would conflict with expectations once held for the method of Pulliam, et.al.:

> "This [NUREG/CR-3331] methodology provides the means by which a new plant or control system can be developed with a more effective (and safer) allocation of tasks to man and to automation. But it does not provide a means by which to evaluate an existing control room..." (Pulliam, 1983)

It seems reasonable to inquire how one can expect to engineer an adequate product from scratch while being unable to recognize similar inadequacies in existing products. However, the general scarcity of concrete, specific design breakdowns ever cited as misallocations suggests that the allocation concept suffers practical weakness from the test viewpoint, as well.

To achieve successful performance, the hypothetical-deductive FA model of Pulliam, et.al. (ibid.) depends on error feedback and iterative processing. It is thus a case of the more general closed-loop control model. The famous *Fitts Law* of skilled movement can be similarly formulated (Keele, 1968) to model a single voluntary motion as a series of increasingly accurate approximations or submovements. But, in practical execution, these submovements are transparent to the mover. What if, due to occasional clumsiness, you were required to make each submovement explicit? It would reduce your speed, but would it increase your accuracy? This model-driven decomposition fallacy suggests another reason that formal FA has been avoided by designers: Though allocations are *always* implicit in design, allocation models are not coherent prescriptions for explicit design behavior.

Thus, consistent with the pragmatic thesis, action demonstrates its essential primacy over reason in the design process. This condition, combined with the socio-economic and open-system complexity of design issues across their lifecycle, leads to the interminable result that top-down FA models cannot be efficiently developed and do not reliably predict the impact of hypothetical allocations on complex system performance. They have little but face validity, and are an artifact of rationalist understanding and explanation taking refuge in ambiguity. Otherwise, this process which engineers seem unwilling or unable to perform "appropriately" would have been (like it or not) increasingly allocated to more cooperative machines.

4. Pragmatic *Post Hoc*

> "The pragmatic method means...looking away from first things, principles, 'categories,' supposed necessities; and looking towards last things, fruits, consequences, facts."
>
> William James

In describing an allocation process as *a priori*, I mean one that claims to specify human and machine roles from the top-down, guided by principles and methods rather than practical needs and constraints. The motto of *a priori* FA could be Louis Sullivan's famed aesthetic that form follows function. Functional decomposition is a classic *a priori* method of design in general, and for allocation in particular.

In contrast, a *post hoc* FA process takes one or more received designs (e.g., concepts, models, simulations, prototypes, prior systems, etc.), from which an improved product is developed or the best option is selected. Its sadder but wiser motto, more typical of everyday engineering issues, could be that misfit follows form.

Despite claims to the contrary, scrutiny reveals purported FA processes as reliably *post hoc* in nature. Typical of such processes is a buildup of preliminary design tools and/or generic analytic processes (function analysis, tradeoff analysis, cost-benefit analysis, risk analysis, task analysis, workload analysis, resource leveling, etc.) around the ethereal FA step, which gains apparent substance and value by subsuming these more practical activities. But look and listen critically to the literature: The allocations themselves remain implicitly developed, either given to, conceived for, or presumed by the design.

Case in point--many FA guidelines begin by advising to allocate to the machine those functions that must be automated by regulation. But, this is neither process nor decision, but merely conformance: *Such allocations are already given.* In general, unanswered FA questions are successfully residualized by "traditional" design activities and constraints. Thus, it is normally best to take Occam's Razor and excise the formal FA "step;" preliminary design will progress unimpeded.

The superfluity of FA is borne out by two related practical trends. One is the continued scarcity of specific FA processes, not only in evolutionary, but in state-of-the-art design. Most notable in this regard is the dynamic allocation literature: While scaling revolutionary heights of man-machine integration, and while engaging a more complex (i.e., dynamic) set of allocation decisions, there is little mention of formal FA processing.[2] Likewise, recent design handbooks allow that the FA process may be expendable. One well-known example is ambivalent by chapter and author, proclaiming FA "a critical step in work system design" (Czaja, 1997), grandly touring its ambiguities (Sharit, 1997), ignoring it in the main discussion of automation (Woods, Sarter & Billings, 1997), and finally deriding it as a failed dichotomy:

---

[2] However, the notion that dynamic allocation actually *resolves* the static allocation problem (Hancock and Scallen, 1996) is perhaps an oversight, since there is no more basis for the new variety of dynamic system states than existed for the simpler static system. And though dynamic allocation will surely yield some useful (if costly) products, it is unlikely to reduce the burden of allocation on designers -- for who else will determine the added switching functions and be accountable for their behavior in the system?

> "It makes no sense to argue the case of 'manual control vs. automation.' One must try to see how to integrate both methods of process control. Indeed, it is misleading to put the question in the form of a balance between manual and automatic control. From a systems design viewpoint one should rather ask the question, 'What are typical tasks which humans are required to perform...and how can human skill be integrated...so that they may be best carried out?'" (Moray, 1997)

Moray's practical and holistic view suggests the other, related trend: The resilient appearance of references to *task* allocation (e.g., on consideration, most chapters in Beevis, Essens & Schuffel, 1996). While this could be just another ambiguity of allocation (i.e., in terms of level of abstraction and detail), I would argue that it is more significant. Task analysis (the essential human factors method) clearly presumes some representation of the operated system as input to the analysis. In contrast to FA, which aims to "keep the solutions at bay," (Price, 1985) task analysis seeks engagement with low-level details and structural particulars. A task, as opposed to a function, is a manageable unit, a pragmatic and action-oriented unit; indeed, *to allocate specific tasks is a natural human activity.* All of these things make it more viable to focus on concrete tasks, *post hoc*, than on abstract functions, *a priori.*

The fact that human-machine misfits are best identified at the task level is one reason that analysis of critical and representative tasks remains an essential design activity. Nonetheless, such descriptive analysis tends to follow the law of diminishing returns, so that beyond necessary and sufficient description lies an increasingly sub-optimal (i.e., inefficient) design process. Such practical realities may have been recognized by another of the famous contributors to Fitts (1951), who elsewhere offered a bluntly pragmatic and *post hoc* view of FA that remains as valid now as ever:

> "Many textbook accounts of this process assume the systems engineer goes through a deliberate cataloging of functions, followed by a careful and reasoned weighing of alternatives: "Should a man do this, or should a machine do it?" This rarely happens. The nature of systems engineering and the economics of modern life are such that the engineer tries to mechanize or automate every function that can be...This method of attack, counter to what has been written by many human factors specialists, does have a considerable amount of logic behind it...machines can be made to do a great many things faster, more reliably, and with fewer errors than people can...These considerations and the high cost of human labor make it reasonable to mechanize everything that can be mechanized. This means, however, that one important job of the engineering psychologist is to ensure that the jobs left over for human operators are within human capabilities." (Chapanis, 1970)

This practical, unpartisan statement reflected twenty years of experience beyond the Fitts List, but it is rarely cited in a positive light.

5. From Pragmatics to Proselytics: Humane Engineering

In 1995 I gave an overview of the man-machine interface design process to a group at the Korean Electric Power Research Institute. After a portion on the textbook FA process, one audience member looked intently dissatisfied. Finally he spoke: "This is not human factors methodology," he announced. "It is human factors *philosophy*." I could scarcely disagree. But if FA is only an engineering philosophy, then is it at least a coherent one? A practical one? One inclined to humane results? Indeed, it is not at all clear what philosophy is embodied by FA.

For if formal FA is at once a rationalist philosophical success and a pragmatic methodological failure, then it is radically ironic that FA is an outgrowth of American applied psychology. Because the American philosophy of *pragmatism*, championed in the new light of Darwinism for a young industrial age by psychologists William James and John Dewey, stood squarely opposed to rationalism.

The renewal of pragmatic, or more inclusively, *praxis* philosophies is a popular contemporary theme. The case of FA presented here demonstrates that praxical tenets can have substantive implications for design:

- the primacy of action/practice/synthesis/technology over reason/theory/analysis/science
- the essential nature of structural coupling and problem engagement in teleological behavior
- the predominant roles of history, context, interpretation and community in human action

These are key concepts in certain philosophical explorations of questions concerning technology and design (e.g, Coyne, 1995; Mitcham, 1994; and Winograd & Flores, 1986). The applicable point here is simply that less formal, *post hoc* FA methods can be practical because they are consistent with these praxical tenets. Conversely, highly formal, *a priori* methods cannot.

To go a step further, consider the humanist/rationalist dichotomy in Western intellectual history:

> "The 16th-century humanists were the founders of the modern Humanities just as surely as the 17th-century natural philosophers were founders of modern Science and Philosophy...If the Two Cultures are still estranged...it is a reminder that Modernity had two distinct starting points...What has yet to be explained is why these two traditions were not seen from the beginning as *complementary*, rather than in competition." (Toulmin, 1990; italics added)

Does any of this sound familiar? It clearly suggests the well-worn envy of social scientists (and engineers) for their more influential counterparts in physical science and engineering. It also may bear on the point that the hypothetical FA process represents, among other things, a recurring demand (by human engineers, to physical engineers) for integral participation in design. This humanist entry under rationalist cover reveals themes of power and politics, and echoes the course of FA discussions that focus on residual social issues (e.g., professional traditions, cultural influences, job satisfaction, etc.) in the name of *design* safety. Such quagmires of

conflicting human interests reprise the so-called 'wicked' design problem[3], and the infinite regress of technological costs and benefits to humanity.

But to expand the scope of the design problem is hardly to solve it. Thus, one is advised to take *neither* common sense judgment for practical human engineering, nor human factors theory for practical systems engineering. Let the buyer of either system beware.

Engineering is an innately pragmatic realm. As in other evolutionary processes, progress comes largely by building on acceptable solutions while continually weeding misfits. The motivations for higher quality and reduced costs, and the resulting tradeoffs among human interests and design alternatives, are largely implicit and somewhat inevitable. However, in evolutionary design, top-down FA will do more to disrupt than to improve the result. Instead, various *post hoc* approaches are necessary, due to the temporal human condition captured so well by Kierkegaard (i.e., that "life must be lived forward, but can only be understood in reverse.")

Thus John Henry battled the steam drill; Gary Kasparov, Deeper Blue. And in the present allocation battle? A standoff exists in which top-down FA is a doubly dichotomous appendage to the design process. Its philosophies are embraced, while its blue-sky methodologies are ignored; and engineers explain what is otherwise useful to do as somehow meeting the relevant requirements. This has been, until recently perhaps, the most practical and humanly acceptable thing to do.

6. FA and Nuclear Power Plant Design

Mishaps and perceptions notwithstanding, the US nuclear power industry has an outstanding safety record. Has formal FA played any part in this success? For an answer, some snapshots of related industry experience are presented.

Before TMI

Before the accident at Three Mile Island (TMI) Unit 2, FA was not widely held as a particular problem for Nuclear Power Plants (NPPs), but a rare example appears in Whitfield (1971):

> "Consider a proposed design for charging and discharging fuel elements from the reactor...[the operator] functioned for most of the time as a relay...*The human factors objections were not to the machine monitoring of the man, but to the way it was implemented*...Modifications were proposed to maintain the monitoring, but to make it less obvious...In addition, of course, the displays and controls of the console were arranged to represent the functional structure of the charge machine system." (italics added)

Note three points from the passage. First, it presents a classic case of *post hoc* design review and iteration. Second, the human engineering described fits well with the pragmatic views of Moray and Chapanis given earlier. Third, and most important, while tasks and design details were revised, *no problems or changes in the original allocation of control functions were suggested.*

[3] A wicked design problem (Rittel, 1972) is a class of social system problems which are ill-formulated, have many decision-makers with conflicting values, in which every problem is a symptom of a higher-level problem, and in which information and its ramifications are complex and confusing.

TMI and Allocation

Before TMI, Kantowitz (1974) correctly observed that, "the most likely outcome of a serious operator error would not be the release of radioactive material thus endangering the public but instead would be the economic loss resulting from a sudden scram." Though TMI-2 was both a less likely and more serious event, it too was limited to economic loss. Multiple levels of system defense-in-depth maintained the health and safety of the public. Nonetheless, TMI is a textbook example of how small (i.e., detailed) design weaknesses may contribute to major operating consequences, particularly under degraded or unanticipated conditions. It is thus interesting to consider the following excerpt from a well-known reference:

> "In the Three Mile Island accident the automatic safety systems worked as designed and turned on the emergency pumps. The human operators erred and manually turned off the pumps. Therefore, one possible conclusion is that the human...role should be minimized. This will, on the average, decrease the opportunity for human error and so increase system reliability. Furthermore, it will lower operating costs...Another philosophy is to keep the human operator involved as much as possible...in case something unexpected goes wrong...*We can't tell the reader which philosophy to believe* or which a company should adopt. However...*we believe that the human factors expert has an obligation to provide meaningful work for people.* This leads us to the next topic, dividing work between people and machines." (Kantowitz & Sorkin, 1987; italics added)

Consider first that in a true *systems* approach, the two philosophies do not form a dichotomy but a continuum of potentially acceptable alternatives. Next, note the turn of attention from system reliability and economy to meaningful human work; though all noble goals, the latter may be the less humane in the view of NPP neighbors and ratepayers. Ironically, this plausible net inhumanity serves as point of departure for the failed methodology of dividing functions. But most of all, consider that *of the many failures and deficiencies found post hoc at TMI, conclusive misallocations were few if any.* In fact, the most significant changes to automatic control addressed flaws in the logic (not the allocation) of containment isolation functions. So while man-machine integration was improved, the balance of human and machine was not changed.

In the broader industry, TMI led to new regulations, including those for Detailed Control Room Design Review (DCRDR) as a condition for NPP licensing and operation. The DCRDR mandate, which led to hundreds of thousands of man-hours and millions of dollars worth of human engineering scrutiny, was a windfall for the human factors profession. To direct the process, DCRDR guidelines (NUREG-0700, 1981) were developed, including SAD, FA and a Fitts List (Exhibit B-3; see Figure 1). And despite the basic irrelevance of SAD to DCRDR, Exhibit B-3 at least gave criteria by which allocations in the existing designs could be challenged, if necessary. Nonetheless, while DCRDRs of the entire NPP population did produce thousands of human engineering discrepancies, the hypothesized "ill-considered allocations" of traditional/implicit FA scarcely materialized. Could it be, they weren't there to be found?

[Place **Figure 1** here →] [← Place **Figure 1** here]

Advanced Designs and FA

In 1985, the utility industry embarked on the development of design requirements for the next generation of evolutionary NPPs, the Advanced Light Water Reactor (ALWR) program. The ALWR requirements specified standard plants with improved safety and economy, incorporating industry experience, lessons learned, and modern digital technology. In 1987 the NRC adopted a policy permitting pre-approval of standard plant designs, commencing the era of "one-step" design certification. Little progress occurred, however, until the NRC issued schedules for ALWR review in 1991.

