INL/EXT-13-30570

Framework for Human-Automation Collaboration: Conclusions From Four Studies

Johanna Oxstrand
Katya L. Le Blanc
Jeffrey C. Joe
April M. Whaley
Heather Medema
John O'Hara

November 2013

The INL is a U.S. Department of Energy National Laboratory
operated by Battelle Energy Alliance

SMR/ICHMI/INL/TR-2013/05
INL/EXT-13-30570
Revision 0

Framework for Human-Automation Collaboration: Conclusions From Four Studies

Johanna Oxstrand[c]
Katya L. Le Blanc[a]
Jeffrey C. Joe[a]
April M. Whaley[a]
Heather Medema[a]

John O'Hara[d]

November 2013

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

[c] Idaho National Laboratory
[d] Brookhaven National Laboratory

ABSTRACT

The Human Automation Collaboration (HAC) research project is investigating how advanced technologies that are planned for Advanced Small Modular Reactors (AdvSMR) will affect the performance and the reliability of the plant from a human factors and human performance perspective. The HAC research effort investigates the consequences of allocating functions between the operators and automated systems. More specifically, the research team is addressing how to best design the collaboration between the operators and the automated systems in a manner that has the greatest positive impact on overall plant performance and reliability.

The ultimate goal of the AdvSMR HAC research effort is to develop design guidance that supports optimal interaction between humans and automated systems as well as an interactive tool for design and evaluation of HAC. The tool is intended to be used by engineers in their process of designing AdvSMR systems.

The HAC research effort is a part of the Department of Energy (DOE) sponsored AdvSMR program conducted at Idaho National Laboratory (INL). All AdvSMR designs will employ advanced digital instrumentation, controls, and human-machine interfaces (ICHMI), technology that is significantly more advanced that existing analog systems in the light water reactor fleet, and there are a number of concerns about how those technologies will affect human performance and the overall safety of the plant. It is expected that AdvSMRs will rely on automation to a greater extent than the current nuclear power plant fleet.

Oxstrand et al. (2013 - March) describes the efforts conducted by the researchers to identify the research needs for HAC. The research team reviewed the literature on HAC, developed a model of HAC, and identified gaps in the existing knowledge of human-automation collaboration. As described in Oxstrand et al. (2013 – June), the team then prioritized the research topics identified based on the specific needs in the context of AdvSMR. The prioritization was based on two sources of input: 1) The preliminary functions and tasks, and 2) The model of HAC. As a result, three analytical studies were planned and conduced; 1) Models of Teamwork, 2) Standardized HAC Performance Measurement Battery, and 3) Initiators and Triggering Conditions for Adaptive Automation. Additionally, one field study was also conducted at Idaho Falls Power.

Before the research team develops guidance on how to effectively design HAC for AdvSMR, it is necessary to understand operating experience regarding HAC in highly automated systems. Because there is little to no operating experience with highly automated systems for day-to-day operations in the nuclear field (let alone the AdvSMR field), researchers have sought input from other process control facilities that use high degrees of automation. The first field study was carried out at Idaho Falls Power (IF Power), which is an electric utility that provides power to 22,400 residential and 3695 commercial customers. According to NUREG/CR-1368 the human operator's primary functions would be to monitor and verify performance of safety systems, maintain communication with appropriate onsite and offsite personnel, and initiate recovery actions following an event (NRC, 1987). This operator role is similar to the operators at IF Power that participated in the field study.

During the field study the research team carried out a review of the plant human-system interface (HSI) where they inspected the HSI elements and asked the operator for additional information where appropriate. The researchers were also provided with a list of interview questions to ask the operators. These questions focus on what tasks the operator carries out on a typical day, how the operator interacts with the automation and what challenges he/she perceives associated with the HAC design (e.g., it is too automated, or the HSI doesn't provide enough information). In addition, the operators were asked to answer a plant questionnaire providing the research team information regarding the plant mission, the approach to function allocation, and information about operators (e.g., number and qualifications) and operator training.

The IF Power operators stated that they managed to stay engaged in the process due to their well-defined and specific roles. The automatic system is not equipped to do the tasks that they do manually, therefore they serve an essential function other than their supervisory control roles. The IF Power operators reported that they fully trusted the automation to correctly execute the tasks assigned to the automation and that they could not recall a time when an operator overrode the automatic system.

The teamwork analytical study focused on identifying the requirements for effective human-automation teamwork that are appropriate in a commercial nuclear power team environment. Once identified, the researchers can elaborate on the specific requirements of human and automation team members, and can use those requirements to develop guidance for what automation characteristics and attributes are needed to be good "team players", which is essential to ensure that guidance on the design of successful HAC in AdvSMRs is developed.

The researchers intended to identify what makes for effective teams, in particular effective teams in the nuclear domain. To do so, the researchers evaluated different models of human-human teamwork. As a part of the evaluation the researchers conducted:

- Review recent research on models of human teamwork.
- Emphasize nuclear power plants and other complex, highly automated systems.
- Identify general principles about what makes for effective teams and the characteristics needed.

The HAC research effort is focused on producing guidance on how to design the automation and human-automation collaboration in new AdvSMR plants to facilitate optimal operator and plant performance in a highly automated system. To accomplish this goal, it is necessary to be able to measure performance of the system, operator, plant, and the human-automation interaction in this specific context. As discussed in Oxstrand et al (2013a), there is a need for a more complete set of performance measures that are applicable in this domain and focus on the relationship between humans and automation, including operator awareness, trust in the automation, and use of automation, as well as measures that reflect multi-agent teamwork.

The analytical study of teamwork suggests that the demands on human-automation interaction might be the greatest when the automation agent is supporting higher-level cognitive process, such as situation assessment and response planning. The result of the teamwork study suggests the following design requirements for designing human-automation interaction:

- The way information is processed is accessible to the operator
- The design of communication functions and features enable operators to obtain the information they need to use automation information, such as assessing the credibility of the results
- Interruptions are minimized
- The level of detail can be controlled
- Operators can provide input to the processing and direct its activities

In order to facilitate optimal operator and plant performance in a highly automated system it is necessary to be able to measure performance of the system, operator, plant, and the human-automation interaction in this specific context. The performance measures analytical study aimed to develop an initial set of performance measures that are applicable in this domain and focus on the relationship between humans and automation, including operator awareness, trust in the automation, and use of automation, as well as measures that reflect multi-agent teamwork.

The main conclusions from the performance measures study are that objective performance measures are the most direct measures of human and system performance, however, they should be used in conjunction with subjective measures to validate the objective measures and to better detect ceiling effects. The subjective measures can also be used to gain qualitative insights useful for future system design. The researchers recommend a set of metrics to be used to measure primary task performance, situational awareness, workload, trust, and system performance.

A review of the adaptive automation literature indicates that there are tradeoffs associated with each triggering mechanism, but that some are more applicable to the AdvSMR domain. The two mechanisms that consistently improve performance in laboratory studies are operator-initiated adaptive automation based on hierarchical task delegation and the Electroencephalogram (EEG) –based measure of engagement. Current EEG methods are intrusive and require extensive analysis; therefore it is not recommended for an AdvSMR control rooms at this time. Researchers recommend that future research focus on comparing performance across multiple triggering conditions and levels of static automation, and investigating the effect of adaptive automation on an operator's ability to resume manual control in the event of an automation failure

Moving forward, the research team will utilize the insights learned from the field study at IF Power and the three analytical studies and design a series of experimental studies to further investigate how to best design the collaboration between the human operator and the automated system in the context of AdvSMRs.

CONTENTS

FIGURES

TABLES

ACRONYMS

AdvSMR	advanced small modular reactor
AMCR	advanced main control room
CIC	combat information centers
CMCR	conventional main control room
COSS	computerized operator support system
CRM	crew resource management
DSS	decision support system
ECG	Electrocardiogram
EEG	electroencephalogram
EOP	emergency operating procedure
HAC	human-automation collaboration
HCT	human-computer trust
HHS	human-human-system
HSI	human-system interface
I&C	instrument and control
ICHMI	instrumentation, controls, and human-machine interfaces
IF	Idaho Falls
INL	Idaho national laboratory
INPO	institute of nuclear power operations
ISA	instantaneous self assessment
LOA	level of automation
LWR	light water reactor
NASA	national aeronautics and space administration
MBM	mutual belief model
MCH	modified Cooper-Harper scale
O&M	operations and maintenance
SA	situational awareness
SACRI	Situation Awareness Control room Inventory
SAGAT	situation awareness global assessment technique
SART	situational awareness rating technique
SPAM	situation present assessment method
SMR	small modular reactors
SWAT	subjective workload assessment technique

TLX	task load index
UAMPS	Utah associated municipal power systems
UAV	unmanned aerial vehicles

1. INTRODUCTION

The Human Automation Collaboration (HAC) research project is investigating how advanced technologies that are planned for Advanced Small Modular Reactors (AdvSMR) will affect the performance and the reliability of the plant from a human factors and human performance perspective. The HAC research effort investigates the consequences of allocating functions between the operators and automated systems. More specifically, the research team is addressing how to best design the collaboration between the operators and the automated systems in a manner that has the greatest positive impact on overall plant performance and reliability.

The ultimate goal of the AdvSMR HAC research effort is to develop design guidance that supports optimal interaction between humans and automated systems as well as an interactive tool for design and evaluation of HAC. The tool is intended to be used by engineers in their process of designing AdvSMR systems.

The HAC research effort is a part of the Department of Energy (DOE) sponsored AdvSMR program conducted at Idaho National Laboratory (INL). All AdvSMR designs will employ advanced digital instrumentation, controls, and human-machine interfaces (ICHMI), technology that is significantly more advanced that existing analog systems in the light water reactor fleet, and there are a number of concerns about how those technologies will affect human performance and the overall safety of the plant. It is expected that AdvSMRs will rely on automation to a greater extent than the current nuclear power plant fleet.

AdvSMRs have been proposed as a way to meet the nation's energy needs while also reducing carbon emissions (Carelli et al., 2010). In order to be economically viable, AdvSMRs will likely need to achieve Operations and Management (O&M) costs that are comparable to the existing fleet of nuclear power reactors on a per/kilowatt basis. If existing concepts of operation for nuclear power plants are applied AdvSMR operations, some estimates indicate that the O&M costs for an SMR would be greater than that of an existing light water reactor (Carelli et al., 2010). Given that most AdvSMR designs will produce 300MWe or less (IAEA, 2005, 2006), they will need to drastically reduce the per-unit O&M costs compared to existing designs. Some features that many conceptual AdvSMR designs have proposed to meet this need are reduced complexity, passive safety features, online maintenance, etc. Another proposed feature of AdvSMR designs is an increase in automation to enable reduced staffing and increased efficiency (O'Hara, Higgins, & Pena, 2011). Though increases in automation can enhance efficiency, many researchers have noted human performance issues associated with increased automation including complacency and loss of situation awareness (Bainbridge, 1983; Vicente, 2003; Miller & Parasuraman, 2007). In order to ensure that proposed increases in automation for AdvSMRs do not have adverse effects, designers need to fully understand the impacts of automation on human performance.

Oxstrand et al. (2013 - March) describes the efforts conducted by the researchers to identify the research needs for HAC. The research team reviewed the literature on HAC, developed a model of HAC, and identified gaps in the existing knowledge of human-automation collaboration. As described in Oxstrand et al. (2013 – June), the team then prioritized the research topics identified based on the specific needs in the context of AdvSMR. The prioritization was based on two sources of input: 1) The preliminary functions and tasks, and 2) The model of HAC. As a result, three analytical studies were planned and conduced; 1) Models of Teamwork, 2) Standardized HAC Performance Measurement Battery, and 3) Initiators and Triggering Conditions for Adaptive Automation. Additionally, one field study was also conducted at Idaho Falls Power, which is described in Section 2. The three analytical studies are described in detail including a discussion of the result and conclusions in Section 3 – Models of Teamwork, 4 – Standardized HAC Performance Measurement Battery, and 5 – Initiators and Triggering Conditions for Adaptive Automation. The last section in the report, Section 6 provides a high-level summary of the general conclusions from the different studies as well as a discussion of future research.

This report describes the activities conducted, mentioned above, to meet the DOE milestone M3SR-14IN1301053 – Conclude Initial Studies for Evaluation of Human-Automation Collaboration Performance.