Three vendors submitted designs under the new one-step process: Combustion Engineering (ABB), General Electric (GE), and Westinghouse. GE was the first formal applicant, and so became the lead for standard plant licensing. However, the old DCRDR method of human factors review was now problematic, as the designs submitted for pre-approval were less detailed and inspectable than as-built facilities. Thus a new method was developed, one for reviewing human factors engineering *programs*. The first drafts of this Program Review Model (PRM) appeared early in 1992 for the GE review. Though our interest here is in the PRM's treatment of FA, note first its grounding in classic SAD presumptions:

> "A generic HFE Program Model was developed based largely on applied general systems theory and the Department of Defense (DOD) system development process...the military has applied HFE for the longest period of time (as opposed to industrial, commercial, or other users), thus, the process is highly evolved and formalized and represents the most highly developed model available." (O'Hara & Higgens, 1992)

The draft PRM had eight elements, including functional requirements analysis, followed by FA, followed in turn by task analysis. With minor modifications, GE accepted all eight elements as licensing commitments for later design *process* activities. In contrast, ABB felt that its design *products* already met the aim of several of these process elements. The need for top-down FA was particularly opposed:

> "ABB-CE...stated that full analyses of functional requirements and function allocation are not necessary because the System 80+ design is an evolution of the System 80 design that was previously reviewed and approved by the NRC and has an operating history...[and] that the definition and allocation of functions for the System 80+ design are largely unchanged from that of its predecessor...The staff agreed, and the HFE PRM criteria were modified accordingly." (NUREG-1462, 1994)

For the licensing process, ABB addressed the PRM's functional requirements analysis and function allocation elements as one, and provided a *post hoc* description of the existing, proven design functions and their control schemes (Fuld, 1993b). This included:

- identification of federally mandated allocations
- review of regulatory and industry allocation guidance
- description of critical safety functions in the design
- description of the operators' role in controlling safety functions & related equipment
- rationale for assigned allocations
- identification of relevant changes from predecessor designs

This approach had sufficient merit for the PRM, in modifying its criteria, to follow ABB suit by collapsing the two functional design elements and by adding the evolutionary option (NUREG-0711, 1994)[4]. Further validation came two years later, when Westinghouse used the evolutionary option to defend its new passive design (an effort which bears acknowledged similarity to the original ABB approach). In doing so, their philosophy was commendably pragmatic:

> "In the Westinghouse design process, functional requirements analysis and preliminary function allocation are largely the responsibility of system designers." (Brockhoff, et.al., 1996)

Of course, it helps to have a precedent. Completely overlooked was one from the post-TMI era, in which an expert human factors panel clearly (if begrudgingly) told the NRC that evolutionary designs don't necessarily need formal allocation:

> "Of all these unacceptable bases for function allocation, only the basis of tradition has any merit...Tradition may provide efficient and reasonably useful guidelines for preliminary gross function allocation...for a system that represents an evolutionary design change...It is not a reliable or valid guide when there are revolutionary changes in system design (as, for example, in the change in design from use of fossil fuel heated boilers to nuclear steam supply systems)." (Hopkins, et.al., 1982)

Actually, these authors could hardly claim to have reliable or valid guides at *either* end of the spectrum, but at least they recognized evolutionary change to be the lesser design problem. (On the other hand, the fact that commercial NPPs in the US evolved *literally* from Rickover's proven shipboard designs was completely missed.) But with no call for FA in the post-TMI era (or perhaps because it was antithetical to formal allocation), the evolutionary 'option' was soon forgotten.

7. Human Engineering Philosophy

> "With the elucidations such "engineers" will give...everyday problem[s] will become incomprehensible...[the notion] that one needs no detailed expertise in a given field, that he can..."solve" a problem by the exercise of his intellect and the use of concepts...may be true in pure science; it certainly is not in engineering. To advocate the contrary demonstrates a lack of insight on how engineering problems are actually solved."
>
> Hyman Rickover

On balance, engineering philosophies must be pragmatic to avoid self-contradiction. This can be no different for human engineering. Within the human factors profession itself, there have long been calls for a more pragmatic focus befitting its practice and practitioners. In contrast, formal

[4] Likewise, the developers of IEC 1839 (1998), a proposed standard specifying procedures for "function analysis and assignment" in NPP control room design, are urged to add the evolutionary/descriptive option to their document.

function allocation is literally a textbook case of the inertia and opposition of theoretical forces to practical reality.

How can a more pragmatic focus for human factors be achieved? If change is sought truly in *practice,* I suspect that, like allocation, it is more likely to occur "bottom-up." To that end, individuals so inclined may find the literature of pragmatism useful (e.g., Diggens, 1994). As a case in point, consider the design process in terms of four dimensions based on Toulmin (1990):

Humanist - Rationalist
informal/oral - formal/written
concrete/particular - abstract/general
diverse/local - uniform/global
timely/transitory - timeless/permanent

The implicit need for balance on these dimensions should be familiar to most engineers. However, for human factors specialists in particular, a *leftward* adjustment in their professional approach is often needed. Paradoxically, in doing so one ventures *away* from general principles of the human, towards engagement with the details of particular *systems*. It may not be entirely circular to suggest that this sort of pragmatic shift should be considered by specialists who seek greater participation in the local design process community.

8. Pragmatic Design Guidance

With the era of one-step NPP licensing, the FA process has become subject to formal design review. What practical guidance can be offered to industry designers? Presently, three references are noted, with caution, for critical readers.

The Descriptive Evolutionary Approach of NUREG-0711 (1994)
Its background proselytics and methodological references aside, Element 3 (Functional Requirements Analysis and Function Allocation) of NUREG-0711 allows the designer to present and defend a design in terms of functional design and allocation issues without a theory-based FA process. The scope of Element 3 review is limited to control of plant safety functions, it adds no unnecessary terminology or structure, and its text is a commendably brief five pages. Key features of the approach include 1) reliance on descriptive analysis; 2) a reduction in the recommended scope of analysis; 3) emphasis on retaining design rationale; and 4) clear use of related predecessor designs for either proof-of-concept or basis-for-revision. A schematic of Element 3 is shown in Figure 2.

[Place **Figure 2** here →] [← Place **Figure 2** here]

The Allocation Process Approach of IAEA-TECDOC-668 (1992)

This document, based directly on NUREG/CR-3331, offers an iteration of the classic hypothetical-deductive method. It is thus subject to most of the criticisms in the present article. Nonetheless, from the more practical standpoint it espouses, TECDOC-668 betters its predecessor by 1) its accumulated experience with a wide range of successful NPP operating philosophies; 2) a shift in emphasis towards evolutionary justification, and away from process-driven generation, of safe design; and 3) a simple two-thirds reduction in length. The document still suffers from FA's dichotomies; and though its different sections show conflicting philosophical perspectives, this at least adds some balance to the overall product. Its struggle for practicality, it is hoped, will have further success in any future revisions.

The Time Response Analysis Approach of ANSI/ANS 58.8 (1994)
This nuclear industry standard was first published in 1984. Actually begun prior to TMI, its self-described purpose is to provide:

> "time response design criteria for safety-related operator actions to be...used to determine the minimum response time intervals [allowable]...to receive credit in the safety analysis for operator actions that initiate or control safety-related functions...[thus] determining whether a particular action...might be accomplished by operator action or must be accomplished by an automatic action." (ibid.)

Incorporating results from numerous studies of simulator and event data, ANS 58.8 is the most relevant and validated method available for resolving a safety-related allocation question early in NPP design. In terms of human factors, it offers the plant safety analyst not specialist expertise, but instead a very crude (i.e., *conservative*) modeling assumption: That human actions each require one minute to be performed.

Thus, event analysts must explicate a basic task analysis to show sufficient time available for manual safety actions. And while this simplistic but practical model is easily criticized for imprecision, it has yet to be faulted for conservatism: Empirical studies have confirmed that its results reliably envelop typical operator performance. It is thus a final irony that ANS-58.8 has been virtually ignored by the industry's own FA literature. Perhaps its simplicity had no appeal for human factors theorists. Or perhaps, lacking the FA label, its obvious relevance was missed.

9. General Recommendations

> "The tool...must, as it were, withdraw in order to be ready-to-hand quite authentically... obtrusiveness and obstinacy...bring to the fore the characteristic of presence-at-hand in what is ready-to-hand."
>
> Martin Heidegger

From a design process standpoint, no mere turn will solve the problems of FA, which are as much ontological as technological. Thus, I will not offer alternative methods, but rather a bit of guidance in terms of the tenets of praxis proposed earlier.

Action-Orientation

Engineering requires efficient and effective action. The benefits of iterative design, for example, depend on effective management, prompt closed-loop communications, efficient design change control, and strategies to reduce the number and impact of iterations. In the human factors area, though formal experimentation is occasionally quite valuable, it is modeling & prototyping that have become the tools of choice (precisely because modern methods permit rapid design *action*). The interest in articulation work (human-human task allocation) and in dynamic allocation both focus on concrete activities (work, design) rather than on abstractions. And of course, the classic guidance for improved action is to simplify the process or product; conversely, their complication must be justified by the expectation of clear benefits.

History, Interpretation and Context

From the *post hoc* perspective, the importance of historical context for engineering in general and FA in particular is clear. This suggests the value of developing and retaining the bases or rationale for a design, including the intended role of the operator. Besides their eventual retrospective value, bases provide an opportunity to project responsibility into future contexts for key considerations of the original design. This suggestion is consistent not only with the descriptive evolutionary approach described earlier, but also with the recommendations of formalists such as Price & Tabachnick, (1968) and in proper measure, surely with common sense as well.

Of course, such linguistic descriptions are innately incomplete, ambiguous, and context reliant, following, as noted earlier, the law of diminishing returns. But, to be accountable, meaningful statements must at least avoid conflicting ambiguity. On the subject of FA, therefore, resolve to neither take nor grant refuge in ambiguity. This suggests general avoidance of the term "FA" in new design activities, and its replacement by more specific terms that uniquely distinguish its various pragmatic incarnations as a philosophy, a problem, a product...or if need be, a process.

Structural Coupling, Engagement, and Community

The common focus of human factors professionals in all industries is symbolized by task analysis. Task analysis activity in its myriad forms reflects engagement with the environmental context, the phenomena, the language, and the actions of humans with machines. Task analysis is thus praxical (and also *post hoc*). However, beyond this immediate point of interest in technology and its users, many human factors specialists have found it difficult to engage with the broader engineering sphere. One component that is often lacking is engagement with, (i.e., knowledge of) the systems and equipment in use. Such knowledge provides not only better understanding of the user's tasks and of the potential breakdowns of the man-machine system, but also a key basis for communication and involvement across disciplines; its value cannot be overstated. Finally, Rechtin (1997) has done much to provide a praxical design paradigm that joins real-world project engineering (including the primacy of form and structure in design) with the contemporary traditions of systems engineering. He calls this paradigm Systems Architecting.

10. Conclusion

> "After such bad experiences, this is the moment to forget ether completely and try never to mention its name."
>
> Einstein & Infeld

Wirstad (1979) has summed up the FA process succinctly: "Although the principle is clear, function allocation has never worked in practice." Fundamental reasons for this failure have been proposed elsewhere by Jordan (1963) and Fuld (1993a). The present article has gone further to claim for the US nuclear power industry, at least, that the resulting "fatal mistakes" (Wourms & Rankin, 1994) of the implicit FA process are literary folklore. The body of actual misallocations has never been found, even after the unprecedented human factors scrutiny of the post-TMI control room design reviews. Against this background, a *post hoc*, evolutionary approach to FA issues has been shown to be compatible with practical engineering activities, and to be acceptable for standard plant licensing. For the nuclear power industry, these facts make the necessity and sufficiency of theoretical FA methods a moot point, at best.

I conclude that the chronic gulf between proponents and critics of function allocation is possible because FA is a useful theory but not a practical method. If either its theory becomes *less* useful, or its methods more practical, then the gulf will be reduced. This article has tried to accomplish a bit of both.

I will close by returning to Jordan's example of a faulty concept--the "ether" of 19th century physics--to extend his analogy between function allocation and the ether. Both concepts received decades of relatively fruitless study. Both concepts were challenged by compelling null results (in the case of the ether, by those of the famous Michelson-Morley experiment). Both concepts led to confrontations of rationalist theory and pragmatic observation. And both concepts we've been advised to discard. The analogy ends here, though; for while history has judged for Einstein on the ether, readers must yet judge FA for themselves.

References

ANSI/ANS 58.8-1994 (1994). Time Response Criteria for Safety-related Operator Actions. La Grange Park: American Nuclear Society.

Bailey, R.W. (1982). Human Performance Engineering. Englewood Cliffs: Prentice Hall.

Beevis, D., Essens, P.J.M.D. & Schuffel, H. (Eds.), (1996). Improving Function Allocation for Integrated Systems Design. Report No. CSERIAC SOAR 96-01. Wright Patterson AFB: Crew Systems Ergonomics Information Analysis Center.

Brockhoff, C.S., Mumaw, R.J., Roth, E.M. & Schulz, T.L. (1996). AP600 Functional Requirements Analysis and Function Allocation. Report No. WCAP-14644. Pittsburgh: Westinghouse Electric Corp.

Chapanis, A. (1970). Human Factors in Systems Engineering. In K.B. De Greene. Ed. Systems Psychology, pp. 51-78. New York: McGraw-Hill.

Coyne, R. (1995). Designing Information Technology in the Postmodern Age. Cambridge: MIT Press.

Czaja, S.J. (1997). Systems Design and Evaluation. In G. Salvendy, Ed. Handbook of Human Factors and Ergonomics, pp. 17-40. New York: Wiley-Interscience.
Diggens, J.P. (1994). The Promise of Pragmatism. Chicago: University of Chicago Press.
Einstein, A. & Infeld, L. (1942). The Evolution of Physics. New York: Simon and Schuster.
Fitts, P.M. (Ed.), (1951). Human engineering for an effective air navigation and traffic control system. Washington: National Research Council.
Fuld, R.B. (1993a). The fiction of function allocation. Ergonomics in Design, 1(1), January; pp. 20-24.
Fuld, R.B. (1993b). Human factors evaluation and allocation of System 80+ functions. Report No. NPX80-IC-RR790-02. Windsor: Combustion Engineering, Inc.
Gray, L.H. & Haas, P.M. (1983). Criteria for Safety-related Operator Actions. Presentation at the Eleventh Water Reactor Safety Research Information Meeting, October 24-28, 1983. Washington: USNRC.
Hancock, P.A. & Scallen, S.F. (1996). The future of function allocation. Ergonomics in Design, 4(4), October, pp. 24-29.
Heidegger, M. (1927). Being and Time. Tr. by John MacQuarrie & Edward Robinson. New York: Harper & Row, 1962.
Hopkins, C.O., Snyder, H.L., Price, H.E., Hornick, R.J., Mackie, R.R., Smillie, R.J. & Sugarman, R.C. (1982). Critical human factors issues in nuclear power regulation and a recommended comprehensive human factors long-range plan, Vol. 2. Report No. NUREG/CR-2833. Washington: USNRC.
IAEA-TECDOC-668 (1992). The Role of Automation and Humans in Nuclear Power Plants. Vienna: International Atomic Energy Agency.
IEC 1839 (1998). Function Analysis and Assignment. Working Group A8 draft standard. Geneva: International Electrotechnical Commission.
James, W. (1907). Pragmatism: A new name for some old ways of thinking. New York: Dover, 1995.
Jordan, N. (1963). Allocation of functions between man and machines in automated systems. Journal of Applied Psychology, 47(3), pp. 55-59.
Kantowitz, B.H. (1974). Human information processing: Tutorials in performance and cognition. Hillsdale: Lawrence Erlbaum Associates.
Kantowitz, B.H. & Sorkin, R.D. (1987). Allocation of functions. In G. Salvendy, Ed. Handbook of Human Factors. New York: Wiley.
Keele, S.W. (1968). Movement control in skilled motor performance. Psychological Bulletin, 70(6), pp. 387-403.
Mitcham, C. (1994). Thinking Through Technology. Chicago: University of Chicago Press.
MIL-H-46855A (1972). Human engineering requirements for military systems, equipment, and facilities. Washington: Department of Defense.
Moray, N. (1997). Human Factors in Process Control. In G. Salvendy, Ed. Handbook of Human Factors and Ergonomics, pp. 1944-1971. New York: Wiley-Interscience.
NUREG-0700 (1981). Guidelines for control room design reviews. Washington: USNRC.
NUREG-0711 (1994). Human factors engineering program review model. Washington: USNRC.
NUREG-1462 (1994). Final safety evaluation report related to the design certification of Combustion Engineering System 80+. Washington: USNRC.

O'Hara, J. & Higgins, J. (1992). Interim Human Factors Review Criteria for the Design Process of an Advanced Nuclear Power Reactor (draft). Report No. L2314-3-4/92. Upton: Brookhaven National Laboratory.

Pierce, C.S. (1958). Collected Papers of Charles Sanders Pierce, Vol. 4. Cambridge: Harvard University Press.

Price, H.E. (1985). The allocation of function in systems. Human Factors, 27(1), pp. 33-45.

Price, H.E. (1990). Conceptual system design and the human role. In H. Booher, Ed. MANPRINT: An approach to system integration. New York: Van Nostrand Reinhold.

Price, H.E., Maisano, R.E. & Van Cott, H.P. (1982). The Allocation of Functions in Man-machine Systems: A perspective and literature review. Report No. NUREG/CR-2623. Washington: USNRC.

Price, H.E. & Tabachnick, B.J. (1968). A Descriptive Model for Determining Optimal Human Performance in Systems, Vol. III: An approach for determining the optimal role of man and allocation of functions in an aerospace system. Report No. NASA CR-878. Serendipity Associates.

Pulliam, R. (1983). Allocation of functions. Eleventh Water Reactor Safety Research Information Meeting, October 24-28, 1983. Washington: USNRC.

Pulliam, R., Price, H.E., Bongarra, J., Sawyer, C.R. & Kisner, R.A. (1983). A methodology for allocating nuclear power plant control functions to human or automatic control. Report No. NUREG/CR-3331. Washington: USNRC.

Rechtin, E. (1997). The Synthesis of Complex Systems. IEEE Spectrum, 34(7), 51-55.

Rickover, H.G. (1970). Memorandum on a proposed human factors program, quoted in the February, 1977 issue of the Human Factors Society Bulletin, 20(2), pp. 1-2.

Rittel, H.W.J. (1972). On the planning crisis: Systems analysis of the first and second generations. Bedriftsokonomen, 8, pp. 390-396.

Sharit, J. (1997). Allocation of Functions. In G. Salvendy, Ed. Handbook of Human Factors and Ergonomics, pp. 301-339. New York: Wiley-Interscience.

Sheridan, T.B. (1996). Allocating functions among humans and machines. In D. Beevis, P.J.M.D. Essens & H. Schuffel, (Eds.), (1996). Improving Function Allocation for Integrated Systems Design. Report No. CSERIAC SOAR 96-01. Wright Patterson AFB: Crew Systems Ergonomics Information Analysis Center.

Swain, A.D. & Wohl, J.G. (1961). Factors Affecting Degree of Automation in Test and Checkout Equipment. Report No. D&A-TR-60-36F. Stamford: Dunlap & Associates.

Toulmin, S. (1990). Cosmopolis. Chicago: The University of Chicago Press.

Whitfield, D. (1971). Human skill as a determinate of allocation of function. In W.T. Singleton, R.S. Easterby & D.C. Whitfield, Eds. The Human Operator in Complex Systems, pp. 154-160. London: Whitfield, Taylor & Francis.

Winograd, T. & Flores, F. (1986). Understanding Computers and Cognition. Reading: Addison-Wesley.

Wirstad, J. (1979). On the Allocation of Functions between Human and Machine. Report No. 13. Karlstad: Ergonomrad.

Woods, D.D., Sarter, N. & Billings, C. (1997). Automation Surprises. In G. Salvendy, Ed. Handbook of Human Factors and Ergonomics, pp. 1926-1943. New York: Wiley-Interscience.

Wourms, D.F. & Rankin, J.W. (1994). Computer-based procedures. Report No. CSERIAC-RA-94-002, for USNRC. Wright Patterson AFB: Crew Systems Ergonomics Information Analysis Center.

Humans Excel In	Machines Excel In
Detection of certain forms of very low energy levels	Monitoring (both personnel and equipment)
Sensitivity to an extremely wide variety of stimuli	Performing routine, repetitive, or very precise operations
Perceiving patterns and making generalizations about them	Responding very quickly to control signals
Detecting signals in high noise levels	Exerting great force, smoothly and with precision
Ability to store large amounts of information for long periods—and recalling relevant facts at appropriate moments	Storing and recalling large amounts of information in short time periods
Ability to exercise judgment where events cannot be completely defined	Performing complex and rapid computations with high accuracy
Improvising and adopting flexible procedures	Sensitivity to stimuli beyond the range of human sensitivity (infrared, radio waves, etc.)
Ability to react to unexpected low-probability events	Doing many different things at one time
Applying originality in solving problems: i.e., alternative solutions	Deductive processes
Ability to profit from experience and alter course of action	Insensitivity to extraneous factors
Ability to perform fine manipulation, especially where misalignment appears unexpectedly	Ability to repeat operations very rapidly, continuously, and precisely the same way over long periods
Ability to continue to perform when overloaded	Operating in hostile environments which are hostile to humans or beyond human tolerance
Ability to reason inductively	

Figure 1. Exhibit B-3: Human/machine capabilities (NUREG-0700, 1981)

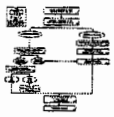


Figure 2. Functional Requirements Analysis and Function Allocation (NUREG-0711, 1994)

RM T-2B3

Good Afternoon. I wish you all a happy 50th Anniversary of "Atoms for Peace" (actually next Monday I believe). If I may introduce myself, my name is Robert Fuld, and I am a human factors professional in the Nuclear Industry. I am currently Certified as a Human Factors Professional by the Board of Certification in Professional Ergonomics. I have a graduate degree in Industrial Engineering from the University of Illinois, and I'm an active member of several professional societies including the IEEE NPEC Subcommittee 5 on Human Factors and Reliability. I've worked mainly in nuclear power since 1976, when I joined the Navy Nuclear Power Program, AND and I'm now employed in Windsor CT by the Westinghouse Electric Company.

While I am observing today's meeting on behalf of my employer, let me be clear that I am making the following statement as a private individual and as an independent member of my profession and my industry. Therefore, I am entirely responsible for the views I express here today, and they should not be taken to reflect the views of my employer or any of my professional affiliates, aforementioned or otherwise, now or in the future.

That said, my statement concerns Chapter 18 of the Standard Review Plan (SRP), and the continued impact of the long-running NUREG-0711 initiative on its contents. NUREG-0711, as you know, is the Human Factors Engineering Program Review Model, or PRM. While I have dedicated most of my career to the study and practice of human factors, I am concerned that the PRM GRADUALLY promotes THIS particular interests OF my profession to the detriment of the interests of my industry and the public good. IN A NUTSHELL, THAT THOSE GROWING COSTS ARE OFTEN NOT MATCHED BY COMMENSURATE BENEFITS. Professionally, of course, this puts me in an awkward position. Integrity requires that only the highest possible levels of technical and scientific conduct be promoted … but, at the same time, none higher than possible. Thus, our reach should not exceed our grasp when we claim what is safe, what is valid, what is [redacted], or what is true.

CHAPTER 18 OF The SRP is being invited to incorporate, and thus validate, the essential rhetoric of NUREG-0711. This will bring NUREG-0711 a [redacted] step closer to insinuating itself into the Federal Regulations. Thus far, its principal means TO DO SO has been to lay claim at every opportunity to the words of 10 CFR 50.34(f)(2)(iii), which states that the applicant must:

<QUOTE>

"Provide, for Commission review, a control room design that reflects state-of-the-art human factors principles prior to committing to

fabrication or revision of fabricated control room panels and layouts." (I.D.1)

The reference to I.D.1, indicates the Control Room Design Review section of NUREG-0660, the Post-TMI Action Plan. It seems reasonable that the post-TMI lawmakers understood the current state-of-the-art to be adequate and to supercede past (or absent) standards, and that design products should thus meet the then-current (i.e. adequate) state-of-the-art. On the other hand, it is not at all clear that lawmakers intended human factors to be a moving target for applicants, or that lawmakers would have found a monumental "state-of-the-art process" to be logically equivalent to an adequate design. The law requires a design, not a process, one-step licensing notwithstanding.

The PRM, ostensibly a model for process review and not for the process itself, is nonetheless easily turned to imposing its particular approach. This should be of concern on technical grounds, since *there is little proof of the general cost-effectiveness of this highly bureaucratic approach to design.* Indeed, consider its slight basis:

> "The HFE PRM was developed largely on the basis of applied general systems theory...and the DOD system development process...Other DOD military standards, guidance documents were utilized as well...Since the military has been applying HFE longer that industrial and commercial system developers, the process is more formalized and contains detailed design process requirements. Thus, the DOD system development process was used as a major input..." (NUREG-0711, Rev. 0, Sec.1.4.3, p.1-7)

Though the preceding evidence was struck from Rev.1 of the PRM, the self-report remains accurate. It also summarizes the collective weight of 19 references offered as evidence of the model's validity. But, the authors' finding was merely that DOD's design model was then (circa 1990) *the oldest and the most formal.* Granting that this may be true *forever*, it is at best a weak argument, at worst a red herring, since it is easily overlooked that:

1. applicability of the DOD model to the nuclear industry was uncritically presumed
2. no alternative models were considered
3. no evidence was ever offered that DOD's experience was successful, efficient, or economical

AND, THAT

4. high costs and bureaucratic inefficiency are DOD traditions.

Little has changed to validate the Systems Approach to Design since it was first offered to the industry in 1981 (BY [redacted] App.B to NUREG-0700). Nonetheless, from such modest and obscure bases have come aggressive and widely publicized conclusions: <AND I QUOTE>

> "The HFE PRM describes the HFE program elements that are necessary and sufficient to develop an acceptable detailed design specification and an acceptable implemented design..." (NUREG-0711, and a decade of daughter documents)

Fortunately, whether or not the PRM is technically *necessary and sufficient*, it is not legally *required*. But it is an increasingly obstructive non-requirement, so much so that Human Factors of the control room is now considered the leading risk to successfully bringing a new plant on line within budget and schedule, even more so than software-based protection systems! This is clearly ironic, given the reduced reliance of new designs on operator responses to ensure safety.

Many other strategies can be identified in the PRM for promoting its authority and approach:

- The use of safety, vaguely defined, as a rationale for INSUFFICIENT [redacted] or unproven methods
- The renaming and redefining of existing terms so as to supplant accepted precedents
- A confirmatory research bias that champions largely pre-ordained conclusions AND AVOIDS CONTRADICTORY EVIDENCE
- Promotional self-reporting ~~such as the PRM's "extensive independent peer review"~~
- An inexorable expansion of process scope and complexity, in contradiction to the NRC mandate to reduce unnecessary regulations

And, while they are too lengthy to cover here, I will submit written attachments to justify that a number of the analyses and constructs being promoted by the PRM are merely theories or philosophies (i.e., principles) that have yet to be connected in an objective, reliable, or efficient way with the assurance of nuclear safety. These include the [redacted] process of

Function Allocation, the measurement of Situation Awareness, and the use of "Quasi-Experimental" Validation methods.

Let me conclude by saying that I would welcome the opportunity to discuss any of these matters in more detail should that interest the members of the Subcommittee. I would also encourage you to scrutinize the comments submitted by NEI on this Chapter 18 revision. Thank you very much for your time and attention.

On System Validity, Quasi-Experiments, and Safety: A critique of NUREG/CR-6393

R.B. Fuld[1]
Westinghouse Nuclear Automation, Windsor, Connecticut, USA

Abstract: Based on the methods of Cook and Campbell (1979), a quasi-experimental validation model for complex human-machine systems has been recommended for the final evaluation of nuclear power plant control rooms (O'Hara, et al, 1997). This model-based approach to validating systems and assuring safety is critiqued in terms of experimental theory and engineering practice. It is concluded that the model may be inappropriate for industry as a general model of engineering validation.

Keywords: Validation, quasi-experiment, NUREG/CR-6393, NUREG-0711

1. Introduction

The inaugural issue of *Cognition, Technology, and Work* presented a Quasi-experimental Validation Model (QVM) for complex human-machine systems (O'Hara, 1999). The QVM is described as "a general approach to integrated-system validation in terms of general principles and methods." It was first developed for the US Nuclear Regulatory Commission (NRC) to support licensing reviews of new power plant designs in the human factors engineering (HFE) area (O'Hara, et al, 1997).

These NRC reviews are performed according to the HFE Program Review Model, or PRM (O'Hara, et al, 1994). Design reviews by the PRM no longer focus on the design product per se, but on an all-inclusive design *process*. The PRM's central premise is that by following its top-down design process, substantial assurance is given that the resulting product will be acceptable (i.e. safe). Naturally, the PRM includes steps to test and evaluate the product itself, and the final phase of such product testing is validation. The QVM extends the PRM by taking the requisite validation methodology to follow a model of behavioral experimentation.

The sole mandate for QVM development and for overall licensing review by NRC is nuclear safety. If safety were not somehow in question, the NRC's efforts to regulate design process and validation methods would be unwarranted. This pedigree is significant because US nuclear capacity has not grown in two decades. As a result, the first chance to employ the QVM is likely to occur with the upgrade of existing control rooms, rather than the construction of new ones. But in either case (and notwithstanding precedents claimed from Department of Defense guidance) the cost-effective virtue of these methods to assure safety and reduce risk is largely undemonstrated.

The present article offers a technical critique of the QVM approach to validating systems and assuring safety. Part 2 of this paper defines some key terms for the ensuing discussion. Part 3 relates the concept of engineering validation to the concept of safety as described by the QVM and others. Part 4 considers the role of human performance experiments in engineering validation. Part 5 distinguishes the QVM approach to experimentation from that of true quasi-experiments (Cook & Campbell, 1979). Part 6 then evaluates the QVM's treatment of experimental validity. Part 7 addresses some anticipated objections to the present critique.