2. HAC FIELD STUDIES

Before the research team develops guidance on how to effectively design HAC for AdvSMR, it is necessary to understand operating experience regarding HAC in highly automated systems. Because there is little to no operating experience with highly automated systems for day-to-day operations in the nuclear field (let alone the AdvSMR field), researchers have sought input from other process control facilities that use high degrees of automation.

Researchers plan to visit multiple process control and power generation facilities. The objective of the field study conducted to date was to determine how process control and power generation industries handle HAC in highly automated systems. Specifically, researchers sought answers to the following questions:

- What is/are the plant mission(s)? How does the plant mission relate to that of a traditional LWR? How does it relate to AdvSMR?
- What functions and tasks are assigned to automation versus human operators? How 'automated' is the system?
- What are some of the difficulties that operators encounter when interacting with these highly automated systems? Are there negative impacts on human-system performance?
- How does the HSI support Human-Automation collaboration? What problems do operators encounter when interacting with automation via the human-system interface (HSI)?

2.1 Method

The first field study was carried out at Idaho Falls Power (IF Power), which is an electric utility that provides power to 22,400 residential and 3695 commercial customers. Approximately 30% of the power is generated by the utility's 5 hydroelectric dams.

When visiting IF Power the researchers conducted the following activities to address the questions related to HAC mentioned above:

Plant Questionnaire. The majority of the questions on this questionnaire were related to the plant mission, the approach to function allocation, and information about operators (e.g., number and qualifications) and operator training. This questionnaire was administered on-site in an informal interview setting.

Review of plant HSI. Researchers were provided with a list of features to review (e.g., alarm system, operating procedures, controls, etc.). They will conduct the review by touring the control room with an operator. They will inspect the HSI elements, and ask the operator for additional information where appropriate.

Operator Interviews. The researchers were provided with a list of interview questions to ask the operators. These questions focus on what tasks the operator carries out on a typical day, how the operator interacts with the automation and what challenges he/she perceives associated with the HAC design (e.g., it is too automated, or the HSI doesn't provide enough information). In addition the operators were asked about any accident scenarios they have encountered, and how the HAC design supported (or hindered) the process. These interviews were conducted in a semi-structured manner during the plant visit.

2.2 Results from Idaho Falls Power

2.2.1 Plant mission

There are two main missions of IF Power. The first is power distribution. Power is distributed via 12 substations that are operated remotely from the control room.

Another main mission of IF Power is electrical generation. The hydroelectric dams at IF power supply about 30% of the power needed for Idaho Falls.

A secondary mission of the facility is to maintain the river level for the waterfalls in downtown Idaho Falls. By government mandate, IF Power must maintain river level within a tolerance of ± 1 foot. However, in order to maintain the waterfalls, they maintain river level within a ± 1 inch tolerance.

The final mission of IF Power is to protect public safety. The main issue related to public safety is the potential for people to fall into the river and be caught in one of the plants.

The missions of IF power are similar to those of LWRs and AdvSMRs in that one of the main missions is the generation of electricity. However, IF power is different in that power distribution, rather than electrical generation, is its primary mission. In the case of AdvSMR, the main mission may be to supply process heat rather than electrical generation, thus the missions of IF power may be more similar to AdvSMR than LWRs due to the fact that electrical generation is a secondary mission. Despite these differences, the missions of IF power are similar to those of both AdvSMR and LWRs.

2.2.2 Approach to Human automation collaboration

The majority of functions performed at IF Power are fully automated. For those functions, the operators serve a supervisory control role (monitoring the system and intervening when necessary). However, there are a few functions that are fully manual. Overall the system is highly automated, with a few manual tasks assigned to the human operators. In general, the humans and automation perform their assigned functions and tasks independently. The humans and automation do not have any tasks that are shared, and the only interaction is that the operator supervises the automatic system.

2.2.2.1 Functions and tasks that are fully automated

Power distribution at IF power is fully automated, however, the operator may need to isolate parts of the grid for maintenance (switching).

The power plants are fully automated; however operators can take manual control actions from the control room, if necessary. The plants will automatically synchronize to the grid, but operators will sometimes synchronize the plants to the grid manually for refresher training/skill retention. Additionally, if a plant has tripped, operators typically synchronize it to the grid manually, not automatically.

2.2.2.2 Functions and tasks that are fully manual

Maintaining the river level is a fully manual task. The operator will do this by:

- Monitoring, watching, and controlling river level and flow
- Watching dam output and predicting impact on river level
- Monitoring weather, temperature, and winds
- River levels are controlled by opening or closing plants (i.e., increasing or decreasing the output of the turbine)

Maintaining river level is a knowledge-based task and it requires estimation and mental calculations, adjusting gates, evaluating river response, and adjusting again.

Another fully manual task is to dispatch trucks and personnel as needed. The operators coordinate efforts during outages from the control room.

Starting and stopping plants is typically executed manually, however once they are started, the plants run autonomously.

The operators also manually perform load forecasting (used to determine amount of power to purchase) with Utah Associated Municipal Power Systems (UAMPS).

Finally, operators manually coordinate work in the field and tag-out/lock-out during maintenance activities.

2.2.3 Description of the human-automation collaboration

The HAC design at IF is best described as human agents and automation agents working to accomplish their tasks independently. The human monitors the automatic process, and dispatches appropriate personnel to deal with any failures. There are no apparent situations in which the human operator and the automation must collaborate, per se. Therefore it is difficult to draw general conclusion regarding optimal human-automation collaboration based on this study. Nonetheless, there are a number of useful insights regarding human automation interaction in highly automated systems that can be drawn from this work.

2.3 Discussion and Conclusions

Though a large body of experimental research has demonstrated issues with trust-in-automation including complacency related to highly reliable automation (Dixon & Wickens, 2006; Wickens et al., 2010), and underutilization of automation (Ruff et al., 2004; de Visser & Parasuraman 2007), the operators at IF power did not report any of these issues. Operators reported that they trusted automation to do what it was designed to do 100% of the time. The operators also reported that they could not recall a time when the human operator overrode the automatic system. While operators reported a high degree of trust in the automation, they stated that they do not trust automation when human life is at stake. They stated they thought that it was good design that maintaining river level is a manual task. River level has the potential to impact human life, and the automation is not designed to value human life.

One of the main issues identified with high levels of automation is a reduction in situation awareness (Endsley, 1996, 1997; Endsley & Kaber, 1999; Endsley & Kiris, 1995). This reduction in situation awareness (SA) may lead to the operator failing to properly monitor the automation, and consequently miss an automation failure. Further, a reduction in SA could also lead to the operator having difficult regaining manual control in the event of an automation failure. It was difficult to assess the issue of reduced SA at IF power. The operators stated that when an automatic system fails in the context of this plant, the operator/dispatcher rarely needs full awareness of what the automatic system was doing in order to recover. They simply send maintenance personnel out to fix the problem.

Another insight provided by the operators is that they manage to stay engaged in the overall process because they have well-defined, specific, and necessary roles to play as part of the process. The automatic system is not equipped to do the tasks that they do manually, therefore they serve an essential function other than their supervisory control roles. Additionally the plant functions are not complex in-and-of themselves, but together comprise a complex process, which serves to keep the operator engaged because he has a wide variety of tasks to execute.

The overall roles and tasks of operators at IF power are different from a typical LWR operator, but could potentially be quite similar to operators of anticipated AdvSMR designs. An IF power operator's main role is to monitor the automatic processes and dispatch personnel and resources in the case of an event (typically and outage). This is very similar to the roles described in NUREG/CR-1368 which states that the human operator's primary functions would be to monitor and verify performance of safety systems, maintain communication with appropriate onsite and offsite personnel, and initiate recovery actions following an event (NRC, 1987). Therefore, the insights gained form this study are applicable to the AdvSMR domain.

2.3.1 Path Forward

The research team plans to conduct several more site visits and combine the results from multiple process control facilities. Other facilities that have been identified as potential candidates are:

- Coal fired plants in Idaho
- Monsanto phosphate facility (Soda Springs, ID)
- Simplot fertilizer facility (Boise, ID)
- Other fossil fuel plants

3. THE DEVELOPMENT OF A MODEL OF HUMAN-AUTOMATED AGENT TEAMWORK FOR SUCCESSFUL TEAM BEHAVIOR

3.1 Introduction and Objective

As described in (Oxstrand et al., 2013), teamwork between human and automation was identified as one of the topics needed to be addressed with further research. It was decided to conduct an analytical study to further broader the teams understanding about teamwork and its implication in the context of HAC and AdvSMRs. Considerable research has been conducted to address how to make automation "team players" (e.g., Christoffersen & Woods, 2002; Land, et al., 1995; Lenox, et al., 1998; Steinberg, 2012). This is a logical result of the findings that poorly designed human-automation collaboration (HAC), such as poor communication between automation and it human supervisor, often leads to problems for operators, such as:

- Undesirable changes in the overall role of personnel
- Difficulty understanding automation
- Poor monitoring, lack of vigilance, and complacency
- Out-of-the-loop unfamiliarity and situation awareness
- Workload to interact with automation and when transitioning to greater manual control
- Loss of skills for performing tasks automation typically performs
- New types of human error

For multi-agent systems, designing automation to be a good "team players" has typically based on an implicit notion of what it means to be a team player and how members of a team should perform to function successfully. General concepts of team characteristics and behavior are employed, such as trust, goal and intention sharing, cooperation, and redundant responsibilities (especially in the case of adaptive automation where shifting of responsibility is a hallmark of the approach). These concepts are based loosely on a sense of how teams of humans perform.

However, the concept of multi-agent teamwork relies considerably on simplifications and popular notions about what is needed to foster teamwork, which do not always transfer well to human-automated agent teams. That is, the work to define how automation should behave (be designed) to be a team player has not been based on the recognition that there are different models of human teamwork, each with its own set of member responsibilities and behaviors (Salas, Cooke & Rosen, 2008). As a result, prior work on identifying how automation can be a team player is fragmented and incomplete at best.

Further, even if a belief about how human teams behave is at the core of this research, it does not sufficiently address the fact that automation agents are not humans and cannot completely fulfill the role of a crewmember nor can we expect that it will fully behave as a human member of a team will. For example, automation agents cannot assume responsibility; automation can be given the authority to act, but humans always maintain responsibility (Pritchett, 2001; Sarter & Woods, 1992). As another example, automation is not "concerned" about the consequences of its actions nor is it as able to innovate as human crews will do when things do not go as planned.

To ensure that guidance on the design of successful HAC in AdvSMRs is developed, the objective of the teamwork analytical study is to identify the requirements for effective human-automation teamwork that are appropriate in a commercial nuclear power team environment. Once identified, the researchers can elaborate on the specific requirements of human and automation team members, and can use those requirements to develop guidance for what automation characteristics and attributes are needed to be good "team players."

3.2 Method and Technical Approach

The researchers intended to identify what makes for effective teams, in particular effective teams in the nuclear domain. To do so, the researchers evaluated different models of human-human teamwork. As a part of the evaluation the researchers conducted:

- Review recent research on models of human teamwork.
- Emphasize nuclear power plants and other complex, highly automated systems.
- Identify general principles about what makes for effective teams and the characteristics needed.

Additionally, the researchers investigated how human-automation teams are different from human-human teams. The researchers identified how automation agents differ from human agents, developed a characterization of human-automation teams, and assessed the implication of agent roles within a team (e.g., in higher LOA systems, autonomous agents may act independently from their human teammates; while in lower LOAs automation may play only a supporting role).

3.3 Results

3.3.1 Human-Human Teams

The research team conducted a review of the literature on human teams and teamwork. This section summarizes the team's findings.

Two books that are generally considered classics in this business leadership genre were written by Larson and LaFasto (1989) and Lencioni (2002). Larson and LaFasto (1989) presented eight factors that are important to successful teamwork: (1) a clear and elevating goal, (2) a results-driven structure, (3) competent team members, (4) unified commitment, (5) a collaborative climate, (6) standards of excellence, (7) external support and recognition, and (8) principled leadership. In a related vein, Lencioni (2002) presents a simple model of five factors that can undermine team performance. The "five dysfunctions" are: (1) absence of trust, (2) fear of conflict, (3) lack of commitment, (4) avoidance of accountability, and (5) inattention to results. According to Lenconi, teams need to overcome these issues in order to be effective. All of these factors make intuitive sense. If team members do not trust one another, are afraid of conflict, lack commitment, etc., the likely outcome is suboptimal team performance.

The factors identified by Larson and LaFasto (1989) and Lencioni (2002) are anthropomorphic attributes that still apply to human agents in human-automation teams. For example, the dysfunction, absence of trust is more appropriately applied to the human in human-automation teams, and in this case, translates to a well-known human factors issue called trust in automation. In addition, in situations where the human operator is in a supervisory role, the automation is highly reliable, and the human operator is not invested in the outcome, other dysfunctions such as lack of commitment, avoidance of accountability, and inattention to results may also apply. Translating these team attributes, however, to the automation agent in human-automation teams is somewhat more challenging. These insights on teams from the business leadership literature are likely to be of limited use to human-automation teams. While these factors can be applied directly to the human agents, there is difficulty in extrapolating them to the automated agents in human-automation teams. Automation's performance is not necessarily improved by having a goal that is elevating versus a goal that is not, and is not motivated to work harder or collaborate more closely with other team members by external support and recognition. Furthermore, automation does not trust or mistrust the human, does not experience fear, does not have an understanding of what commitment or accountability is, does not attention issues in the same way humans do, and is agnostic about results.

3.3.1.1 Cognitive Models of Teams and Teamwork

A considerable amount of human factors research has been conducted on teams, team performance, and in particular team cognition and team collaboration. The research on cognitive models of team and teamwork is based on the premise that in addition to the factors that Larson and LaFasto (1989) and Lencioni (2002) identified, how the team collectively thinks and coordinates cognitive and physical efforts are important elements to team success. A number of models of team have been developed from the human factors perspective to help elucidate this premise.

Perhaps the most well known model of teams is the Crew Resource Management (CRM) model Helmreich and Foushee (1993). CRM originated in the 70's in response to concerns for safety in aviation and evaluates teams from an interpersonal perspective. Helmreich and colleagues at the University of Texas and NASA/Ames Research Center established research to analyze flight crew behavior, develop measures to evaluate crew performance and create training to enhance pilot team coordination and improve aviation safety. Later, Helmreich and Foushee developed a three-factor CRM model. The model focused on three critical factors of group behavior: input factors, crew performance or group process factors, and outcome factors. Input factors include characteristics of individuals, groups, organization, and the operational environment. Group process factors include the nature and quality of interactions among group members. Outcome factors include primary outcomes such as safety and efficiency of operations as well as secondary outcomes such as member satisfaction, motivation, and attitudes. The model has an underlying assumption in that the input factors provide the framework and determine the nature of group processes that lead to the determined outcome. Also significant to the model is the presence of feedback loops between the individual components (See Figure 1).

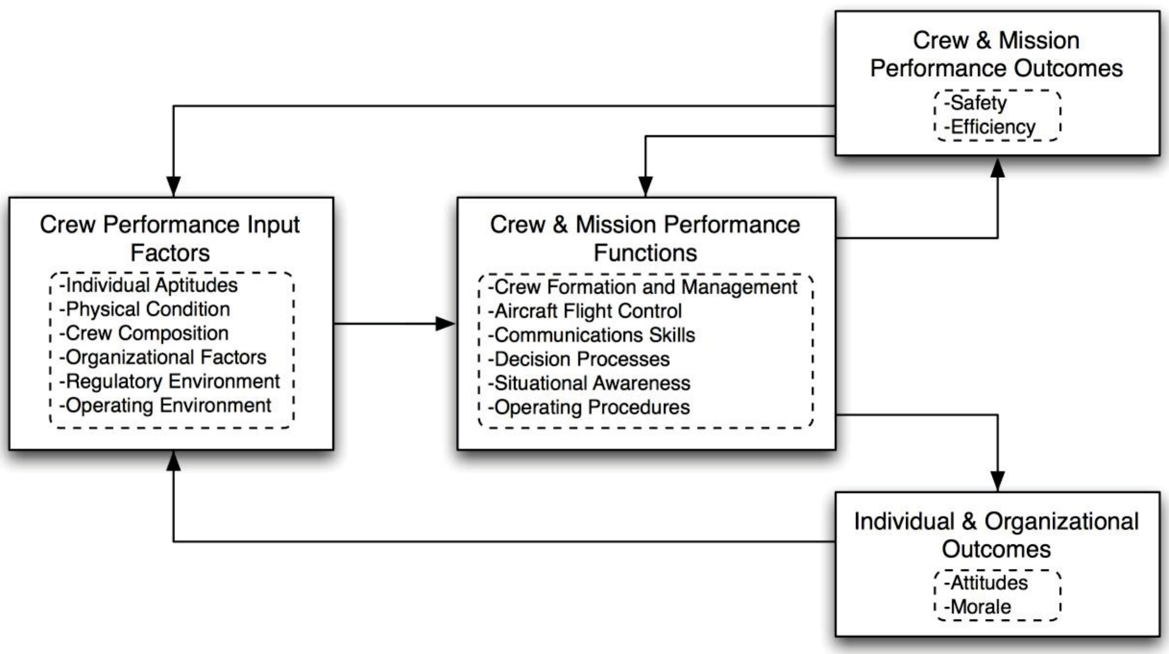

Figure 1. Crew Resource Management Model (from Helmreich & Foushee, 1993).

Another model by Dickinson and McIntyre (1992) called, the Teamwork Model, was sponsored by the Naval Training Systems Center and was based on results of the TEAM Model (Morgan, Glickman, Woodard, Blaiwes, & Salas, 1986) along with a comprehensive literature review of teamwork literature. This work was carried out at one of the Naval Combat Information Centers (CIC) aboard a Navy vessel. The researchers describe the CIC as the location in which "information from a number of sources must be processes and acted on by many individuals before a final decision can be made to deal with a threat. Based on their extensive literature reviews, the researchers identified seven core components of teamwork (see Figure 2) including:

a. *Team Orientation* – attitudes that team members have towards one another and the team task.

b. *Team Leadership* – providing direction, structure and support for other team members (not necessarily one leader over others; can be team leadership)
c. *Communication* – exchange of information between two or more team members, possibly to clarify or acknowledge receipt of information.
d. *Feedback* – giving, seeking, and receiving of information between team members.
e. *Backup Behavior* – assisting the performance of other team members.
f. *Monitoring* – observing activities and performance of other team members with understand of other members' tasks.
g. *Coordination*

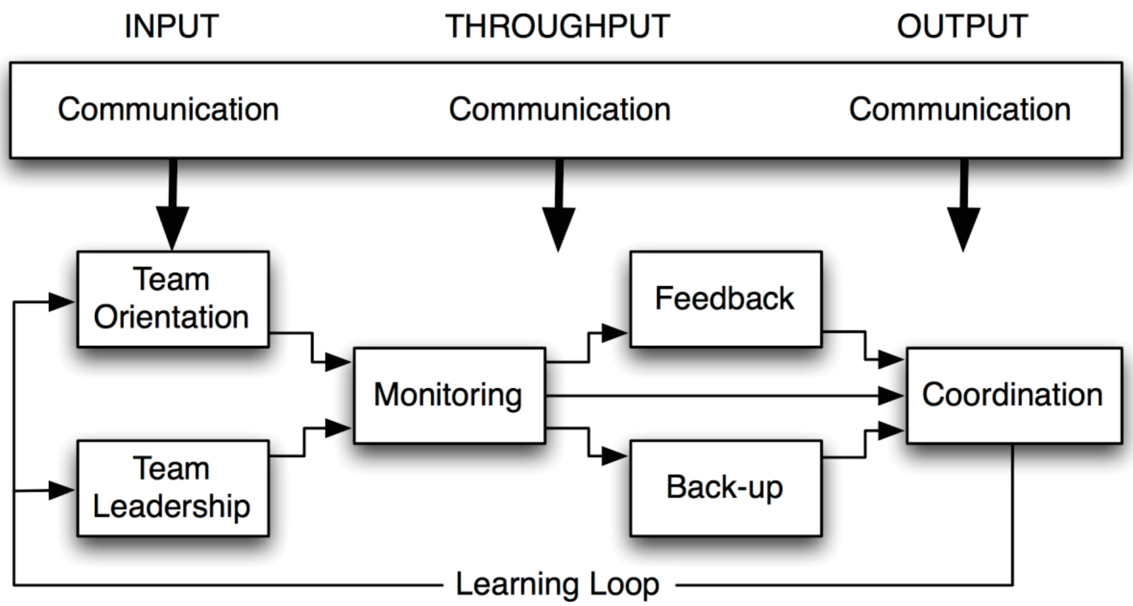

Figure 2. Teamwork Model (from Dickinson & McIntyre, 1992).

Letsky and his colleagues (Fiore, Smith-Jentsch, Salas, Warner, & Letsky, 2010; Letsky, Warner, Fiore, Smith, & Salas, 2007; Letsky, Warner, Fiore, & Smith, 2008) have been the main developers of a model of team collaboration that includes meta-cognitive and macro-cognitive attributes. This model of team collaboration was also developed based on military research, and it identifies meta-cognitive and macro-cognitive processes that are specific to the types of teams that are often seen in military work: asynchronous, distributed, multi-cultural, and hierarchical. The core of this teamwork model includes four major "team collaboration processing" stages, which are very similar to the information processing stages of many models of individual cognition. The four stages are: 1) individual knowledge construction, 2) collaborative team problem solving, 3) team consensus, and 4) outcome evaluation and revision. These four stages are depicted in the ovals in Figure 3 below. "Individual knowledge construction" is the idea that all members of the team have sensory perception capabilities (and limited attention resources) that allow them to detect or notice salient stimuli in environment (but miss others), and based on the stimuli that is detected, construct meaning or knowledge of what the situation or environmental context is. "Collaborative team problem solving" is the emergent team process whereby individual team members communicate and share their interpretations and understanding of what the situation is, as a means to solve the problem they believe is being presented. Specifically, this usually involves a) individuals making assertions that the stimuli presented are symptomatic of a problem, b) the team 'comparing notes' to determine if there is agreement that the symptoms presented are consistent with their understanding of the problem, and c) individuals proposing solutions that can mitigate this problem. The next stage, "team consensus" involves the team (or sometimes just the team leader) deciding which of the proposed solution to implement. The last stage, "outcome evaluation and revision", involves all team members monitoring the ‘system’ response to the solution they implemented to evaluate whether the perceived problem is corrected.

It is also important to note that this model of teamwork highlights the fact that the four different team collaboration processing stages are the result of both meta-cognitive and macro-cognitive processes, and are also affected by a number of different “problem area characteristics” (i.e., performance shaping factors). Meta-cognitive processes are higher-level processes than macro-cognitive processes in that they are the cognitive efforts the team puts forth to synthesize macro-cognitive activities. Meta-cognitive processes appear to correspond to the conduct of operations and standard operating procedures for the team. As such, meta-cognition directs the overall team problem solving effort. Macro-cognition is a comprehensive term that describes both individual cognitive processes (i.e., individual mental model construction) and team cognitive processes (i.e., team shared understanding development and team negotiation of solution alternatives) that emerge at various points in the stages of team collaboration. The “problem area characteristics” include situational factors, individual differences in team members, and unique characteristics of their operating context that can affect team collaboration and overall system performance.

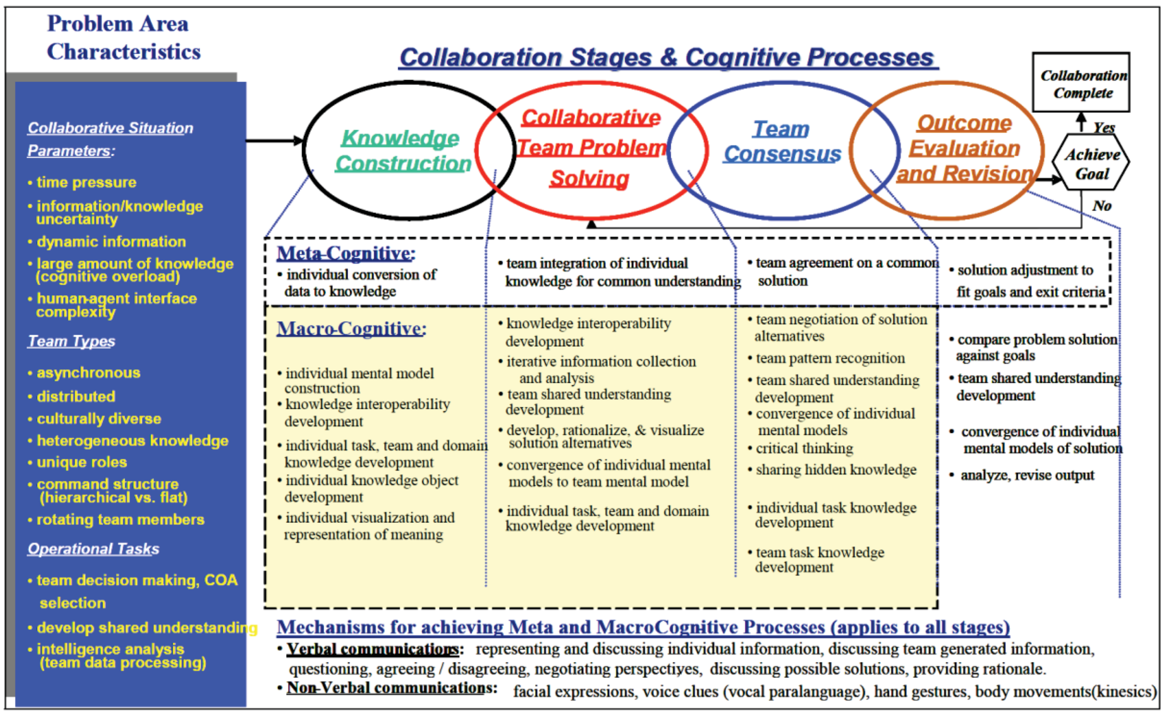

Figure 3. Model of Team Collaboration (from Warner, Letsky, & Cowen, 2004).