2. Terminology

The QVM makes frequent use of various forms of the terms "logic", "experiment", and "validity", and these terms are examined first to clarify the remaining discussion.

Experiment – The term "experiment" is considered to have two relevant interpretations. Informally, an experiment may be synonymous with the more general term "test" or "trial". Formally, an experiment is an hypothesis test under some form of controlled conditions. In this paper, the term experiment will be used only in the formal sense, to distinguish hypothesis tests from other types of tests. An experiment is always a test, but a test need not be an experiment.

---

[1] This article presents the opinions of the author, which are not necessarily those of his employer, the Westinghouse Nuclear Automation segment of the Westinghouse Electric Corporation, Pittsburgh, PA.

Logic – This term, from the Greek *logos* for speech, word, or reason, is oven invoked by the QVM without clarification. Yet, depending on its context, the term "logic" may be ambiguous and loaded with conflicting implications, such as:

1) the science and formal principles of valid inference
2) a demonstration of practical reasonableness, relevance, or justification
3) that which leads to a decision, even apart from or contrary to reason

In a practical sense, engineering validation must distinguish among these three: The necessary, the reasonable, and the rest. Often this question is begged by statements in the QVM and its parent documents which are made in the verbiage of formal logic and logical positivism (e.g. valid inference, necessity, sufficiency, scientific principles, proof, etc.) Some such statements will be examined later in the discussion. But just to be clear, the present paper uses various forms of the term "logic" only in the sense of (2) above. Any reference to (1) will use the term "formal logic"; any reference to (3) will use the term "rationale".

Validity – Validity may generally be defined as the characteristic of *strength* or legitimacy in an open system; validation then is the process of confirming it. In contrast, verification is a confirmation of *truth,* and can only occur in a closed system such as formal logic (Oreskes, et al, 1994). These two qualities and their differences define a range of confirmation practices (Fuld, 1997). Unfortunately, the two are often confused, as when the QVM explains that, "a design can never actually be validated, i.e. it cannot logically be proven that a design will perform acceptably in all actual operating conditions" (O'Hara, 1999; p.39).

Logical Validity – In formal logic, validity addresses the structure, as opposed to the content or meaning, of a given inference. Because formal logic is a closed system, and because such inference structures (whether meaningful or not) can be confirmed unarguably as either correct or incorrect, validity in formal logic is, ironically, a verification problem. It thus conflicts with the general notion of validity described above, and with the particular notions of validity (experimental and engineering) compared below.

Experimental Validity – Experimental validity is the formal and logical strength of an experiment to support its conclusions. Systematic frameworks for experimental validity have been developed, notably among others by Campbell (1957). Campbell seeks improved capability to infer behavioral causation in field studies, so as to expand the bounds of fruitful research methods. As compared to inferences based on laboratory experiments, inferences based on field studies are hampered by the difficulty of meeting prerequisite statistical assumptions, and by a more limited range of possible experimental designs and controls. In light of these constraints, Cook and Campbell (1979) present threats to valid experimental design in order to develop, prioritize, and defend their own departures from such formality.

Engineering Validity – Engineering validity is the practical strength of a system to provide acceptable operating service. In this paper, the term "validation" refers to engineering validation (i.e. of-the-system) unless otherwise specified. The engineering validity of a complex system design is determined, by testing and by day-to-day operations, via the well-known logic of Popperian falsification. A system that fails is proved invalid, while a successful system (pending future failures) is in a weaker sense proved valid by a demonstrated lack of invalidating evidence. The practical problem remains that, to be a pre-condition of operation, engineering validation must be a finite testing process. This requires complex system operations, which are innately open, to be artificially closed by the scope of the tests. Such closure typically consists of demonstrating (i.e. verifying) that the system meets applicable operating requirements for a representative range of test conditions. If testing verifies that these requirements are met, *and no other drastic problems are revealed,* this is typically taken to validate the system. Adequate test strength thus depends essentially on the test conditions selected. However, since the point of final validation is to infer adequate system operation in the future, successful validation is inevitably a weak formal result (Fuld, 1997).

3. Engineering Validation and its Relation to Safety

We return now to the relation between engineering validation and system safety. Safe operation, the NRC's mandate for licensing review, is likewise the point of departure for the QVM. But right away, the QVM extends its charter from safety to operational requirements (O'Hara, 1999; p.37). Of course, to meet operational requirements is crucial to builders and purchasers of the system, and so it may be an issue for engineering validation. However, any causal relation between operational effectiveness and safety is a key inference that cannot be presumed. And lacking such a relation, if the mandate for engineering validation is safety, then operational effectiveness is out-of-scope.

If the point of testing is to assure safety, then valid and efficient testing depends on clear notions of what safety is and how it can best be confirmed. And efficiency here is not a matter of reducing costs, but one of maximizing safety given the reality of limited resources. Yet, behind the QVM is the premise that plant safety is somewhat vague and unverifiable:

> *The rationale underlying the HFE PRM is that "plant safety" is a concept that is not directly observed, but must be inferred from available evidence.* (O'Hara, et al, 1997, p.1-1)

This view of safety is remarkably similar to the notion of a psychological construct, and reveals safety serving as a mere rationale for the current approach. By minimizing the objective physical basis of nuclear safety, for example, the door is thrown wide to prospective investigations, however flawed (since safety is so important and something useful may turn up). But after lifetimes in safety analysis and testing, many experts would be incensed by this vague view of safety. *What is vague in high reliability systems is not so much safety itself as it is the relation of human performance to safety.*

Vague relations in systems are often ascribed to the problem of complexity. The QVM includes a brief tutorial on complexity and its implications, with the relation of parts and wholes suggesting two basic but inadequate system validation approaches. The part- or sub-system validation approach is ruled out by the potential for complex interactions among multiple parts, while the whole- or integrated-system validation approach is dismissed as a limited scope (i.e. design basis) demonstration of likewise inadequate complexity. Thus the stage is set for the requisite "new" methodology of the QVM.

Of course, there can be no objection to the notion that complex events are required to challenge the validity of a system. But the simplicity and limited scope of events given to typify integrated-system validation methods was logically presumed by the QVM. And the intent to use a representative set of complex events for validation testing is neither an innovation of the QVM, nor is it ruled out by other integrated-system approaches.

4. Engineering Validation and Human Performance Experiments

Citing a series of industry standards documents, the QVM claims to be based on "current approaches to design validation". It is fair to say that the QVM is compatible with these standards' recommendations that final designs be tested for overall conformance to requirements. However, readers can confirm that, if any particular approach to such testing is given by the cited documents, experiments are not recommended there.

In fact, for all its talk of scientific bases and logical methodology, *the QVM uncritically presumes that validation tests should follow an experimental model.* This approach likens engineering validation to social science experiments, and permits the recycling of introductory experimental design, hypothesis testing, and Cook and Campbell's (1979) threats to experimental validity. What is true here is not new—but is it applicable? It is hardly a small point that formal human performance experiments are found by many to be, at best, exploratory tools for complex system evaluation.

Experiments are well-suited to scientific research problems because, on the basis of probability theory, under proper conditions of experimental design and control, and given valid measures of hypothetical effects, experiments can detect whether differences measured between treatment groups are greater-than-chance alone.

But engineering validation is not a research problem. Indeed, if the criterion value for a valid measure is known, we do not speak of performing an experiment, but rather of running a test. In engineering validation tests, measured system performance is compared to criteria derived from system requirements. In some cases, these are safety criteria. But valid safety-related criteria rarely exist on human performance,

per se; only on system performance. This limits the role of human performance measures in safety validation to a descriptive or speculative one. Human workload may be relatively high, but when is it clearly unsafe? And if the system is to be highly reliable, should it not be somewhat insensitive, safety-wise, to human performance? Understandably, researchers are frustrated by the insensitivity of system performance to human performance. But that alone just not justify more acceptance criteria.

To circumvent the criterion gap and to save the experimental thesis, one call echoed by the QVM has been for the experimental "benchmarking" of new systems to the de facto criteria of existing systems. But proponents of benchmarking tend to avoid such questions as:

- how existing systems are judged safe in the first place
- how the old and new systems are, and are not, experimentally comparable
- why human performance in new systems must necessarily be better than in old systems
- whether, if benchmarks don't yet exist, twice the testing effort offers twice the safety benefit

Unfortunately, to the other problems that complex systems present to human performance research, it must be added that proper conditions of experimental design and control are practically impossible; hence would-be experimenters must embrace something less than proper experiments.

5 Quasi-experiments and Pseudo-experiments

Momentarily presuming the need for experiments, consider the QVM recommendations on experimental method. These are said to be based on a distinguished body of social research methods summarized by Cook and Campbell (1979) and to which we can scarcely do justice. But briefly: Problems in field study motivated Cook and Campbell to rigorously define and develop methods for *quasi-experiments*. As opposed to laboratory research, field studies typically must be made in available situations with non-equivalent groups (e.g. factory A vs. factory B). This reduces researchers' ability to design a test with a relatively high degree of proper experimental control. A key example is that randomization generally cannot be used to equate test groups (e.g. in the assignment of subjects to test conditions). Lack of randomization is a basic difference between quasi-experiments and proper experiments. A further distinction in Cook and Campbell is that of *non-experiments*, i.e. correlational studies. As most first-year students of behavioral experimentation are taught, correlational methods are a notorious invitation to invalid inference.

In fact, readers can confirm that the methods of the QVM are not quasi-experiments as just described. Furthermore, simulator testing is not a field situation in the sense of Cook and Campbell, so that the very problem which warrants quasi-experimentation in field studies (i.e. non-equivalent subject groups within a test) is not an issue in simulator testing. The QVM analogizes that such non-equivalence exists between sampled conditions (within the test) and population conditions (outside the test). But this is just the standard problem of generalization, which does not warrant quasi-experiments. Arguably, non-equivalent groups may exist within a test when the course of events varies in "identical" operating scenarios. However, QVM methodology offers nothing to actually address this problem.

As the next section shows, what the QVM presents as quasi-experiments are, in essence, merely inadequate ones: Those with insufficient power, noisy and invalid measures, intentional bias, arbitrary criteria, and a rising sea of possible test conditions. The quasi-experiments of Cook and Campbell do not allow for these inadequacies, so that the QVM is without basis in doing so. We therefore add the term *pseudo-experiments* to distinguish the QVM approach from rigorous quasi-experiments.

6 Experimental Validities

By further analogy to Cook and Campbell (1979), the QVM explicates four types of pseudo-experimental validity, along with their corresponding threats. Since the four validities are given as "inference conditions" which must be met for the process to be accepted and for the product thus to be validated, the QVM's showcased threats-to-validity offer a well-equipped toolkit for finding fault *in principal* with the process, thereby invalidating the product by rationale alone. In light of their ostensible necessity, each of these four validities is discussed below in terms of their analogous concepts in Cook and Campbell, and the contradictions embodied by QVM guidance.

System Representation Validity (SRV) – While almost 20 percent of the QVM is spent on SRV, its basic requirements for a representative test can be briefly summarized:

a) test beds, if not the actual system, should provide high fidelity/full-scope simulation of relevant features
b) test subjects should be actual users or actual teams of users with a representative range of ability
c) test conditions should be challenging and should envelope the full range of anticipated events

While these requirements seem reasonable, they are also sufficiently well-known so as to add nothing to modern test methods. At the same time, that SRV remains inherently fallible is perhaps the central problem of engineering validation. Thus it is worth examining further what the QVM suggests.

As for item (a), federal requirements for nuclear plant-specific control room simulators were a post-Three Mile Island result. Use of these simulators has clearly enhanced plant operation, training, and testing. But the validity of such tests is inherently limited: Operators are not under realistic stress in simulators because they know that they are not operating a real plant with genuine risks. However, since complex and dangerous scenarios cannot be run on real plants, and since operators are not deceived by simulators, such limited validity is all that the state-of-the-art can offer.

Item (b) is standard guidance for usability research and testing. However, though actual users with a full range of ability may be desirable subjects, they are not necessarily required for valid acceptance testing. Non-conservative test bias exists only if selected subjects make acceptable test results *more* likely than would typical subjects. A similar bias or advantage can be sought by "over-training" typical subjects. But professional complex system operators are normally over-trained, and if training is sufficient to redress the perceived inadequacies of a complex system, then the system may not be inadequate after all. Indeed, unless "ability" is poorly defined, subjects with *minimal* ability provide the strongest test of system adequacy. Thus, full-ranging subject ability is not a general concern for valid safety testing, and unless the selection or preparation of subjects somehow cheats the test, item (b) is a moot point for final system validation.

It is item (c) which, in the practical sense, is most crucial for a valid system test. But the issue is not whether events should be representative; rather, it is how to make them so. In keeping with its experimental analogy and process bent, the QVM requires a "sampling process" for events. But in high-reliability systems, the rare event combinations of safety concern are not random, so that to sample for them is not statistically valid. Moreover, there is no statistical basis for sizing such a sample, so pseudo-experiments can only call for more: More complexity, more types of tasks, more combined failures. This increases the size, but not the representative validity, of such samples.

Instead, to establish a representative test set, events are not sampled but *selected*, based on ad hoc technical understanding of the system and its credible failure modes. The bases for selected events must justify that the event set is reasonably conservative in its coverage of credible challenges to system safety. Event selection provides the essential basis for generalizing from test results. Given such bases for the selected events, the number of tests and replications is a subordinate issue.

A final note on SRV regards its analog, *external validity*. Though the QVM offers all types of experimental validity as requisites for generalization, this is not necessarily true. It is external validity that most serves generalization (external validity and generalizability were synonyms in Campbell, 1957). Thus, the QVM's main contribution to valid test generalization is SRV, and as discussed before, the elements of SRV are well-known to industry.

Performance Representation Validity (PRV) - This issue, analogous to *construct validity* in Cook and Campbell (1979), pertains mainly to human performance measures and criteria. The main objection to the QVM in this area has already been raised—the vague relations between human performance measures and safety. In psychometric terms, the measures lack *criterion validity* (Anastasi, 1982). In practical terms, a lack of criterion validity would tend to close the case for such measurement.

However, system safety in the QVM is a rationale rather than a basis for measurement. Thus, begging the question of their validity, the QVM merely states that "criteria must be established", and that "four basic approaches are available" (O'Hara, 1999; p.42). In this way, the QVM does little to discourage the use of invalid criteria: *As a matter of process, readers are directed to define criteria without questioning whether valid criteria are even feasible.* Only afterwards does the QVM quietly state that what

it calls "underspecification of performance criteria" is a threat to PRV. That relieves the QVM of responsibility for deficient criteria, placing it instead on the design and its acceptability. Absent relations between human performance and safety are no longer a behavioral science problem—they are now a deficiency of system test methods.

The QVM defines the remaining threats to PRV as deficiencies of "measure comprehensiveness" and of their "acceptable psychometric properties, such as validity and reliability" (O'Hara, 1999; p.43). That many human performance measures are well-known to be psychometrically deficient (if nevertheless useful) suggests of course that better measures are needed. But the QVM again begs the question, allowing acceptable methodology to be "summarized and defined elsewhere", and diverting responsibility for deficient measurement to the design.