Fiore et al. (2010) further expanded upon this model of team collaboration by further elucidating the team knowledge building process that is also occurring during collaboration. The knowledge building process is an emergent process in that each individual on the team is involved in his or her own independent process of collecting Data, converting that data into Information, and ultimately into Knowledge (referred to as the D-I-K process). The fact that individuals engage in the D-I-K process in parallel makes it emergent. As Figure 4 shows, parallel individual processes called “individual knowledge building processes” generate multiple versions of “internalized knowledge,” which then merges together and become the “team knowledge building processes.” This is a more detailed explanation of one of the emergent aspects of team coordination.

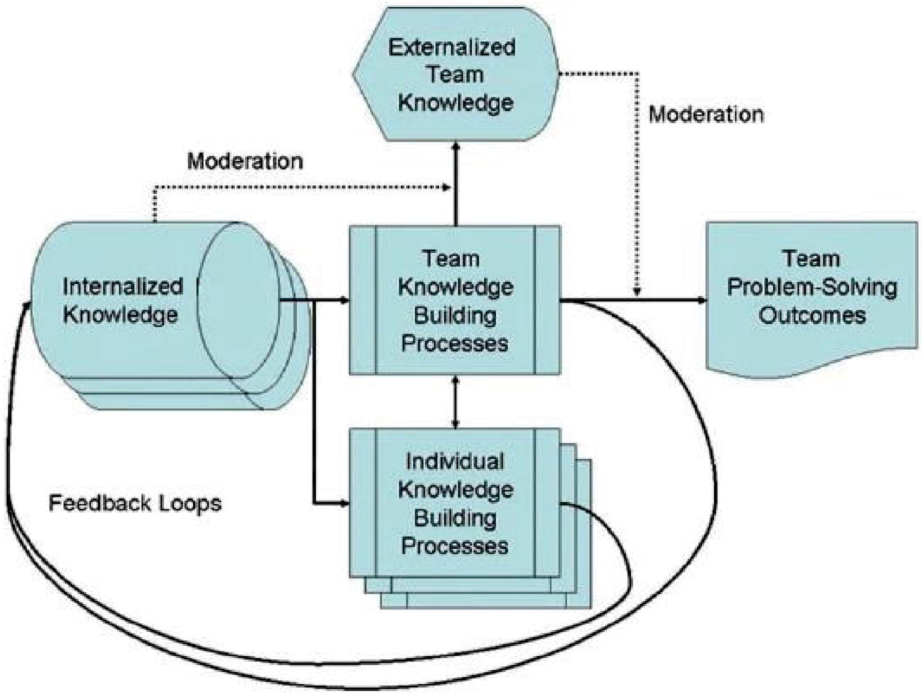

Figure 4. Knowledge Building Process Within Team Macrocognition (from Fiore et al., 2010).

That is, within this model, there is a clear distinction between individual cognitive processes and the truly emergent team cognitive processes.

3.3.1.2 Human Factors Research on Specific Aspects of Team Cognition or Teamwork

The three models of team cognition presented, CRM, the Teamwork Model, and Macrocognition, all are relatively high-level models of team cognition. They all share a common feature of depicting team cognition much in the same way individual cognition is depicted, which in its more basic form is a linear "input – cogitation – output" process model. While these models do go into more specific detail underlying each stage of the linear process model and include feedback loops and other details (e.g., performance shaping factors), there is also some additional research on specific team information processing stages that is worth mentioning. Work on team sensemaking is a prime example, as it is another way of explaining the emergent cognitive processes that occur in teams.

Team sensemaking is defined as the process by which a team manages and coordinates its efforts to explain the current situation and to anticipate future situations, typically under uncertain or ambiguous conditions. Team sensemaking is very similar to individual sensemaking, but Klein, Wiggins, and Dominguez (2010) argue that in many ways, it is more critical than individual sensemaking, because team sensemaking poses additional emergent coordination requirements and has additional ways for sensemaking to break down. They argue team sensemaking is more difficult to accomplish, and it may be a larger contributor to accidents than failures at the individual level.

Like individual sensemaking, Klein, Wiggins, and Dominguez (2010) describe team sensemaking as a process by which teams reconcile data with their existing frames (i.e., mental models of the situation). The steps for team sensemaking are:

1. Identifying a frame to given meaning to the data deemed important enough to be captured
2. Questioning the appropriateness of the frame identified, given the data captured
3. (a) Preserving and elaborating on the original frame, or (b) Re-framing/Creating a new frame.

However, Klein, Wiggins, and Dominguez correctly point out that there are emergent aspects with team sensemaking, which pose additional coordination requirements and provide additional ways for sensemaking to fail. Specifically, how the team decides to work together as they are trying to make sense of their situation is an emergent feature. That is, in team sensemaking, individual team members are engaged in their own internal sensemaking process, but ultimately the team must decide on how the different individual sensemaking processes will be reconciled. The general strategies in which teams work together are Hierarchical, Collaborative, and Opportunistic (i.e., everyone on the team has the same rank and responsibility to detecting problems and prescribing corrective actions). Furthermore, the emergent team process of how to reconcile individual sensemaking processes applies to all steps of team sensemaking—Identifying, Questioning, and then Preserving/Elaborating or Re-framing.

For example, if the team is identifying a frame, one strategy for reconciling the different individual's frame identification is hierarchical, where the team leader will decide the official conclusion. Another strategy is to use consensus (Collaborative). This reconciliation/resolution process is emergent in teams, because individuals will be going through their individual sensemaking processes in parallel, and these parallel processes must eventually be merged and/or narrowed down to one output. This kind of parallel sensemaking processes does not occur in individuals.

The researchers elaborated on this emergent concept by explicating a set of "behavioral markers" for team sensemaking, which are specific examples of the general strategies teams can employ (i.e., Hierarchical, Collaborative, and Opportunistic) for each step of the sensemaking process, and can serve as the basis for inferring performance shaping factors that could affect team sensemaking. Table 1 below lists the team sensemaking behavioral markers.

Table 1. Team Sensemaking Behavioral Markers (from Klein, Wiggins, & Dominguez, 2010).

Team Sensemaking step	Behavioral markers
Identifying a frame	Team formulates criteria or rules used to identify the frame A team member announces the frame Team collaborates to identify the frame
Questioning a frame	Appoint a team member to play devil's advocate and raise doubts about the suitability of the frame Team creates rules or tripwires to alert them that the frame may be unsuitable Team members voice and discuss what might go wrong using the current frame
Preserving and Elaborating a frame	Team discusses and rejects anomalous data as transient signals or otherwise insignificant The team's data synthesizers direct the activities of the data collectors to seek new data to verify the frame The team's data synthesizers and data collectors collaborate to discover new relationships that preserve or extend the frame
Reframing: comparing frames	Team compares frames and votes for one Team forges consensus on which frame is most appropriate Leader announces, which frame is most appropriate
Reframing: creating a new frame	Individual suggests a frame and it is adopted, modified, or rejected as the team compares frames Team speculates on data and suggests causal beliefs; leader or a team member combines viewpoints into a frame Team collaborates to synthesize competing frames

Research by Soraji et al. (2012) on a team's Mutual Belief Model (MBM) is related to the research on team sensemaking, except the sensemaking effort is not about external stimuli per se; the sensemaking effort in MBM is the process by which individual team members compare and reconcile their 'frames' or in MBM terminology, cognition and beliefs (about the other's cognition). In the simplest case of a two-person team (i.e., Person A and Person B), in order for the two people to be on the same ‘page,’ MBM posits that the following steps must occur:

1. Person A and Person B perceive an external stimuli that based on their training and experience is, for example, an indication that the ‘system’ is in a particular undesired state (e.g., unexpected transient), and that a particular action or set of actions will return the system back to a desired state. All of this parallel mental activity occurring independently in Person A’s and Person B’s mind is collectively defined in MBM as Person A’s cognition and Person B’s cognition.

2. If Person A and Person B are going to work effectively as a team, Person A then has to formulate a belief about Person B’s cognition, and Person B has to formulate a belief and Person’s A cognition, as a first step to determining if their independently formed cognitions about the situation are consistent or inconsistent with each other. The formulation of these beliefs about the other’s cognition would be based on what behaviors each person was able to observe in the other (e.g., what alarm(s) did my teammate attend to?), and their past experiences working with each other (e.g., the last time we had this set of annunciators, my teammate concluded the system was correctly experiencing an unexpected transient), and then confirmed through communication.

3. The next step in the process of reconciling Person A’s and Person B’s cognitions and beliefs about the other person’s cognition and belief is the formation of a second set of beliefs whereby: a) Person A forms a belief about Person B’s beliefs about Person A’s cognitions, and Person B forms a belief about Person A’s belief about Person B’s cognitions. This is essentially a mutually confirmatory and recursive process to verify that, ‘Person A knows that Person B knows what Person A was thinking,’ when the external stimuli presented itself, and ‘Person B knows that Person A knows what Person B was thinking,’ when the same external stimuli presented itself.

The MBM steps described above are depicted in Figure 5 below.

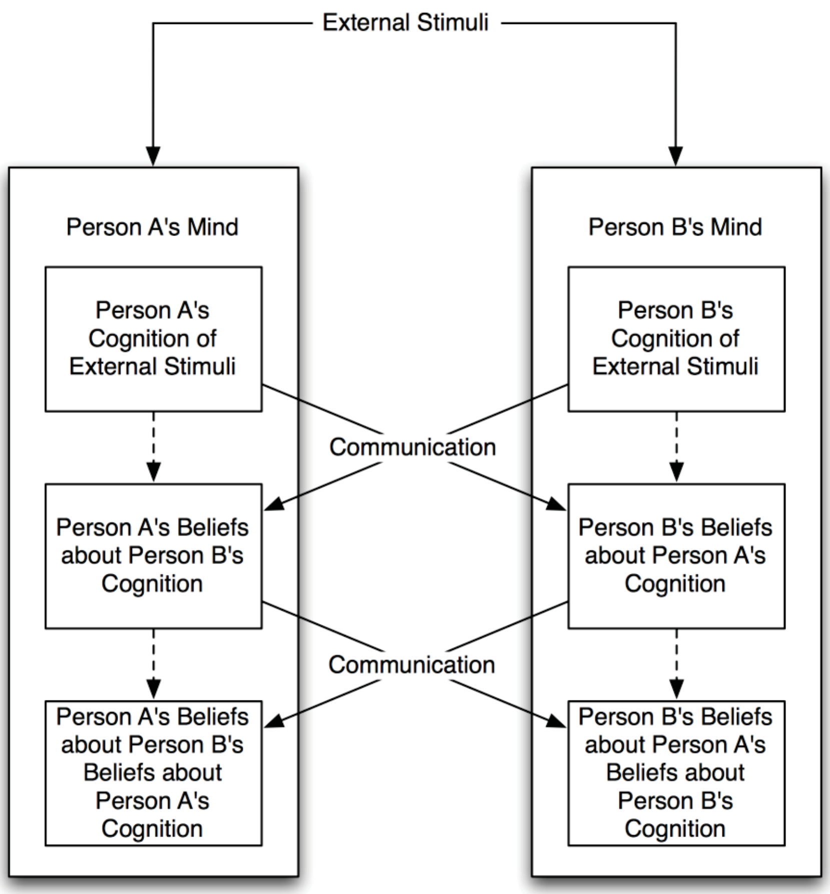

Figure 5. Mutual Belief Model (from Soraji et al., 2012).