And nonetheless lacking valid measures, the QVM urges measurement anyway. The clear case in point is situation awareness (SA), which remains a hot research topic largely *because* its measurement and validity as a psychological construct remain tenuous. Still, the QVM recommends SA measurement. And if the meaning of that suggestion is safely vague, consider the SA tutorial in Section 5.6.2.3.1 of O'Hara, et al (1997) which concludes that, "direct query techniques have been used successfully in numerous applications, including simulated nuclear power plant operations...discussed by Hogg, et al." But to look at the researchers' own report of several studies shows a different picture:

> *SACRI's application lies at a somewhat coarse...level of system evaluation. It can be used alongside other performance measures in generating a more complete evaluation of competing designs. (Hogg, et al, 1995, p.2409)*

It seems there that SACRI's behavior is basically of the two-state variety—showing when SA exists, and when it is lost (e.g. at the occurrence of an event). So, despite a rather laborious data collection method, SACRI's contribution to understanding scenario performance is only comparative and quite modest. Thus, the conclusions of Hogg, et al, while reasonable, do not suggest a criterion measure for safety testing.

The QVM, however, is undeterred by deficient measures and spins this weakness into strength: Valid conclusions must be based on "convergence of multiple measures" (O'Hara, 1999; p.45). However, *convergent validity* is not achieved by combining invalid measures. Rather, convergent validity must be established for as yet invalid measures and constructs through psychometric research (e.g. Campbell & Fiske, 1959). Nevertheless, the role of engineering validation is not psychometric research, and the development of valid measures must logically precede their use in acceptance tests.

Invalid quantification is typical of shoddy research (Cook & Campbell, 1979; p.93). Furthermore, even the addition of valid measures has a premium price:

> *A major practical reason for not [employing multiple measures of an attribute] is that it is so frequently a frustrating experience, lending hesitancy, indecision, and a feeling of failure to studies that would have been interpreted with confidence had but a single response measure been employed. (Campbell, 1957, p.310)*

In the QVM, multiplication of invalid criterion measures multiplies the unjustified obstacles to design acceptance. These obstacles are leveraged through the presumption of experiments: By challenging the design to statistically reject a null hypothesis for multiple measures that lack valid criteria, *what each added measure most ensures is not safety, but another chance for test failure, plus the costs added by the measure's implementation.* This reflects a basic bias of the QVM which will be explored further, below.

Test Design Validity (TDV) – Of the QVM's four validities, TDV receives the least emphasis. This is curious in light of the fact that its analog in Cook and Campbell (1979) is *internal validity*, the major focus of their book. Internal validity is closely related to the concept of experimental control. As such, internal validity and external validity are often conflicting interests. In the case of scientific experimentation, internal test validity clearly has priority (Cambell, 1957). For system validation, in contrast, external test validity (i.e. generalizability) is more important. This reflects a key distinction between the goals of science and engineering.

The QVM treatment of TDV also relates to experimental control, offering recommendations on certain test implementation traits. Good traits include independent multidisciplinary teams, balanced

experimental blocks and presentation orders for scenarios, detailed test procedures, double-blind test administration, and training for both testers and subjects. Bad traits are underspecified test procedures, sources of data noise, sources of bias, confounding, and masking, or things that alter "the relationship between the integrated system and observations of performance."

Though most of these are just broad if miscellaneous guidelines, some detailed objections could be raised, such as the crucial difference in an experiment between random (noise) and systematic (bias) error. (Only systematic error necessarily erodes the validity of a test, and may do so in either the conservative or non-conservative direction.) But the main objections here are only two.

On the one hand, the QVM is defending the inferential integrity of tests for A-B relations that are not evident (i.e. between system safety and human performance measures). On the other hand, pseudo-experiments (unlike true quasi-experiments) are allowed by presumption to be fundamentally invalid test designs in terms of statistical assumptions, power and criteria. In addition, the inherent deficiencies of experimental control in behavioral studies on complex systems are well-known. These deficiencies are largely unavoidable for the complex system representation to be good, and for the generalizability of test results to be high. But on balance, miscellaneous TDV considerations assure neither a valid test nor an adequate test design, and the underlying question of a pseudo-experiment's internal validity is moot.

Statistical Conclusion Validity (SCV) – The last of the QVM's four validities is named the same as its analog in Cook and Campbell (1979). The QVM describes SCV as addressing the relationship between performance data and criteria. The description of criteria as "established" reflects the earlier point that the QVM must presume valid relations between human performance and safety. But the main objection to its treatment of SCV is that, *beyond being technically invalid, the QVM is thoroughly biased against acceptance of the design product.* The recurring rationale for such bias is, ironically, the technical inadequacy of pseudo-experiments. Two basic facets of experiments are used to examine these issues, below.

a) Formulation of the Null Hypothesis: All tests require a decision model, and the presumption of experiments imposes the model of hypothesis testing. By their design, hypothesis tests are conservatively biased to favor the null hypothesis. This necessary conservatism is expressed as an arbitrary alpha level which is grounded in probability theory and research tradition. Unfortunately, pseudo-experiments can have no valid alpha level, because pseudo-experiments have no grounds in probability theory.

   If that deficiency poses no obstacle, one must ask whether, for system validation, the to-be-tested null should be of the weak form, "innocent until proven guilty," or of the strong form, "guilty until proven innocent". Declaring the weak form "incorrect", the QVM follows the strong form: "Maintain a null hypothesis that performance is unacceptable" (O'Hara, 1999; p. 44). The design must then experimentally prove the alternative hypothesis of acceptability.

   Were validation the only means to assure design adequacy, then the strong null might seem reasonable. But in the context of QVM development, validation is the last of the many design process activities that the PRM declares "necessary and sufficient" (O'Hara, et al; 1994). Admittedly, the QVM does not assume that the system in question was developed according to the PRM or any reasonable design process. But complex system designers need reasonable design processes, and these processes should not be presumed to be inadequate. Furthermore, safety-related systems are usually subject to substantial safety, risk, and failure analyses. Finally, a design should always be verified to conform to its requirements. So by insisting in addition on the strong null, *the QVM in effect says that neither conformance to requirements, nor accepted analyses, nor an adequate design process give any assurance of design product adequacy.* Not only does that seem unreasonable, but it also contradicts a central premise of the PRM.

   The strong null is also impractical in light of probability. To statistically reject this strong null in a complex, highly reliable system is infeasible, due to the massive samples of behavior needed to overcome random error and prove the absence of rare events for widely varied conditions. Sample size is one part of the question of an experiment's adequate statistical power. The cost and uncertainty of adequate power are well-known obstacles to valid behavioral experiments in complex systems. And, textbook methods notwithstanding, the QVM has no practical means to judge whether or not the power

of a proposed test is acceptable. But since experiments are already presumed, the QVM can once more shift the burden of deficient methods back onto the design:

> *If a null hypothesis is adopted that performance is acceptable, low power and test insensitivity would work in favor of validating the design. (O'Hara, 1999; p.44)*

The preceding statement, though made in specious argument, is true. But while reviewers must be circumspect of the safety or acceptability of a design product, the case remains that true engineering validation in an open system (i.e. beyond verification of conformance to requirements) is by definition a weak test. So if other evidence of acceptability exists, and other invalidating evidence is still lacking, the weak null that "system performance is acceptable", far from being incorrect, is quite logical for validation. Then, significant findings of unexpected events (i.e. problems) would cause the null, and the design, to be rejected. Such a test does not relieve the system of its need to meet requirements, but neither does it add an unjustified burden to them. Instead, under the weak null, the burden returns to test designers to specify a worthwhile test. And though the question of adequate power remains, this logic is equally applicable to non-statistical, i.e. qualitative evaluations. If the evaluator cannot evidence actual problems with the design, then performance remains acceptable and the design remains valid.

b) Tradeoff of Type I and Type II Errors: The fallibility of statistical decisions is formalized as Type I and Type II error. Under the QVM's strong null hypothesis, an erroneous rejection (Type I error) may seem unsafe; an erroneous retention (Type II error) merely uneconomic. But in fact, there is a direct tradeoff between Type I and Type II errors that is not just explicitly mathematical, but also implicitly practical. Either error has a cost, *and if a successful system is sought*, then the goal in tolerating such errors (since they cannot be avoided, even in theory) is to set the tradeoff so as to maximize the cost-benefits, in this case, to safety.

However, the difficult textbook problem of balancing the cost-benefits between the two is ignored by the QVM. Indeed, irrespective of value, the QVM maximizes their *im*balance: By routinely avoiding Type I errors (i.e. mistaken conclusions of acceptability) it thereby inflates Type II errors (i.e. mistaken conclusions of *un*acceptability). The strong null hypothesis makes acceptance errors inherently unlikely, and as discussed before, such conservatism may itself be unreasonable. But the QVM hedges further against design acceptance, such as by calling for "comprehensive" measures and criteria (each adding to the family-wise probability of test failure), by claiming that "too-narrow performance margins" to acceptance criteria are themselves unacceptable, and by admitting the use of "qualitative comparisons" between performance and criteria where inferential statistics are unjustified. (This latter item seems to have replaced the unabashed recommendation in O'Hara, et al, 1997, to apply criteria to descriptive statistics.) Each of these hedges contradicts the notion of SCV in essential ways.

But in terms of overall evaluation, the probability of false design rejection is not just statistically inflated by the QVM, but is raised indefinitely by its process-orientation and showcased threats-to-validity. Since the mere existence of a threat gives logical grounds for rejecting the process (irrespective of the product), errors of false product rejection are added at reviewer discretion: Any experiment is easily challenged, because threats to validity are finally a tradeoff. And pseudo-experiments, as repeatedly shown, are fundamentally invalid. Thus, using the QVM, the likelihood of false design rejection (typically in the form of interim obstructions to eventual acceptance) is unacceptably high.

The ability of statistics to serve as a valid basis for decisions belongs to inferential statistics alone. Thus, students of statistical science should know its nature from the start:

> *...method[s] for summarizing or describing numerical data...known as descriptive statistics...are contrasted with the modern approach where generalizations are made about the whole which we call the population, by investigating a portion which we call the sample...Such predictions or estimates are generalizations, which we call inferences. The study of how these are made from numerical data is thus called inferential statistics...Inferential statistics is based on probability*

theory [which] measures the chances that an untypical sample will be selected from a population whose characteristics are known. (Lapin, 1973; pp. 3-4)

This introductory passage revisits many of the terms that have been analogically but improperly employed by the QVM. The lack of a basis in probability for pseudo-experiments renders statistical inference from them invalid. In addition, as they substitute test bias for test sensitivity and improperly apply statistical criteria, QVM-based tests are a wholesale addition of decision errors and a gross violation of SRV.

7. Anticipated Objections

Certain generic objections to this critique of the QVM and its supporting documents may be anticipated, and are answered as follows.

The contents of the QVM have been misunderstood – It is common that two plausible interpretations of a single basis lead to opposing interpretations. Interested readers are thus urged to judge these arguments independently and with care. It is hoped that the present critique reflects enough cognizance of its material so that it cannot easily be dismissed as mere misunderstanding. But it is acknowledged that the ambiguity of the QVM's text permits much clarification. Such clarification, if provided, will be beneficial in that misinterpretations of the QVM can be avoided.

The intentions of the QVM have been misunderstood – All claims aside, an author's intentions, even if clear to that author, remain unseen by the public eye. Aiming for verifiable statements in this critique, effort has been made to avoid attributing intentions. But critical readers should regard authors' intentions speculatively and should be suspect of any such claims. In any case, intentions are neither science nor method (nor logical, in any formal sense) and so should have slight impact on the technical merits of either the QVM or the present critique.

The contents of the QVM are being misrepresented, e.g. by being selectively quoted, taken out-of-context, etc. – This could be viewed as an interaction of the preceding objections, an "intentional misunderstanding" as it were. But one aim of this critique has been to place the QVM more clearly *into* the main contexts of its self-justification: Theoretical principles of experimentation, and safe engineering test practices. In these larger contexts, it is argued that the QVM has little basis in the latter other than by analogizing to the former. And as to its basis in the former, readers are encouraged to compare the QVM with the whole of Cook and Campbell (1979), to consider who misrepresents whom and to what extent.

The QVM is somehow being treated "unfairly", e.g. too rigorously – Since the QVM's initial *raison d'etre* was to formalize scientifically informal methods and more firmly impose them on industry, this would seem like the pot calling the kettle black. But, this critique has done no more than consider the QVM logically and practically on its own scientific terms. And if it is unfair to hold the QVM up to itself, then it is likewise unfair to hold anything else to it.

Nothing available is adequate, and nothing better is offered – Acceptance tests are a standard part of the engineering process. These may be more adequate, in some cases, than the QVM suggests. Even so, existing validation methods are subject to incremental improvement, and though this critique did not aim to offer improvements, a few insights may be found in it. But, methods aside, the central burdens of engineering validation remain that test designers must specify worthwhile and representative tests, and that test evaluators must be able to find real problems with the system. As always, both aspects require extensive knowledge of the system and its operation. The rest is detail.

8. Conclusions

Before the QVM, experimental validity and engineering validity were not acquainted, but were related in name only. Using complex system safety as its justification, the QVM takes this superficial similarity to suggest an analogy in which one validity is redefined in terms of the other. This allows human performance experiments to be seen as a necessary condition of final design acceptability.

In opposition to the QVM, this critique argues that validation-of-the-system and validation-by-the-experiment are fundamentally different entities. Furthermore, it shows that the QVM fails to justify, but merely presumes all its key logical points:

- that a causal relation exists between system safety and day-to-day operations
- that a valid relation exists between system safety and human performance criteria
- that experiments are an appropriate form of final system test
- that the QVM is more representative than other forms of final test

The QVM abuses, to the point of self-contradiction, the principles it adopts from scientific literature. In contrast to careful quasi-experiments, the QVM method is shown to be pseudo-experimental, as characterized by the following:

- no established A-B relation between safety and performance
- no basis in probability theory for sampling or generalization
- no basis in experimental effect size for the power or extent of testing
- use of statistically improper measures, criteria, and evaluations of performance

The QVM displaces these deficiencies from itself to the design by laundering them in the didactic wash of threats-to-validity. Such deficiencies are thereby attributed to the inadequate implementation of its methods, rather than to the overall method itself.

The QVM is less scientific or logical than it is scien*tistic* and rhetorical. Assuming that engineering safety is unverifiably vague, and obscured by the fog of valid inference, the QVM builds an ambiguous edifice of behavioral research methods in the name of safety requirements. It offers little to define representative test conditions other than to expand their scope within the desired field of research. Lacking a clear notion of safety and a sensitive test for it, the QVM's principal means to avoid unsafe design acceptance (Type I error) is to bias such testing towards arbitrary design rejection (Type II error). Either bias is, by definition, uneconomic. But if the QVM is not a valid method of safety testing, then to impose it could even make final designs *less* safe, since results will be unreliable, and the required resources could be better spent.