3.3.1.3 General Requirements or Principles for Effective Human Teamwork

Based on all of the literature reviewed the researchers derived a set of general requirements or principles for effective human teamwork.

Belief in the Concept of Team

The first principle of teamwork is that individuals who form the team believe (or they are told that their compensation depends on) the idea that there is a mutual benefit to working together (i.e., that what binds them together is greater than what drives them apart, and that if they act on that belief they will achieve better results than if they were working separately). Dickinson and McIntyre (1997) referred to this as "team orientation," which relates to whether team members have positive or negative attitudes towards the each other, the leader, and their task(s) and whether each individual believes it is worth it to stay in the group or leave. The translation of this principle to human-automation teams is a challenge in that automated agents do not have motivations like humans do to achieve better results, nor do that gain a sense of camaraderie that comes from achieving a goal through a collective effort. They also do not have the ability to decide or choose capriciously (i.e., free will) whether to act as a fully engaged team member, a partially engaged team member, a disengaged team member, or as an agent that is actively working against the team. They perform in the way that they are programmed, or fail (hopefully with some predictability) to function as programmed.

Effective Communication

The second principle of teamwork is effective communication. Communication is the central behavior team members engage in to function as a team. It is central to:

- Exchanging information
- Establishing team situation awareness
- Coordinating and regulating individual efforts (i.e., communication links monitoring and feedback on other members' performance)
- Building trust among team members (e.g., inaccurate, confusing, and/or erroneous transmissions, and/or difficulties in comprehending or interpreting the information transmitted can decrease trust).

The translation of this principle to human-automated teams is also a challenge to varying degrees. While most automated agents are capable of exchanging information in a manner that is technically correct, only well designed and well programmed agents exchange information in a manner that also effectively improves team situation awareness (i.e., enhances the human's comprehension of the situation and other team members' intentions and actions), and facilitates (i.e., coordinates and regulates) teamwork. Furthermore, trust in automation is a well-established issue in human factors, and research by Sarter and Woods (1997) have shown that communication problems attributable to the automation are a significant factor in the human operator experiencing "automation surprises," and subsequently losing trust in the automation.

Team Leadership

The third principle of teamwork is that it requires effective team leadership. Strong leadership by individuals formally appointed as leaders, and informal leadership by other team members, is well established as a key component to teamwork. Larson and LaFasto (1989) argue that good leadership is principled and consistent, and has both transactional (i.e., the management of subordinate team members through monitoring performance against criteria defined in formal work agreements), and transformational (i.e., leading the team by creating a vision and inspiring team members to strive for higher level goals) qualities. Furthermore, they argue that leaders who do not let their personal egos overshadow the contributions of the rest of the team tend to be more effective than those who take more than their share of the credit and less than their share of the blame. The translation of this principle to human-automated teams is also potentially challenging in that automation is indifferent to these attributes of team leadership, and does not respond to these aspects of teamwork in the same way that humans do. In the event that automation is in a leadership position (which is currently difficult to envision) it is unlikely that the automated leader will effectively convey these leadership attributes to the human subordinates in a convincing manner, or that the human subordinates' ability to perform well as a team will benefit from these 'soft skills'.

Monitoring Individual and Group Performance and Providing Feedback

Monitoring and providing feedback on both 1) overall group performance and 2) individual contributions to the overall performance is the fourth principle of teamwork, and is very important. The popular management maxim, "If you can't measure it, you can't improve it," relates to the idea that it is important to monitor team performance as a prerequisite to obtaining feedback on performance. Feedback is key the systematic adjustment of process parameters that are within the span of control of operators in order to achieve the process outcomes that are desired by the decision makers. It is how the agents learn and adapt processes and system parameters to achieve desired outcomes. Monitoring individual contributions to overall performance is also key to mitigating a well-known social psychological phenomenon called social loafing, where individuals working in groups tend ride on the coattails of the collective team performance by putting forth less individual effort than they would if working alone or if their individual work contributions could be measured (Latané, Williams, & Harkins, 1979). Furthermore, both the communication and team leadership principles (from above) have an element of monitoring and providing feedback on performance. Team leadership is the principle that gives the authority to the leader to monitor and provide feedback on the performance of others on the team, and communication is the means through which feedback on the activities performed by the subordinate and monitored by the supervisor is conveyed to the subordinate.

This principle is conceptually fairly easy to translate to a human-automation team, but can be difficult to implement. In the case of the automated agent, monitoring and feedback require that the automated agent have sensors available and capable of monitoring the human operator, the algorithms programmed to know how to evaluate the operator's performance, and the correct decision logic programmed to provide the correct feedback to improve the operator's performance. In the case of the human operator, it can be difficult to monitor what the automated agent is doing, as most of its processing occurs quickly and out of 'view' of the operator. Furthermore, unless the automated agent is programmed with learning algorithms, providing feedback to the automated agent will not change its behavior. Their behavior will only change when the operator or programmer changes their programming parameters.

Coordination and Assistance

The fifth principle of teamwork is the coordination of individual efforts. It is when teams are well coordinated that performance above and beyond what each individual could achieve on their own can be realized. As Dickinson and McIntyre (1997) put it, "Successful coordination implies the effective operation of other components of teamwork (e.g., communication, monitoring, and backup). In this way, the actions of individual members are merged to produce synchronized team performance" (pg. 22). Included in the principle of coordination is the extent to which team members have overlapping capabilities and are willing and able to seek and provide assistance to one another (i.e., provide redundancy or defense in depth) and increase overall system resiliency. The principles of coordination and assistance are not inherently difficult to translate to human-automation teams. However, within highly complex industrial process control systems, coordination and assistance are typically organized through clearly defining the roles and responsibilities, as well as assigning accountability and authority, among all agents. Given this approach, the translation of this principle to human-automation teams becomes somewhat more challenging; depending on what aspect of coordination is being applied. Defining roles, responsibilities, and who has authority is a fairly straightforward process with human-automation teams, but accountability (i.e., a sense of responsibility for the outcome) is an anthropomorphic concept that does not apply to automated agents. Automation is not differentially motivated by the idea that its performance can and will be evaluated and subsequently rewarded or punished depending on how well it performed.

Awareness of Internal and External Performance Shaping Factors That Affect Team Processes

The sixth principle of teamwork is that the team is aware of how internal and external performance shaping factors affect team processes. As noted in research previously summarized, personality traits and cognitive abilities can vary from person to person, and how the team not only manages, but leverages individual differences to enhance team performance is important. Similarly, external performance shaping factors, such as task complexity, time pressure, and information/knowledge uncertainty can affect the team's ability to collaborate, which in turn can affect overall system performance. This principle is not, in theory, difficult to translate to human-automation teams. All of these performance shaping factors would affect the human team member regardless of whether their fellow teammates are humans or automated agents, though there are probably some nuanced differences in how theses performance shaping factors affect the human when his or her teammate is another human versus an automated agent. Strictly speaking, automated agents do not have different personalities, but they do vary in their physical and cognitive abilities and can vary in how they are programmed with respect to other aspects, such as communication etiquette. Automation agents also do not feel anxiety in the same way humans do when under time pressure, but an automated agent's performance can be affected by available time, task complexity, and information uncertainty.

Awareness that each individual's mental model is unique and that it is difficult to create shared mental models in teams.

The seventh principle of effective teamwork is the necessity for individual humans to form accurate mental models of the situation, and to clearly communicate their mental models to others. Yet, humans construct their knowledge based on an attentionally constrained ability to detect and process external stimuli, and a less than perfect long-term memory system, which leads to the formation of an idiosyncratic understanding (i.e., mental model) of the situation. Furthermore, humans have difficulty communicating or sharing their mental models with other humans such that they and one or more team members are subsequently 'on the same page of the playbook.' The translation of this principle to human-automation teams is not difficult. Humans will have these challenges regardless of whether their teammate is another human or an automated agent. With respect to applying this principle to the automated agent, much of automation's processing and interpretation of external stimuli detected via their sensors is essentially the automation's cognition, and the automation's 'mental' models of the situation can and are often markedly different from the human's mental model. Furthermore, automation is considerably limited in its ability engage in processes whereby differences between the two mental models can be discussed and reconciled to created a shared mental model that both the human and it can agree or reach consensus on.

In summary, the literature on human-human teams is varied in terms of the genres that have contributed to this literature (e.g., business leadership and human factors), and in terms of the depth of analysis that has been on performed to understand how to improve human-human team performance. Regardless of what depth the analysis was performed at, however, the literature provides some good insights on the challenges and issues with human-human teams, and often provides a range of practical and workable solutions. The extent to which these insights and solutions translate to human-automation teams, and automated agents in particular, varies depending on the finding or principle. Generally speaking, most of the issues with, and recommendations on how to improve human-human teams stem from an understanding of seemingly universal human foibles that automated agents do not necessarily share. For example:

- Humans have beliefs, motivations and emotions that affect their performance.
- Human and human-human team performance is affected by the quality of communication, the extent to which they trust others, the morale of the group, the quality of leadership, and how well coordinated the team is in monitoring performance, providing feedback, and timely assistance.
- There is significant between-person variability in people's values and mental and physical abilities that affects team performance.
- External situational and contextual factors affect individual and team performance.
- The mental models humans develop of complex situations are idiosyncratic, often inaccurate, and are difficult to communicate to others such that they share the same understanding.

For all these shortcomings humans have (and more), the literature on human-human team performance addresses, and provides insights on how to mitigate them. However, automated agents, as the literature review described in section 3.3.3 will show, have some but not all of these shortcomings, and the ways in which a particular shortcoming manifests itself in an automated agent is often different than how it manifests itself in humans. Generally speaking, given that these human foibles do not directly apply to automated agents means that the well-known team dynamics of human-human teams fundamentally change when teams are comprised of human and automated agents. That is not to say that none of this literature is applicable to the new human-automation team dynamic. Rather, the change in team dynamics means that the applicability of the insights from this literature on the new problem space defined by the pairing of humans and automated agents is not simple and straightforward. A corollary to this is that the problems associated with the inherent limitations of automated agents, and the problems their limitations further create in teams, are not directly addressed by this literature's insights and solutions. Additional thought must be applied if the valuable insights from the literature are going to add any value or meaningful guidance to designers contemplating the use of human-automated teams to operate their systems.

3.3.2 Human Teams in the Nuclear Domain

The research on teams, teamwork, and team cognition summarized so far has not taken into consideration the specific nuances and challenges that teams in nuclear power plants must confront. This next section summarizes the research on nuclear power plant crews in conventional large light water reactors, and the unique challenges they confront.

In commercial nuclear power plants, the coordinated activity of multi-person teams accomplishes operations. In today's plants, several reasons dictate the need for teams, including the distribution of workload, and the physical layout of the control room where tasks are completed via dozens of control boards.

The Institute of Nuclear Power Operations (INPO) is one of the main industry drivers for nuclear power plant control room operator training and licensing in that they help establish the requirements and standards for operator training and licensing. According to INPO ACAD 10-001, along with the expected requirements that operators have the fundamental knowledge of technical topics such as nuclear physics and reactor theory, electrical science, and chemistry, INPO also has guidelines on teamwork, and related interpersonal issues such as control room supervision and operational decision-making (INPO 88-003 and SOER 96-1). Specifically, these INPO reports and guidelines cover a variety of teamwork issues including, as directly quote from INPO ACAD 10-001 (pp. 28-29):

- Communications
- Interaction among team members of different personality types
- Anticipation and recognition of error-likely situations (for example, self- and peer-checking)
- Questioning to obtain needed information
- Advocacy of a position or concern
- Command, control, and leadership, and a commitment to achieve team goals
- Constructive conflict management within the control room shift team and with other plant personnel
- Self-critique of individual and team performance

According to INPO ACAD 10-001, INPO expects licensed reactor operators to be well versed in all of these aspects of teamwork as they recognize that teamwork is an important facet of the ability for the crew to function effectively, particularly in novel or emergency situations. INPO also recommends that teamwork training not only be taught in the classroom, but "should be continually reinforced during day-to-day work and training activities." (pg. 28).

According to Stubler, O'Hara, Higgins, & Kramer (1999), a nuclear power plant operating crew is a team in which the members share information and perform their tasks in coordination to satisfy specific operational and safety goals. The crewmembers interact with each other and with other personnel in the plant. The human system interface (HSI) provides the context within which personnel convey information and coordinate actions. The four generic cognitive tasks (i.e., monitoring and detection, situation awareness, response planning, and response implementation) each require interactions between crewmembers. The following discusses these tasks from the perspective of crew performance, and how performance may be affected by upgrades.

Monitoring and Detection

The ability of crewmembers to communicate information about the plant's status to the primary decision maker (e.g., senior reactor operator or shift supervisor) is a major consideration of monitoring and detection. Operators must communicate in a way that allows decision makers to rapidly incorporate this information into their developing situation models.

Communication failures can have many causes. Cognitive causes stem from a lack of a common understanding. Much spoken communication relies on a shared context between the speaker and the listener to support the correct interpretation of the intended meaning. Typical failures involve hidden differences in frames of reference and assumptions (Mumaw et al., 1994). Communication may be impaired to the degree that personnel have different situation models. For example, due to differences in their understanding of the plant's state, an operator may misinterpret a request for information from the decision maker, or the decision maker may misinterpret information provided by an operator. Communication may also be affected by physical considerations. For example, upgrades can change operators' tasks, thus increasing the physical distance between operators (e.g., an operator may be required to spend more time at back panels, or away from the main control panel). Other physical considerations include modifications that affect the line-of-sight between operators, and ones that increase background noise making speech less intelligible.

Situation Awareness

During a transient, the primary decision maker attempts to develop a coherent understanding of changes in the plant's status from available information. Due to limited cognitive resources for seeking information and drawing inferences, the decision maker relies on other crew members for input. Operators in the CR are likely to develop different interpretations of the plant's state from the information they access, and differences in their individual representations of system status. However, if crewmembers are aware of the decision maker's interpretation of events, they can seek confirming and disconfirming evidence. Thus, the interpretation of plant status depends upon the ability of decision makers to convey their interpretations, and the ability of crewmembers to communicate their findings and interpretations. Another element of situation awareness, determining the implications of plant state, may require the decision maker to mentally simulate events to predict future states. Team performance depends upon the decision maker's ability to communicate the intermediate results of these mental simulations to crewmembers so they may provide input (Mumaw et al., 1994).

Response Planning

During response planning, the decision maker formulates plans based on specific goals. Here, interactions between personnel are important, especially when the situation cannot be addressed by straightforwardly applying an existing procedure (e.g., incidents not explicitly addressed by emergency operating procedures; severe accidents). One important skill is the ability of the decision maker to communicate the intermediate results of response planning to others, so they may help evaluate it and contribute additional information. Another is the ability of crewmembers to convey evidence to the decision maker showing that the plan is no longer feasible, such as when plant conditions change quickly (Mumaw et al., 1994).

Response Implementation

Response implementation requires the coordinated actions of crewmembers to execute the selected plan. People must maintain awareness of each other's actions, so that parallel activities are coordinated. In addition, operators may monitor each other's actions to detect errors in responding.

Crew performance is supported when personnel can develop a shared view of the plant's status and control room activities. This may be supported by HSI characteristics that provide them with access to similar sets of information about plant conditions, support their ability to communicate plant information to each other effectively, support coordination of activities by allowing them to maintain an awareness of each other's actions, and support monitoring of personnel's performance to detect errors. Hutchins (1990) identified three characteristics of the work environment that support team performance:

- *Horizon of Observation* – This refers to the portion of the team's task that can be seen or heard by each individual. It results from the arrangement of the work environment (e.g., proximity of team members) and is influenced by the openness of tools and the openness of interactions. By making portions of a task observable, other team members can monitor for errors of intent and execution, and for situations in which additional assistance may be helpful.

- *Openness of Tools* – This refers to the degree to which an observer is able to infer information about another crewmember's ongoing tasks by observing a tool's use. Open tools show characteristics of the problem domain that provide an observer with a context for understanding what has been done, and the possible implications. For example, a control task carried out by directly manipulating a graphical interface (e.g., mimic representation of a system) may be more useful for an observer (by showing which component is being operated and the effect on the rest of the system) than if the same task was done by other means, such as with text commands via a keyboard.

- *Openness of Interaction* – This refers to the degree to which the interactions between team members provide an opportunity for others with relevant information to contribute. Openness of interaction depends on the type of communication (e.g., discussing actions or decisions in the presence of others) and the style of interaction (e.g., the degree to which unsolicited input is accepted). Openness of interaction is also influenced by characteristics of the work environment (e.g., openness of tools, horizon of observation) that provide other team members with an opportunity to see and hear the interaction. Conventional control room designs typically provide a broad horizon of observation that facilitates the observation of team activities (Stubler and O'Hara, 1996b). Conventional hardwired controls and displays may be "open tools" in the sense that an observer can infer information about actions (e.g., which plant system was involved, which control was operated, what action was taken) by observing the operator's location at a control panel and the action performed. Interactions may be considered "open" when most are verbal ones that can be heard by others in the control room. Practices such as calling out requests for information and calling back the responses allow other control room personnel to maintain awareness of operational activities. The crew's performance may be impaired when upgrades change these HSI features and operational practices, as may occur when the display or interaction features of HSI upgrades interfere with the ability of personnel to share a view of the plant. An example is an individual operator console, which gives individual views of the plant that are not easily seen by others. If other crewmembers cannot view similar information, coordination difficulties may arise. For example, when interacting with the operator, greater opportunities for failures in communication may occur because other crewmembers do not understand the operator's frame of reference. If such individual consoles are used to execute control actions, error detection may be compromised because other crewmembers may not be able to observe those actions while performing their own. Thus, subtle mechanisms for maintaining awareness of control actions and detecting errors may be lost (Stubler and O'Hara, 1996b). A related example is decision aids that provide special information to an individual, but do not make it accessible to the crew. Consequently, the user may experience difficulty in communicating complicated concepts to the rest of the crew because members do not share a common view of plant conditions. The physical characteristics of HSI upgrades also can affect the crew's performance. For example, face-to-face communications may be impaired by factors that interfere with their line-of-sight, or new equipment may requires operators to spend more time away from the main control panel (e.g., monitoring back panels). Another characteristic of the HSI that may impair crew coordination is background noise that can affect the intelligibility of speech.

In summary, nuclear power plant teams coordinate in sharing information and perform their tasks to maintain safe operation, and to restore the plant to a safe state should there be a process disturbance. Important human factors engineering aspects of teamwork include having common, coordinated goals, maintaining shared situation-awareness, engaging in open communication, and cooperative planning. O'Hara and Roth (2005) showed that successful teams monitor each other's activities, back each other up, actively identify errors, and question improper procedures.

NUREG/CR-6208 identified three types of cognitively demanding situations where specific types of crew interaction appeared to contribute positively to successful crew performance from a technical perspective. There were:

- Cases where operators needed to pursue multiple objectives. Specifically, cases where they had to manage dual requirements to (1) proceed through the EOPs to cool down the plant and bring it to a more stable state in a timely manner and (2) engage in extra-procedural activities to handle aspects of the situation that were not covered by the EOPs
- Cases where situation assessment required integration of information that was distributed across crew members
- Cases where crews had to evaluate the appropriateness of a procedure path and/or decide whether to take actions not explicitly specified in the procedures

Similarly, Roth (1997), Mumaw, Roth, Vicente, and Burns (2000), and Vicente, Roth, and Mumaw (2001) performed field studies of nuclear power plant crews, and found that operators often work outside of their narrowly prescribed roles as a way to try to improve their holistic understanding of the situation and improve their overall performance. According to the Mumaw et al. (2000) study, they "found that what makes monitoring difficult is not the need to identify subtle abnormal indications against a quiescent background, but rather the need to identify and pursue relevant findings against a noisy background. Operators devised proactive strategies to make important information more salient or reduce meaningless change, create new information, and off-load some cognitive processing onto the interface." (pg. 36). According to the Vicente, Roth, & Mumaw (2001) study, there are cases where a crewmember made a decision or took an action that was not strictly identified as being within his or her purview, but did so because he or she was attempting to:

1. Enhance information extraction by increasing the salience of important indicators and reducing the background "noise"
2. Create new information
3. Offload some of the cognitive processing onto the interface (e.g., creating external aids and reminders for monitoring).

Carvalho, Vidal, and de Carvalho (2007) and Carvalho, dos Santos, and Vidal (2006) also performed field studies of nuclear power plant crews, and showed that communication, shared cognition, and plant culture can affect crew coordination, and ultimately plant safety. Their analyses showed that communication was key in developing a shared understanding of the situation among the crewmembers (i.e., shared cognition), such that the correct decisions and courses of action could be taken. Perhaps more importantly, they showed that this emergent process is messy. The process is messy in that:

1. The communication process can iterate numerous times
2. It is not always possible to predict how many iterations it will take for a shared understanding to be created
3. Numerous cultural factors, including the leadership style of the senior reactor operator and the plant's safety culture can affect communication and the process of forming a shared understanding, which other research (i.e., Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000) has shown is key to effective Team Coordination.

Toquam, Macaulay, Westra, Fujita, and Murphy (1997) assessed performance differences between different nuclear power plant crews, and noted that there are three primary antecedent factors that contribute to variability in performance between teams. The antecedent factors are task characteristics (e.g., how routine and simple versus unusual and complex is the task the team must perform), team member characteristics (e.g., intelligence, personality types, specific cognitive abilities, etc.), and team dynamics, or team characteristics, such as group cohesion and communication practices. In general, there is a considerable amount of literature that supports the observed relationship between the antecedent factors and variability in team performance. For example, Gertman, Haney, Jenkins, and Blackman (1985) found that the emotional stability of individual crewmembers' personalities was related to their future performance. Itoh, Yoshimura, Ohtsuka, and Matsuda (1990) showed that perceptual speed and memory (i.e., cognitive abilities) are also related to individual nuclear power plant crewmember performance, and as a result, overall team performance. Similarly, many studies show that group cohesion or lack thereof, can have an effect on team performance (for a review, see Evans & Dion, 1991).

Chung, Yoon, and Min (2009) argued that communication errors in nuclear power plant crews are important to understand, and that communication protocols will vary depending on if the crew is operating in a conventional main control room (CMCR) or an advanced main control room (AMCR). To research communication errors and how they can lead to other human errors, they developed a framework to analyze communication among crewmembers of highly complex industrial processes. Their Human-Human-System (HHS) framework allows communication exchanges between CMCR or AMCR crews to be deconstructed and analyzed such that communication errors can be identified and their effects on crewmembers' cognitive processes, what they call abstractions and de-abstractions, can be ascertained. The researchers observed and videotaped the conversations of crews in a full-scale dynamic simulator to test their HHS framework, and found that it was effective at identifying how communication errors among crews are an important contributor to other human errors (e.g., errors in decision making), which further affect overall system performance.

With respect to how the differences between CMCR and AMCR can affect crew communication, the researchers made some astute observations about how the HSI and division of labor between humans and automated systems differ in CMCRs and AMCRs. They note that the CMCR design philosophy tends to follow the single-sensor-single-indicator design philosophy and requires the human operator to do the heavy cognitive processing required to synthesize the single indicator data points the HSI provides them into meaningful information, whereas the AMCR design philosophy uses automated systems programmed with advanced algorithms to do more of the information processing and synthesis of data into meaningful information. This difference changes the role of the operator from data synthesizer and "sensemaker" in the CMCR to more of a supervisor in the AMCR. This change in the roles and responsibilities of the crewmembers can affect not only the nature and content and of their communications, but also the types of communication errors they are likely to commit.

Kim, Park, and Byun (2009) noted that between 2001 and 2006, Korean nuclear power plants experienced 120 reactor trips where human error was attributed as a contributing factor. Wanting to improve overall nuclear power plant performance, they conducted a study using Korean nuclear power plant crews on whether CRM training would improve 1) attitudes, as measured by operators management attitudes questionnaire, 2) individual performance, as measured by workload and situation awareness, and 3) team performance, as measured by coordination demand analysis and behaviorally anchored rating scales. They found that CRM training improved crew attitudes, individual, and crew performance. Compared to the crew that did not receive any CRM training (control group), the crew that received the CRM training had a more positive attitude towards performing their tasks, perceived their tasks as requiring less work (i.e., reduced workload), had greater situation awareness, and behaved more effectively as a cohesive unit.