The QVM states its intent "to advance the systems research and engineering literature on validation" (O'Hara, 1999; p.38), but it mostly advances process requirements for human performance experiments in system development. Such experiments are activities of R&D, not of final testing. And notwithstanding the possible benefit of more human performance research, researchers' credibility is diminished by specious arguments and exaggerated claims. Practical engineering, human factors or otherwise, depends on cost-effective methods.

In the larger framework of the PRM, the QVM is more clearly a process audit than a product evaluation method (i.e. was the correct test methodology applied?) This increases the role of independent evaluators even as it distances them from the product and any culpability for its failure (i.e. we only review the process). In the nuclear industry, the QVM arrives as a closing entry in the era of so-called "deterministic" review guidance. The new era is to be one of "risk-informed" review guidance. The risk-informed regulatory initiative aims to ensure safety while reducing unjustified burdens on industry, such as undue conservatism and concern with non-safety considerations (Jackson, 1999). And whether or not final testing can be reduced via risk assessment, the present critique suggests that the QVM is just the sort of burden that improved regulation should minimize.

To conclude: The QVM is less than the sum of its parts. Even if its individual statements are defensible, its overall approach is not. Its presumption that engineering validation should follow an experimental model is unjustified and impractical. Its approach to validation inference is highly biased and is technically improper. Ultimately, its process diverts reviewers from the constructive job of specifying problems with a near-finished design into the technocratic busywork of criticizing methodology on the grounds of general principles. Thus, the QVM should not be accepted by industry as a general model of engineering validation.

7. References

Anastasi, A. Psychological Testing. Macmillan Publishing, New York, 1982.

Campbell, D.T. Factors relevant to the validity of experiments in social settings. Psychological Bulletin, 1957:54:297-312.

Campbell, D.T., Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959:56:2:81-105.

Cook, T.D., Campbell, D.T. Quasi-experimentation: Design and analysis issues for field settings. Houghton Mifflin, Boston, MA, 1979.

Fuld, R.B. Verification and validation: What's the difference? Ergonomics in Design, 1997:5:3:28-33.

Hogg, D.N., Folleso, K. Strand-Volden, F., Torralba, B. Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. Ergonomics, 1995:38:11:2394-2413.

Jackson, S. A. Transitioning to risk-informed regulation: The role of research. Nuclear News, 1999:42:1:29-33.

Lapin, L. L. (1973). Statistics for Modern Business Decisions. Harcourt, Brace Jovanovich, Inc. New York.

O'Hara, J.M. A quasi-experimental model of complex human-machine system validation. Cognition, Technology, and Work, 1997:1:37-46.

O'Hara, J., Higgens, J., Stubler, W., Goodman, C., Eckenrode, R., Bongarra, J., Galleti, G. Human factors engineering program review model (NUREG-0711). US Nuclear Regulatory Commission, Washington, DC, 1994.

O'Hara, J., Stubler, W., Higgens, J., Brown, W. Integrated System Validation: Methodology and review criteria (NUREG/CR-6393). Brookhaven National Laboratory for US Nuclear Regulatory Commission, Washington, DC, 1997.

Oreskes, N., Shrader-Frechette, K., Belitz, K. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science, 1994: 263: 641-646.

United States Nuclear Regulatory Commission

ACRS Subcommittee on Human Factors Engineering
December 2, 2003

Standard Review Plan (SRP) Chapter 18, Human Factors Engineering

Meeting Purpose

- Request ACRS Review, Comment and Endorsement of Revision to SRP Chapter 18, **"Human Factors Engineering"**

- Request ACRS Review, Comment, and Endorsement of Principal Reference Documents :

  NUREG-0711, Rev. 2, **"Human Factors Engineering Program Review Model;"**

  NUREG-0700, Rev.2, **"Human System-Interface Design Review Guidelines;"**

  NUREG-1764, **"Guidance for the Review of Changes to Human Actions."**

Presenters

- James Bongarra, NRR/DIPM/IROB
- Paul Lewis, RES/DSARE/REAHFB
- J. Persensky, RES/DSARE/REAHFB
- Susan Cooper, RES/DRAA/PRAB

Agenda

- SRP, Chapter 18, Rev.2.A
- NUREG - 0711, Rev.2
- NUREG - 0700, Rev.2
- NUREG - 1764
  - Risk-informed screening method
  - Human factors engineering review criteria

SRP Chapter 18 Background

- SRP Chapter 18 is the principal guidance document for human factors engineering
- Summarizes and refers to other detailed guidance documents
- Last revision,1996
- Since 1996, many updates to related documents

SRP Chapter 18 Overview

- SRP Chapter 18 provides a high level framework for all HFE reviews

- Applications:

  Review aspects of –

  - o New Plants
  - o Control room modifications
  - o Modifications affecting human actions

SRP Chapter 18
Overview (2)

- Three Technical Review Areas
  - II.A. Review of New Plants
  - II.B. Review of Control Room Modifications
  - II.C. Review of Changes to Human Actions

SRP Chapter 18 Overview (3)

- Three Technical Review Areas
  - II.A. Review of New Plants

    NUREG-0711
  - II.B. Review of Control Room Modifications

    NUREG-0700 (NUREG-0711)
  - II.C. Review of Changes to Human Actions

    NUREG -1764 (NUREG-0711)

SRP Chapter 18
Review Philosophy

- HFE reviews are performed to provide reasonable assurance of safe plant operation
  - 10 CFR Part 50
  - 10 CFR Part 52
- Review of Design Process and Products
  - HF-related problems often related to faulty early design decisions
  - V&V criteria depend on design process quality
  - Timely feedback to applicants and licensees
  - Perform reviews prior to design completion

SRP Chapter 18
Review Approach

- HFE Program follows a systematic approach
  - Functions => Tasks => Detailed Design
  - Assures consideration of all performance shaping factors and human reliability
- HFE is a Life-Cycle Process
  - Concept Planning
  - Design
  - V&V
  - Performance monitoring after design/modifications
  - Graded, partially risk-informed approach

SRP Chapter 18 Revisions

- Modified review elements and acceptance criteria to agree with NURE-0711, Rev.2
- Added review of plant modifications and credited human actions
- Added a graded approach to HFE review based on risk insights

SRP Chapter 18
Technical Basis for Revision

- Address feedback from applications
  - ALWR reviews
  - Plant modernization reviews performed in other countries
  - Feedback from staff and international users
- Incorporate NRC research on human factors engineering

SRP Chapter 18 Summary

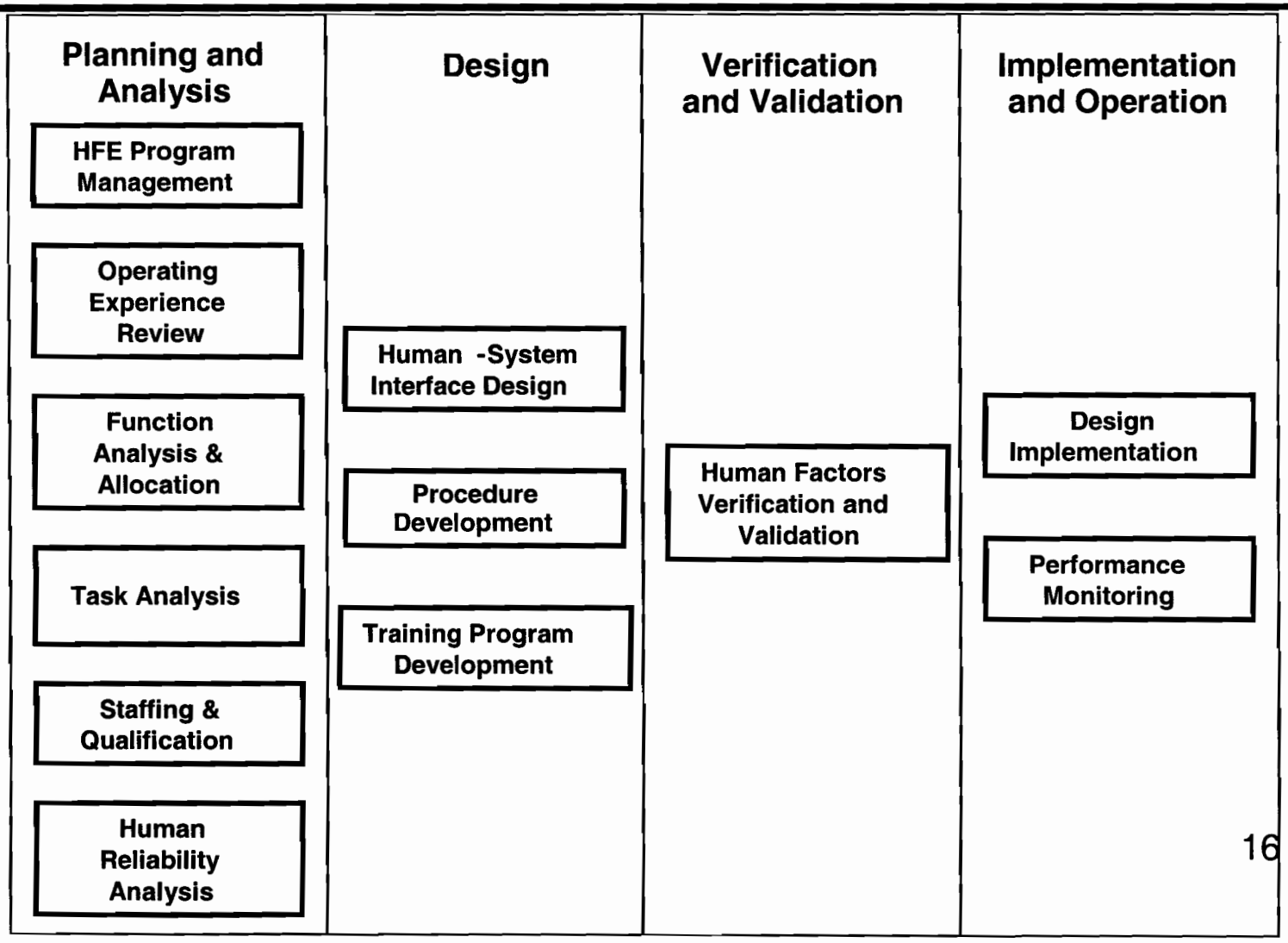
- SRP Chapter 18:
  - Existence Since Early 1980's
  - Last Revised – 1996
  - Principal NRC HFE Guidance
  - Refers to Several HFE Related Guidance Documents
  - Latest Revision Upgraded, Partially Risk-informed

NUREG-0711, "Human Factors Engineering Program Review Model"

NUREG-0711
What is it?

- Complete set of basic HF review elements.
  - All HF reviews.
  - Complete life cycle.
  - Includes reviews of the design process and the design product.
- Elements from NUREG-0711 are adapted in other documents for specific types of review.

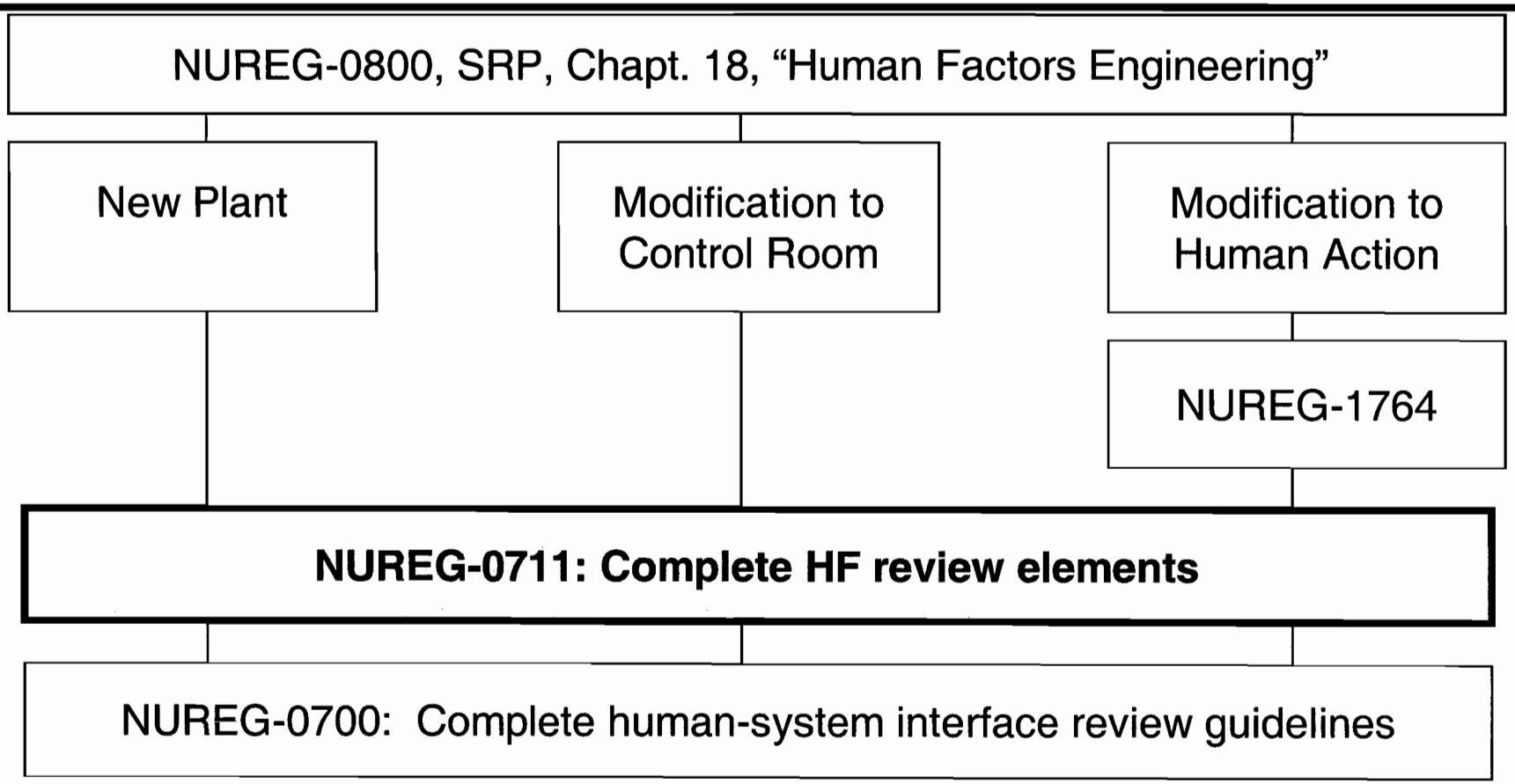
NUREG-0711
Review Elements



NUREG-0711
Format of Elements

- Background
  - Brief explanation of the rationale and purpose.
- Objective
  - The review objective(s) of the element are defined.
- Licensee submittals
  - Reports, design files review, and observations used by staff.
- Review criteria
  - Acceptance criteria for design process and products and for the final design review.
- Reference documents