Park, Jung, and Yang (2012) also studied how characteristics of communication among nuclear power plant crews during simulated emergencies affected crew performance. Specifically, they found that communication characteristics such as: 1) a tightly coupled communication structure (which is an indicator of good team cohesion), 2) increasing the amount/density of communication to increase team situation awareness (e.g., crew members speaking up when observing changes in the system state), and increasing the thoroughness of communication to make shared understanding more explicit (e.g., greater adherence to three-way communication practices) improved overall crew performance during simulated emergency scenarios.

In summary, existing nuclear power plants are highly complex systems that require teams of human operators to be well trained on both the technical aspects of operations as well as the 'soft-skills' of effective teaming. That is, the literature summarized in this section showed that in order for nuclear power plant operators to perform well as a team, they not only need to be technically proficient, they also need to be well trained on how to manage the coordination of both their cognitive processing and physical efforts. Moreover, effective communication was identified as one of the key factors to the effective management of team coordination. New nuclear power plants, including AdvSMRs, are not likely to be any less complex (and could be even more complex), and will still require teams to operate them. The composition of these teams, however, is likely to change from all human teams to human-automation teams. How human-automation teams will effectively manage both their 'cognitive' processes and coordinate their efforts such that they operate these new plants optimally is an important question that needs to be addressed by this study.

3.3.3 Human-Automation Teams

In this section the research on human-automation teams is examined from four perspectives. First the issue involved in modeling human-automation teams is studied. Next, the insights derived from types of automation that are highly interactive and functions much like a human team mate (computer operator support systems) are presented. Third, the effects of the introduction of automation on human-human teamwork and examined. Finally, the research focused on defining principles of human-automation teamwork is described.

3.3.3.1 Models of Human-Automation Teams

As the role of automation expands in new and advanced systems, one goal of research is to make automation a team player. One motivation behind this effort is the finding that humans relate to automation in similar ways to the way they relate to human teammates. This approach uses human-human models of teamwork to look at human-automation teams. However, automation agents cannot behave and interact with humans in the same way other humans can. Thus, an understanding of the differences between human and automation agents is important for designing automation to be team players subject to their differences and limitations. Where agent capabilities can mimic those of a human teammate, principles for doing so can be developed. For those teammate characteristics that agents cannot mimic, principles for designing alternative approaches for accomplishing the characteristic can be developed.

Steinberg (2012) identified two significant limitations in multi-agent, or human-automation teamwork. First, automation cannot fully capture the human operator's intention in performing tasks. Second, automation cannot flexibly adapt to situations that have not been considered by the designer. Similarly, human operators often do not have a good understanding of how automation functions and find its interactions disruptive. Thus, Steinberg concludes that it may not be realistic, in the near-term at least, to think automation can be a teammate in the sense that another human operator is. Steinberg suggests that to create more effective multi-agent teams, researchers need to think more broadly about how human-machine relationships. A means of doing this is to look at different alternative teaming models. For example, while humans and dogs work together in teams, there is no expectation that the dog will exhibit all of the social and cognitive skills as a human teammate would. Similarly, adults and children can work together as a team. However, adults do not expect that a child will function like another adult. Steinberg's point is that there are alternative models of teamwork based on teammates that have different capabilities.

Klein et al. (2004) made a similar observation:

> We can imagine in the future that some agents will be able to enter into some form of a Basic Compact, with diminished capability. Agents might eventually be fellow team members with humans in the way a young child or a novice can be—subject to the consequences of brittle and literal minded

> interpretation of language and events, limited ability to appreciate or even attend effectively to key aspects of the interaction, poor anticipation, and insensitivity to nuance. (p. 94).

Fiore et al. (2005) also questioned the use of human-human models of teamwork for designing human-automation teams:

> Because the execution of many complex tasks is increasingly relying on human-agent teams (e.g., law enforcement, emergency response, military operations), it is critical that we understand the cognitive and coordinative processes arising from such interaction and the specific conditions affecting them. We need to fully understand the degree to which theory and methods applicable to human teams are appropriate to human-agent team environments. Although the utility of human-agent teams seems promising, such systems have been applied at a rate that vastly outpaces the research necessary to fully capitalize on their strengths while limiting their weaknesses.

Sycara and colleagues (Sycara & Lewis, 2004; Sycara & Sukthankar, 2004) discussed integrating agents and human in teams. They viewed teamwork from the human teaming perspective of Sycara & Lewis (2004) which identified the four critical dimensions of: information exchange, communication, supporting behavior, and team initiative/leadership. In addition to these teamwork skills, agents must also have domain knowledge.

Sycara & Lewis, 2004 stated that the human-automation interaction requirements differ based on the role of the automation in the team. The identified three different types of contributions automation can make to teamwork

- To work independently on tasks and in that capacity function like a human teammate
- To collaborate with human team members and support their task performance
- To support teamwork itself such as facilitating communication and coordination of human team members (see Lenox et al., 1998, below)

Sycara et al. noted that the greatest obstacle to integrating machine agents with human teams is communicating a human's intent. Further, if automation is made more flexible and autonomous, it becomes more difficult for humans to monitor it and evaluate its performance.

Sycara & Sukthankar (2004) identified three important factors in human-automation interaction:

- Mutual predictability of teammates
- Team knowledge (shared understanding)
- Ability to redirect and adapt to one another

Agent predictability and shared understanding is made more difficult because automation often does not communicate its intent (this is consistent with Steinberg, 2012) They suggest that predictability can be fostered by:

- Consistently pairing simple observable actions with inputs
- Making the causes and rules governing an agent's behavior accessible to the human
- Making the purpose, capability, and reliability of the agent known to the human

Directability and mutual adaptation are important aspects of teamwork and enable teams to be flexible in different task contexts. Sycara and Sukthankar define directability as "assigning roles and responsibilities to different team members" and mutual adaptation as "how team members alter their roles to fulfill the requirements of the team." Researchers acknowledge that the most effective agents will need to change their level of initiative, or exhibit adjustable autonomy, in response to the situation to be most effective. For agents to appropriately exercise initiative, they must have a clear model of the team's goals, member roles and team procedures.

Related to directability and mutual adaptation is power distribution among agents. Power distribution is a critical characteristic of the team setting with some research suggesting that introducing automation to a team setting significantly affects the human team member dynamic (Wiener, 1989). Introducing automation into the cockpit changed the human-human interaction with increased attentional demands placed on the crew. Distinctions between flying and monitoring duties were ambiguous as Bowers, Salas, and Jentsch (2006) pointed out, "The 'flying' pilot's primary duty is to monitor the automation, but what does the monitoring pilot (non-flying) keep track of: the flying pilot or the automation?" Furthermore, the power distribution is not limited to the humans, with automation taking on both leader and subordinate roles depending on the given task. There are a number of options. First, the automation may assume the leadership role, which may occur with the blessing of the human team members. Next, the automation may relinquish power to the human team members. Finally, the automation may bypass the human team members and relinquish power to additional automated team members. In addition, there are certain automated systems that maintain complete control, with neither direct control or override capabilities granted to the human. As these researchers point out, a lack of clearly defined roles, boundaries and authority significantly affects the power distribution within the team.

Communication between agents is key to achieving mutual predictability, shared understanding, and an ability to redirect and adapt. Communication structure between automated and human team members, however, is significantly different than standard human-human communication. While human-to-human communication includes face-to-face communication and nonverbal cues, automation is limited to transmitting and receiving information via specific formats within an array of displays (Bowers, Salas, & Jentsch, 2006). Suchman (1990) attributed problems with modern automated system to the limits of syntax and semantics in automation. The automated system is at the mercy of the programming, specifically, the certain types of information the system can provide and request. These limits then extend to the human teammate, who must anticipate how to acquire and format the information received. This sharing of information is critically different than the "spontaneous, free-flowing nature of human-human communication" (Bowers, Salas, & Jentsch, 2006). Good communication, therefore, is a cornerstone of achieving effective human-automation interaction.

Lenox et al. (1998) investigated the use of automation to support teamwork (the third function discussed by Sycara & Lewis, 2004 above). Using the teamwork dimensions identified previously, the authors identified opportunities for agents to support teamwork (see Table 2). These teamwork skills are independent of individual competencies, yet the performance of teams, especially in tightly coupled tasks, is highly dependent on them.

Table 2. How Agents Can Support Human Teamwork.

Teamwork Dimension	*Ways to Provide Support*
Situation assessment	▫ act as an information provider ▫ ensure the exchange of relevant knowledge among all members ▫ provide a shared representation of the situation for common knowledge ▫ notify people when all available information sources are not being used ▫ notify people when information is currently disable or unreliable
Supporting behaviors	▫ make other's actions and decisions visible to detect potential errors ▫ alert members to potential errors, constraints and conflicts ▫ support self-reflection of the team for self-correcting teams
Communication	▫ alert members to ambiguity ▫ support translation of terminology among various subgroups ▫ serve as a repository for messages between team members
Team leadership/initiative	▫ communicate intent behind commands to allow lower levels of command to meet intent under evolving conditions ▫ communicate priorities and notify members when priorities change

Johnson et al. (2012) emphasize that "increased autonomy can eventually lead to degraded performance when the conditions that enable effective management of interdependence among team members are neglected." Johnson et al. (2012) conducted an experiment predicting that high levels of automation would not be associated with the highest team performance level. Participants performed a simple block retrieval and delivery task with the help of an agent teammate. The agent had four levels of autonomy. At the lower levels, the participant had to instruct the agent in most aspects of the task. At the higher level, it performed with minimal instructions from its human teammate. On task performance measures, time to complete the task becomes shorter as automation increased to Level 3, but rose with level four, consistent with their hypothesis. Errors were also highest in the high automation condition. Also, with increasing autonomy, workload went down and ratings of "opacity" went up. Participants also were asked in which condition they felt their performance was best and which they preferred. On both measures, Level 3 had the highest ratings.

Thus, task requirements, such as how much interaction in needed to jointly perform a task, impacts the type of team processes that are desirable.

Woods et al. (2004) also noted that designers often fail to appreciate the need for humans and automaton to interact even when automation is designed to be mostly autonomous. One reason for this is that designers often underestimate the complexities of operational environments. Further, when problems are encountered, there is a tendency to treat as an isolated glitch that needs to be addressed, rather than as an indicator of broader issues. This led Woods and his colleagues to postulate a variant on Murphy's Law:

> … any deployment of robotic systems will fall short of the target level of autonomy, creating or exacerbating a shortfall in mechanisms for coordination with human problem holders (p. 210)

While the expectation is that as autonomy goes up, the need for human interaction goes down, actually the demands for more sophisticated forms of communication go up.

Characteristics that lead to these demands, even in highly automated systems, are automation's brittleness and literal mindedness, which limits its ability to adapt to changing situations.

Rather than human-human teamwork models, Fiore et al. (2005) suggested the focus should be on generic and specific team competencies. Team competencies are knowledge, skills, and attitudes that foster effective team interaction behaviors and performance. Some of these competencies are generic regardless of the specific situation, (e.g., communication and leadership). Other competencies are team-specific and are dependent on the team's current situation and its task requirements, (e.g., knowledge of other team members' abilities and team cohesion).

Fiore et al. state that human-agent team interaction introduces a number of issues with respect to both team and individual cognition such as humans having to deal with an increased level of abstraction that may place unique demands on their information processing and difficulties coordinating team members given that differing communication patterns may be necessary to share cues.

Fan & Yen (2011) made a similar observation. They note that human teams rely on global situation awareness and shared mental models. When automation agents are involved, developing this awareness and shared mental models can create additional costs centered on communication, resolution of conflict, and social acceptance.

Pritchett's and colleagues identified ways in which automation agents are different from human agents (Feigh & Pritchett; in press; Kim, 2011; Pritchett, Kim, & Feigh, in press a & b).

- *Behavior when outside boundary conditions* – human team members will continue to attempt effective performance in unfamiliar circumstances, while automation generally cannot.
- *Anticipating the needs of a teammate* – Good teammates anticipate each other's information needs and provide information. Automation is limited in this capability.
- *Managing interruptions* – Humans time their interactions based on the current situation, e.g., another teammate workload. Automation can be "clumsy" in this regard and interrupt human teammates at inopportune times.

- *Responsibility* – Automation does not have motivation and a sense of responsibility.

The design of automation teammates should address these differences.