NUREG-0711
Relationship to Other NUREGs

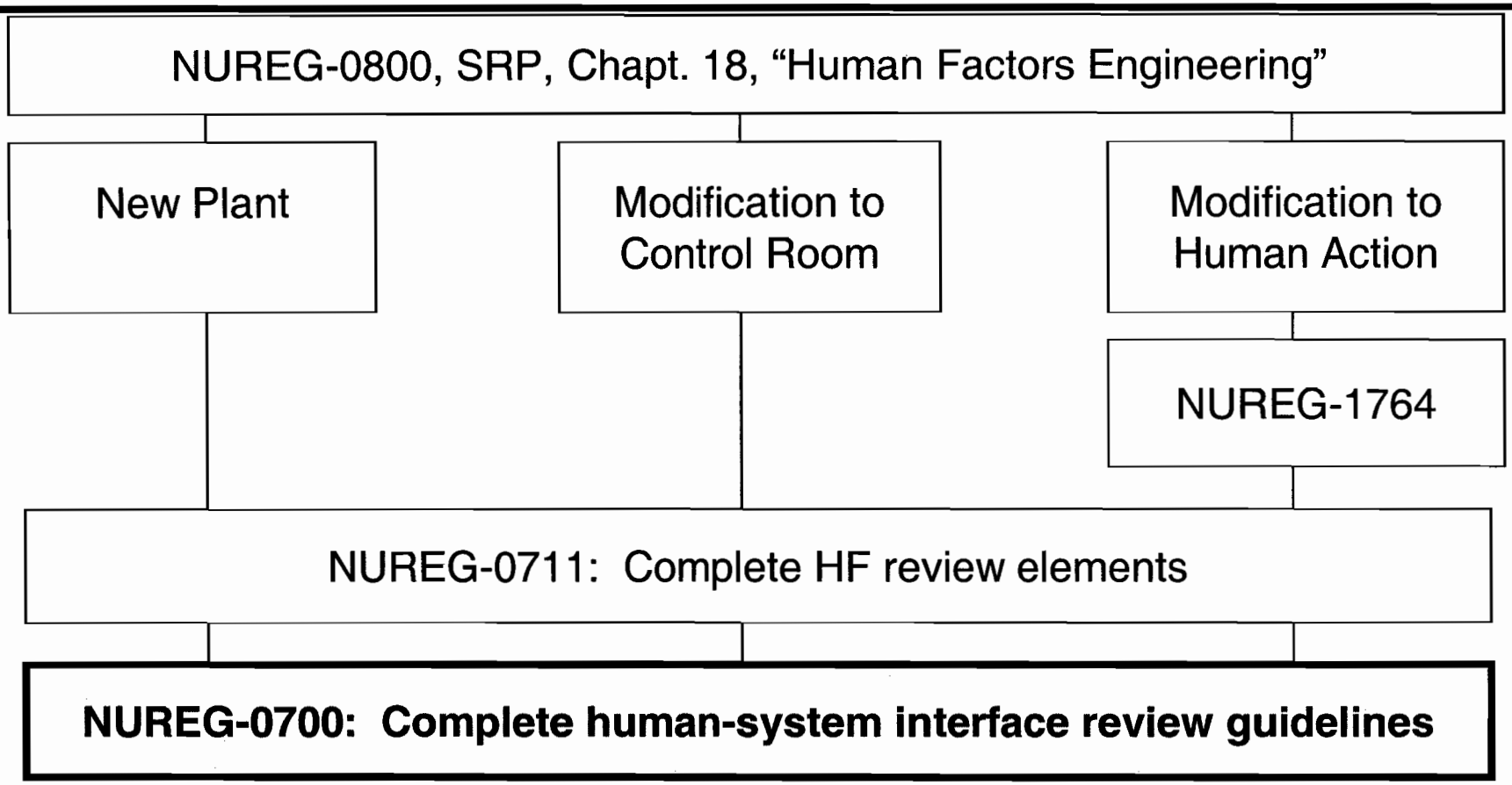


NUREG-0711
Changes From Prior Version

- Applies to all HF reviews, not just advanced reactors.
- Complete set of basic HF review elements by adding two elements:
  - Design Implementation.
  - Performance Monitoring.
- Changes made to the following elements
  - Function Analysis and Allocation.
  - Human Reliability Analysis.
  - Human-System Interface.
  - Verification and Validation.
- Most of this review guidance previously existed.

NUREG-0700, "Human-System Interface Design Review Guidelines"

NUREG-0700
Relationship to Other NUREGs



NUREG-0700
What is it?

- Complete set of guidelines for reviews of human-system interfaces
- See next slide for review topics

NUREG-0700
Review Topics

- Basis HSI Elements
  - Information Display
  - Interaction and Interface Management
  - Basic Controls
- HSI Systems
  - Alarm Systems
  - Group-View Display System
  - Soft-Control System
  - Computer-Based Procedure Systems
  - Computerized Operator Support Systems
  - Communication Systems
- Workstations and Workplaces
- HSI Support
  - Maintainability of Digital Systems

NUREG-0700
Changes From Prior Version

- Fills gaps in review guidance for digital systems
  - General computer-based information system interfaces
  - Soft controls
  - Computer-based procedures and alarm systems
  - Interface management and navigation

NUREG-0711 and -0700 Significance (1)

- Culmination of much work
- State-of-the-art technical basis
- Up-to-date review guidance for digital HSIs
- Wealth of information and guidance
- These NUREGs establish a new state-of-the-art in comprehensive review guidance for NPPs.

NUREG-0711 and -0700 Significance (2)

- Outside users of NPP regulation and design review
  - KINS: development of Korean next generation reactor regulatory criteria (KOREA)
  - SKI: basis for review of control room upgrades of several plants (Sweden)
  - CSN: basis for review of control room modifications (Spain)
  - IAEA: Temelin NPP design review (Czech Republic)
  - GE: Lungman ABWR NPP control room design review (Taiwan)
  - BNFL: design review of expert system (UK)

NUREG-0711 and -0700 Significance (3)

- Outside users of NPP design
  - EPRI: basis for design of control room modifications for digital I&C upgrades
  - AECL: plant modifications and validation
  - KOPEC: design of new plant and control room upgrades (Korea)
  - KEPRI: design of HSI prototypes (Korea)
  - KAERI: design of advanced control room simulator (Korea)
  - ABB: design of control room upgrades of several plants (Sweden)
  - Halden Reactor Project: design of control room upgrades (Norway)
  - Westinghouse: design of upgrades to NPP (Switzerland)
  - TVA: HF guidance development

NUREG-0711 and -0700 Significance (4)

- Non-NPP Outside users
  - Savannah River Lab: control room design and evaluation
  - PNNL: Hanford HSI design
  - DoD: Army Research Laboratory design reviews
  - DoD: one basis for situation assessment guidance development
  - Navy: submarine and aircraft carrier
- Incorporation into Standards Activities
  - International Standards Organization (ISO)
  - Institute for Electrical and Electronic Engineers (IEEE)
  - International Electrotechnical Commission (IEC)

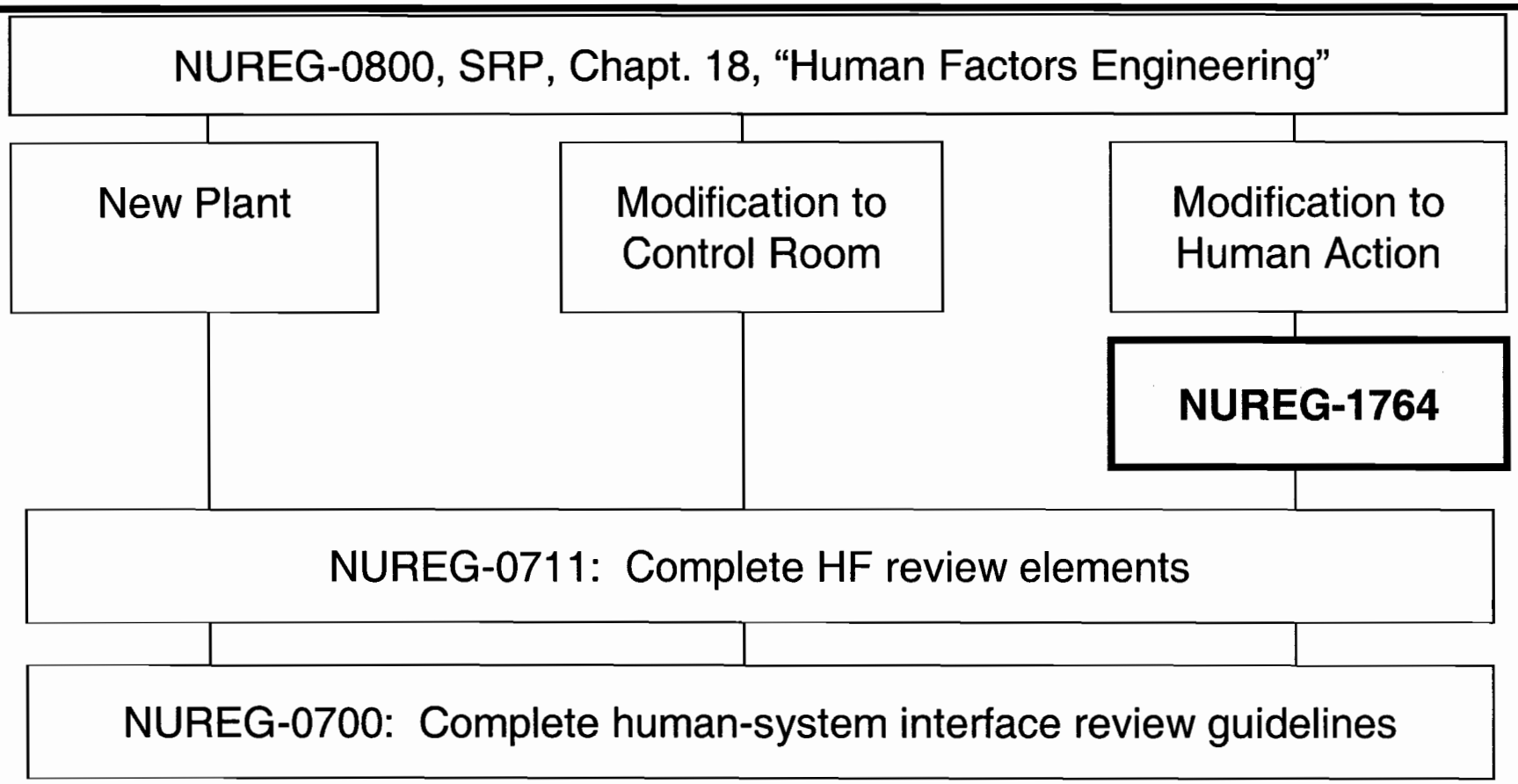
NUREG-1764, "Guidance for the Review of Changes to Human Actions"

NUREG-1764
What is it?

- Guidance for the review of changes to human actions.
  - New actions (e.g., substitution of a human action for an automated action, when the automated equipment fails.)
  - Modified actions (e.g., due to new or modified system components.)
  - Modified task demands (e.g., change in amount of time available, or in environment.)

- Risk-informed review guidance
  - The risk screening method determines the level (detailed, medium, brief) of human factors review.

NUREG-1764
Relationship to Other NUREGs



NUREG-1764
Three Phases

1. Risk-screening method
2. Human factors review
   - Level 1 review (most detailed and thorough)
   - Level 2 review (medium)
   - Level 3 review (brief)
3. Results of human factors review is submitted to Integrated Decision-Making (See RG1.174, Section 2.2.6) and to Safety Evaluation Report

NUREG-1764
Phase 1
Risk Screening Method

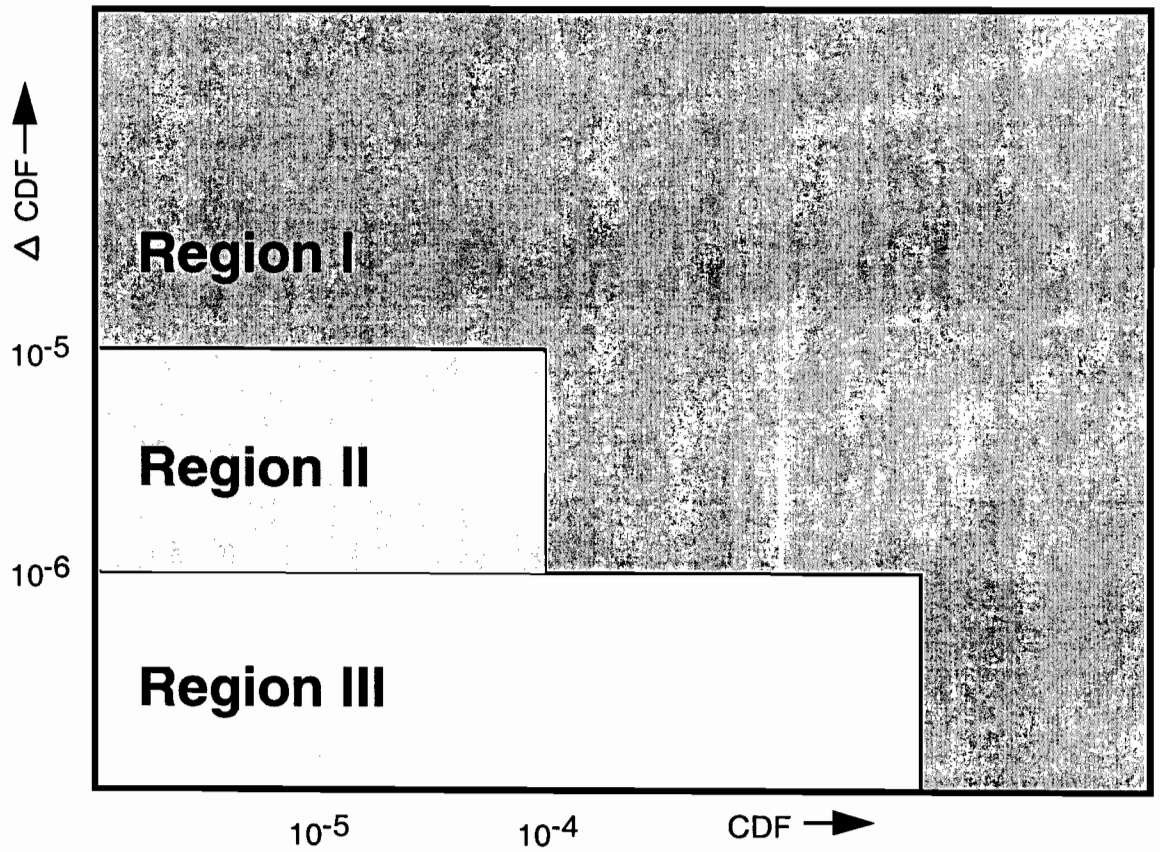
NUREG-1764, Phase 1
Risk Screening Method Has Four Steps

- Step 1: Change in risk due to modification per RG 1.174.
- Step 2: Evaluation of risk-significance of human action not being performed correctly.
- Step 3: Qualitative evaluation.
- Step 4: Integrated assessment.

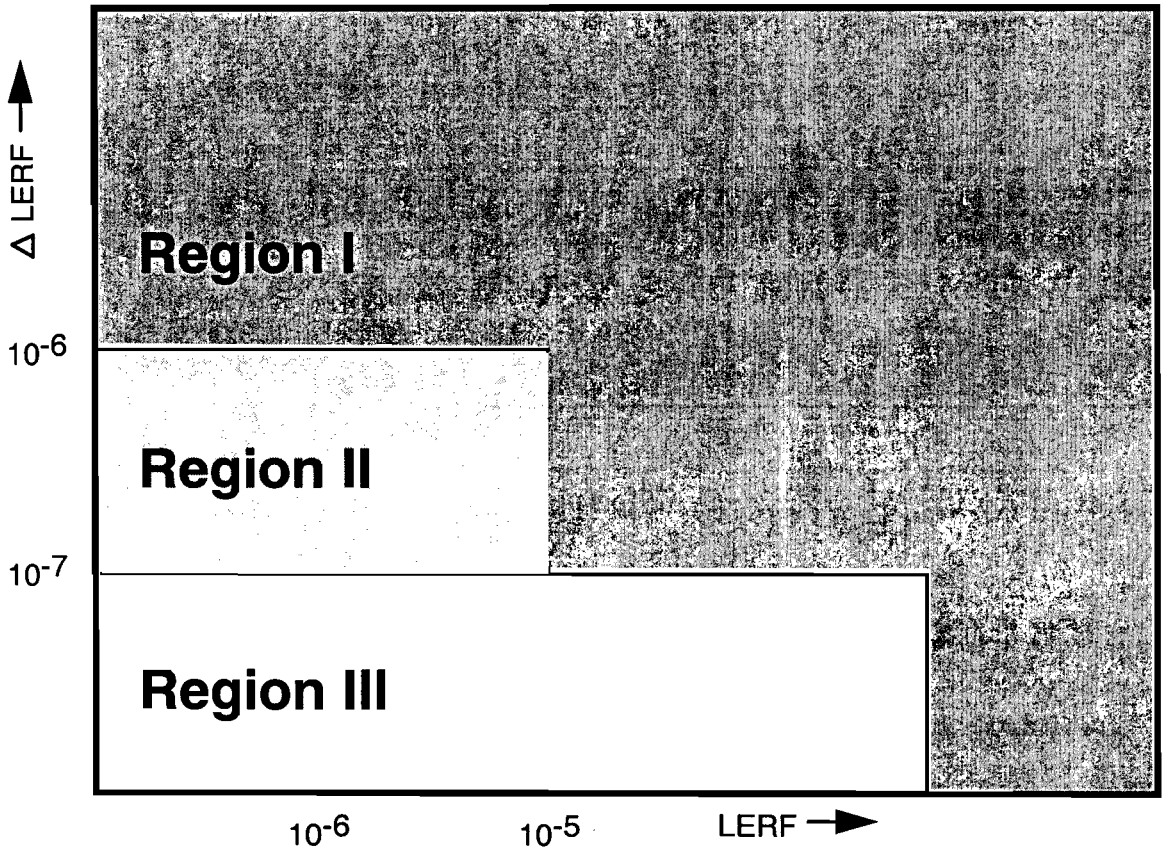
NUREG-1764, Risk Screening Step 1

- Uses a four step process
- Step 1 - change in risk due to modification per RG 1.174
- Δ CDF_{mod} = [new CDF (with modification in-place) - current baseline CDF]

NUREG-1764, Risk Screening, Step 1 CDF Screening Guide



NUREG-1764, Risk Screening, Step 1
LERF Screening Guide



NUREG-1764, Risk Screening, Step 1 (cont.)

- If HA only and Region I – Do a Level I HFE review.
- Otherwise - go to Step 2 to evaluate risk-significance of human-action not being performed correctly

NUREG-1764, Risk Screening
Step 2

- Step 2 - evaluation of risk-significance of human action not being performed correctly

- Evaluates risk importance of HA based on both RAW and FV importance measures.
- Preliminary determination of Review level for HA as Level I, II, or III.

NUREG-1764, Risk Screening
Step 3

- Step 3 - Qualitative Evaluation
  - Allows the screener to reduce or elevate the Level of HFE Review
  - Based on factors such as:
    - Personnel functions and tasks.
    - Design support for task performance.
    - Performance shaping factors.

NUREG-1764, Risk Screening
Step 4

- Step 4 Integrated Assessment
  - Integrates the results generated in Steps 1 through 3
  - Provides a Table that gives the Level of HFE review based on screening.
  - Conclusion of Risk Screening: The level (I, II, or III) of human factors review.

NUREG-1764
Phase 2
Human Factors Review

NUREG-1764, HF Review
Three Levels of HF Review

- Level I is the most detailed review
  - Review areas are adapted from NUREG-0711
- Level II is a moderately detailed review
  - Review areas are adapted from NUREG-0711
- Level III is a brief review

NUREG-1764, HF Review Previous Guidance

- Most of the HF Review guidance previously existed.
  - Information Notice 97-78, "Crediting Operator Actions in Place of Automatic Actions and Modifications of Operator Actions, Including Response Times."
  - Information Notice 91-18, "Information to Licensees Regarding Two NRC Inspection Manual Sections on Resolution of Degraded and Nonconforming Conditions and on Operability."
  - NUREG-0711

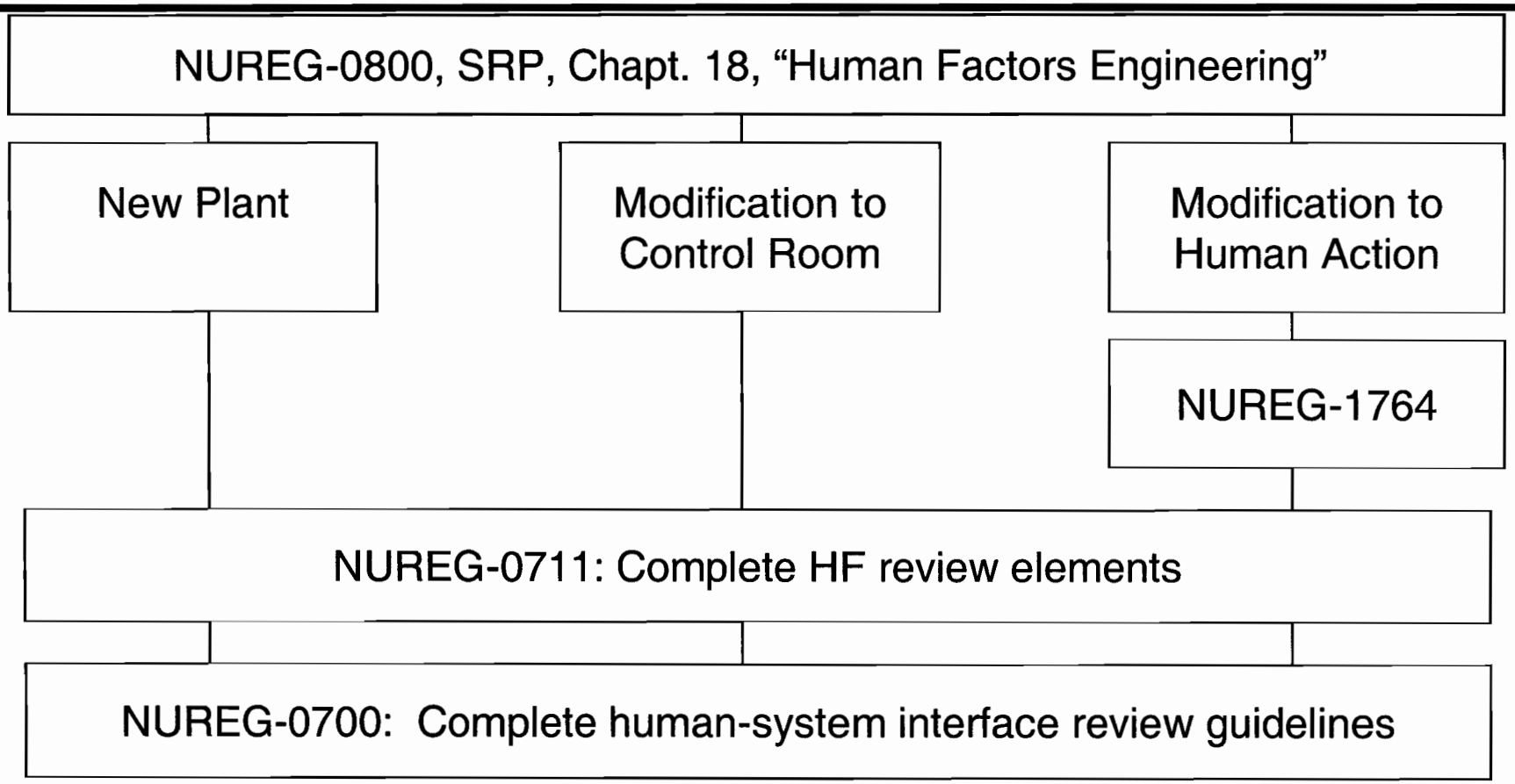
NUREG-1764, Phase 3 HF Review Decision

- Result of human factors review is submitted to Integrated Decision-Making (See RG1.174, Section 2.2.6) and to Safety Evaluation Report

SRP Chapter 18 and Related NUREGs Summary

- SRP Chapter 18 now has three applications
  - Most of the review guidance previously existed
- NUREG-0711
  - Application scope expanded from advanced reactors to all reactors
  - New review guidance for existing review elements
  - Two new elements
- NUREG-0700 added review guidance for specific HSI topics
  - Computer-based procedures
  - Soft controls
- NUREG-1764
  - This is a new document.
  - It contains (1) a risk screening method and (2) graded HF review criteria.
- Most of the added review guidance previously existed.

SRP, Chapter 18, "Human Factors Engineering" Summary



SRP Chapter 18 and Related NUREGs Significance: Promote NRC Performance Goals

- Reduce unnecessary burden
  - NUREG-1764 has a risk screening method
  - They are guidelines, not requirements
- Improve regulatory efficiency
  - Clear, detailed review guidance
  - Detailed, because users want detail
  - Organized package
    - Role of each NUREG is well defined
    - Standardized formats
- Maintain safety
  - Risk screen provides detailed review for risk important human actions.
  - Reduces regulatory uncertainty, which can cause licensees to delay safety improvements
  - Contains review guidance for new (digital) technologies
- Further implement NRC policy on risk-informed regulation

Technical Reports Related to the Development of HFE Review Guidance

Brown, W. (2001). *Update of NUREG-0700 control room and work place environment review guidance* (BNL Technical Report E6835-T5-1-6/01). Upton, New York: Brookhaven National Laboratory.

Brown, W., O'Hara, J., and Higgins, J. (2000). *Advance alarm systems: Guidance development and technical basis* (NUREG/CR-6684). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Echeverria, D., Barnes, V., Bitner, A., Durbin, N., Fawcett-Long, J., Moore, C., Slavich, A., Terril, B., Westra, C., Wieringa, D., Wilson, R., Draper, D., Morisseau, D., and Persensky, J. (1994a). *The impact of environmental condition on human performance: a critical review of the literature* (NUREG/CR-5680, Vol.1). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Echeverria, D., Barnes, V., Bitner, A., Durbin, N., Fawcett-Long, J., Moore, C., Slavich, A., Terril, B., Westra, C., Wieringa, D., Wilson, R., Draper, D., Morisseau, D., and J. Persensky, J. (1994B). *The impact of environmental condition on human performance* (NUREG/CR-5680, Vol.2). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Higgins, J. and Nasta, K. (1996). *HFE Insights For Advanced Reactors Based Upon Operating Experience* (NUREG/CR-6400). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Higgins, J., O'Hara, J., Stubler, W., & Deem, R. (1999). *Summary of credit of past operator action cases* (Report No. W6022-T1-1-7/99). Upton, New York: Brookhaven National Laboratory.

Higgins, J. and O'Hara, J. (2000). *Proposed Approach for Reviewing Changes to Risk-Important Human Actions* (NUREG/CR-6689). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J. and Brown, W. (2002). *The Development and Revision of the NRC's HFE Design Review Guidance: NUREG-0711 and NUREG-0700.* (W6546-1-9/02). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J. (1994). *Advanced human system interface design review guideline* (NUREG/CR-5908). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J. and Brown, W. (2002). *The effects of interface management tasks on crew performance and safety in complex, computer-based systems.* (NUREG/CR-6690). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J. and Brown, W. (2001). *Human-system interface management: Human factors review guidance* (BNL Technical Report W6546-T6A-1-3/01). Upton, New York: Brookhaven National Laboratory.

O'Hara, J., Brown, W., Hallbert, B., Skråning, G., Wachtel. J., and Persensky, J. (2000). *The*

Effects of Alarm Display, Processing, and Availability on Crew Performance (NUREG/CR-6691). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J., Brown, W., Higgins, J., and Stubler, W. (1994). *Human factors engineering guidelines for the review of advanced alarm systems* (NUREG/CR-6105). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J., Brown, W., and Nasta, K. (1996). *Development of NUREG, 0700, Revision 1* (BNL Technical Report L-1317-2-12/96). Upton, New York: Brookhaven National Laboratory.

O'Hara, J., Higgins, J., and Kramer, J. (2000). *Advanced Information Systems: Technical Basis and Human Factors Review Guidance* (NUREG/CR-6633). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J., Higgins, J., Stubler, W., and Kramer, J. (2000). *Computer-Based Procedure Systems: Technical Basis and Human Factors Review Guidance* (NUREG/CR-6634). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J., Stubler, W., Brown, W., and Higgins, J. (1997). *Integrated System Validation: Methodology and Review Criteria* (NUREG/CR-6393). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J., Stubler, W., and Higgins, J. (1998). *The development of HFE design review guidance for hybrid human-system interfaces* (BNL Report J6012-T6-12/98). Upton, New York: Brookhaven National Laboratory.

O'Hara, J., Stubler, W., and Higgins, J. (1996). *Hybrid human-system interfaces: Human factors considerations* (BNL Report J6012-T1-4/96). Upton, New York: Brookhaven National Laboratory.

Roth, E. and O'Hara, J. (2002). *Integrating digital and conventional human system interface technology: Lessons learned from a control room modernization program.* (NUREG/CR-6749). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Stubler, W., Higgins, J., and Kramer, J. (2000). *Maintenance of Digital Systems: Technical Basis and Human Factors Review Guidance* (NUREG/CR-6636). Washington, D.C: U.S. Nuclear Regulatory Commission.

Stubler, W. and O'Hara, J. (1996a). *Group-view display support document* (BNL Report E2090-T4-4-4/95, Rev. 1). Upton, New York: Brookhaven National Laboratory.

Stubler, W. and O'Hara, J. (1996b). *Human-System Interface Design Process and Review Criteria* (BNL Report E2090-T4-5-11/95). Upton, New York: Brookhaven National Laboratory.

Stubler, W., O'Hara, J., Higgins, J., and Kramer, J. (2000). *Human-System Interface and Plant Modernization Process: Technical Basis and Human Factors Review Guidance* (NUREG/CR-6637). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Stubler, W., O'Hara, J., and Kramer, J. (2000). *Soft Controls: Technical Basis and Human Factors Review Guidance* (NUREG/CR-6635). Washington, D.C.: U.S. Nuclear Regulatory Commission.