In summary, a number of factors have been identified that call a strict human teamwork model into question when designing human-automation teams. Some of these factors are:

- Failure to anticipate a human's intention
- Limited flexibility to handle novel situations
- Poor communication facilities
- Difficulty with mutual predictability
- Difficulty establishing shared cognition such as shared situational and mental models
- Limited ability to redirect and adapt
- Limited interaction with high levels of automation
- Failure to consider generic and specific competencies
- Failure to anticipating the needs of teammates
- Failure to manage interruptions
- No sense of responsibility

As noted earlier, where agent capabilities can mimic those of a human teammate, principles for doing so can be developed. For those teammate characteristics that agents cannot mimic, principles for designing alternative approaches for accomplishing the characteristic can be developed.

3.3.3.2 Insights for Human - Automation Interaction Requirements from Studies of Computerized Operator Support Systems Limitations

A noted in the discussion above, human-automation interaction is needed even when the level of automation is high. When automation is applied to higher-level cognitive functions, such as situation assessment and response planning, the level of interaction required is much higher. This type of automation is frequently referred to as decision support systems (DSSs) or computerized operator support systems (COSSs). The term COSS will be used in this document. By contrast, the requirements of interacting with automation of lower-level cognitive functions, such as monitoring/detection and response implementation, are typically lower.

Wickens and Hollands (2000) found when higher-level cognitive functions are automated; operators are more prone to experience out-of-the-loop unfamiliarity with the ongoing situation. The reasons for this finding related to the design of the COSS and their interaction functions.

Several studies have identified numerous human-automation design issues that limit the effectiveness of current COSSs. O'Hara et al. (1996) identified the following:

- Poor integration with operators' task performance
- Complexity of COSS information-processing
- Lack of transparency of the COSS decision process
- Inadequate explanatory information to address personnel's needs for verification
- Lack of communication features that permit people to query the system or obtain a level of confidence in the conclusions it has drawn.

Some of these issues relate to the operators interaction with automation: Lack of transparency and communication features.

Other studies identified similar issues (IAEA, 1995; Malin et al., 1991a & 1991b; Land et al., 1995; Rook & McDonnell, 1993). These studies point to the importance of operator understanding of the reasoning processes used by COSSs and communication with the COSSs, especially when they do not comprehend its results.

Malin et al. (1992) detailed case studies of the design of 15 intelligent systems intended for a variety of aerospace projects; most were real-time fault-management systems. They deemed as problematic the interfaces between the human operator and the intelligent systems. Some specific concerns Malin et al. identified were similar to the concerns identified above: Providing visibility into the system's reasoning, understanding its reasoning, the system's ability to respond to the context in which a question is asked, and its ability to distinguish hypotheses from facts. Additional problematic areas were determining the credibility and validity of information, handling interruptions and changes in planned activity sequences, distinguishing between modes of operation, gaining control over the system's actions, and identifying its errors. The systems also had many problems associated with the general design of the interfaces with the systems. Operators often did not get the information they needed, or it presented in confusing formats that were unsuited to their task requirements. Other times, excessive detail was provided. This made it difficult for operators to "…visualize the intelligent system's situation assessment and recommendations in relation to the flow of events in the monitored process."

Other studies also identified the importance of visibility of the reasoning process. Rook and McDonnell (1993) experimentally manipulated the interface to an expert system that was designed to support fault diagnosis during simulated space-station problem-solving situations. One version of the HSI clearly depicted the reasoning; in the other, little information about the reasoning process was available. The results demonstrated that the former led to more effective use of the aid.

Similar findings were obtained by Roth, Hanson, Hopkins, Mancuso, and Zacharias (2004) in evaluating the extent to which operators of unmanned aerial vehicles (UAVs) understood the plans developed by automated controllers. Humans and automatic agents cooperated as a team in planning and executing missions. The authors' findings supported the need to communicate more effectively the rationale behind the automation's plan elements, and to provide "levers" enabling the operators to modify the plan. Without such information, operators could not properly assess the plan's appropriateness, and assess whether changes to it were needed.

Dien and Montmayeul (1995) surveyed operating experience with COSSs in the nuclear industry. They concluded that few improvements in operations were achieved with their use. COSSs often offer guidance for situations that operators already are equipped to handle. That is, they are designed for situations that previously were analyzed, and designers and operators are familiar with. Such aids add little value for operators, except perhaps, to confirm their decisions. Further, the systems were unable to effectively manage unforeseen circumstances and, therefore, did not provide appropriate guidance to operators during them. Another problem was that aids were "acontextual", i.e., their guidance had little reference to the current situation. COSS guidance was given without appropriately communicating what led to its issuance, what parameters were analyzed, and what sequence of reasoning was followed. When the reasoning process was provided, it often conflicted with that of the operators, (i.e., It was seemingly based on the designer's theoretical understanding and not on the operator's practical experience). Dien and Montmayeul (1995) also found that poor integration with the existing HSI systems is problematic, (e.g., differences between systems in dialog principles, such as command vs. direct manipulation interfaces, and coding, such as the meaning of color usage). Poor integration with existing practices has led operators to abandon COSSs (Reed, Hogg, & Hallbert, 1995).

Consistent with these studies, Roth, Bennet, and Woods (1987) concluded that the interface of the COSS must support cooperative dialogue so the operator can better understand and utilize the system. In general, they found that these aids tend to be technology-driven, and do not address the operator's needs; the systems are developed by finding an application for a given technology, rather than being designed to meet users' needs. Proper design requires an adequate analysis of user's requirements to identify information needs. The system should accurately provide the needed information, and minimize or eliminate extraneous material. Roth et al. (1987) stated that COSSs should support the operator's cognitive processes and reinforce their existing approach to operations that was formulated through training and experience. Operators should be clearly informed of the limitations of COSSs (Bernard & Washio, 1989a, 1989b; Terry, 1989).

The design issues discussed above may lead to a fundamental issue underpinning human-automation interaction – trust. Madhavan and Wiegmann (2007) analyzed operator's perceptions of COSSs. They concluded that humans enter into relationships with automation in a manner very similar to the way they enter into relationships with other humans. A key factor is trust. Trust begins with a basis in performance or reliability, progresses to the level of dependability and finally evolves to faith. Trust in automation begins with faith in the initial stages, which is followed by dependability, and then predictability COSSs should be designed to:

- Mimic human interactions, including the use of human language structures if possible
- Possess unique knowledge and functional algorithms that may be inaccessible to the human teammate, thus carrying out tasks which humans cannot do alone

These studies provide insights concerning the general design of COSSs. The researchers divided the issues into three classes: structural design (how COSSs accomplish the tasks), human-automation interaction (how COSSs interact with the operators they are supporting), and moderating factors. The results are summarized in Table 3.

Table 3. COSS Design Issues Limiting Their Effectiveness.

Structural Design	Human-Automation Interaction	Moderating Factors
Complexity of COSS info processing (intelligibility)	Lack of transparency (visibility of the reasoning process)	Poor integration with task demands (lack of context)
Identifying modes	Poor Communication features	Trust
	Interruptions	Confidence
	Assessing or determining the credibility of COSS results	Integration with general HSI design (including the interaction/dialog approaches)
	Providing input and gaining control	
	Control over level of detail	

In summary, the human-automation interaction issues identify design requirements for human-automation teaming. The human-automation interaction should be designed so:

- The way information is processed is accessible to the operator
- The design of communication functions and features enable operators to obtain the information they need to use automation information, such as assessing the credibility of the results
- Interruptions are minimized
- The level of detail can be controlled
- Operators can provide input to the processing and direct its activities

3.3.4 Effects of Automation and Automated Aids on Human Teamwork

Automated aids also affect team performance. The design of human-automation teams needs to account for these effects.

Roth and O'Hara (O'Hara & Roth, 2005; Roth & O'Hara, 2002) observed the introduction of a computer-based emergency operating procedures (EOPs) system as part of one utility's digital instrumentation and control (I&C) upgrade of a nuclear power plant. This system automated the functions of information acquisition and analysis and gave some support to decision-making. The crews handled disturbances on a training simulator. Following each scenario, our interviews focused on the impact of the procedures on operations.

Before this upgrade, the usage of EOPs involved the entire control-room crew of three operators. The supervisor read the procedure and decided at each step how to proceed. The board operators retrieved the needed data for each step and told the supervisor about it. At the supervisor's instruction, they took any required actions. This work demanded team communication and coordination. The computer-based EOP automated many of the tasks that the crew performed with paper procedures, including

- Retrieving data and assessing its quality
- Resolving step logic
- Tracking the location in the procedure
- Tracking the steps to assure continuous applicability
- Assessing cautions, safety-function status trees, and foldout page criteria.

Thus, the system automated many of the crew's information acquisition, analysis, and decision-making activities.

By introducing the new system, the procedure management workload was reduced to the point that procedure use became a one-person activity. The board operators were far less engaged in this, except to take occasional control actions at the request of the supervisor. Consequently, the operators felt they were out-of-the-loop, had lost situation awareness of EOP activities, and were unsure what to do. Our main point was that the introduction of the automation impacted teamwork, a finding that was unanticipated by the designers and plant staff. This was not a negative finding per se. The situation was addressed by the plant's operations and training staff who redefined the roles and responsibilities of the individual crew members, and took steps to foster teamwork. First, operators were told to manage alarms and check key parameters on their side of the plant (reactor and balance-of-plant). Then, specific stop points were added into the procedure where the supervisor updated the operators on the procedure's status and the crew shared their assessments and informed the supervisor of key findings. Because the shift supervisor and the crew worked more independently than before, and attended to separate sources of information, the stop points gave them the opportunity to keep each other informed and to ensure a common understanding of the event. EOP training reinforced this new approach to emergency management.

Kaber and Wright (Kaber et al., 2001; Wright, 2002; Wright & Kaber, 2005) examined the effects of automation on team performance and teamwork. The objective of this research was to evaluate the effects of automation as applied to these different stages of information processing on the performance and coordination of teams in a complex decision making task. Two-person crews performed a simulated mission to protect a home base from enemy attack. There were two task difficulty levels. Automation was provided and accomplished information acquisition, information analysis, or decision selection. Performance measures included team effectiveness, quantity of team communication, team coordination ratings by outside observers, and task and team workload ratings.

The results showed that the effects on teamwork differed based on the generic tasks that are automated and are summarized in Table 4. The findings suggested that automation of early and intermediate stages of processing may have benefits with respect to teamwork, while automation of decision selection may be more limited in the contexts in which it will be beneficial.

Table 4. Summary of the Effects on Task Automation on Teamwork (adapted from Wright, 2002).

Generic Task	Finding
Monitoring and Detection	Decreased the quantity of communications
	Increased the anticipation ratio
	Led to the selection of a better task strategy
Situation assessment	Increased ratings of team coordination
	Increased the quantity of communications related to situation assessment, decision-making, and leadership
	Led to the selection of a better task strategy
	Increased workload under low difficulty
Response planning	Decreased ratings of team coordination
	Increased performance under low difficulty
	Increased workload

In summary, an important consideration of human-automation interaction is the effects of the introduction of automation on human-human teamwork. The design of human-automation teams needs to account for these effects.

3.3.5 General Principles of Human-Automation Teamwork

From his extensive study of automation in aviation systems, Billings (1991, 1997a, 1997b) identified general principles for supporting personnel interaction with automated systems:

- Humans are responsible for outcomes in human-machine systems; thus, humans must be in command.
- Humans must be actively involved in the processes undertaken by these systems, and where appropriate, be given options on decisions and actions. Humans are needed for their flexibility, creativity, and knowledge. They are able to cope with incomplete knowledge and uncertainty.
- Humans must be adequately informed of human-machine system processes, so they can track what automation is doing.
- Humans must be able to monitor the machine components of the system.
- The activities of the machines must be predictable.
- The machines must be able to monitor the performance of the humans, to alert them to potential errors.
- Each intelligent agent in a human-machine system must know the intent of the other agents.

Billings' principles describe a human-machine system in which humans are in charge and work is accomplished interactively with machine agents. Each agent involved in the process is aware of and monitors the performance of other agents. Billings' view is essentially that automation is one of the team, a fellow crewmember.

Many authors used human teamwork as a model to identify the general characteristics of desired human-automation interaction (Klein et al., 2004 & 2005; Lee & See, 2004; Parasuraman & Riley, 1997; Woods & Hollnagel, 2006). In fact, in their study of human trust in automation, Lee and See (2004) noted, "…many studies show that humans respond socially to technology, and reactions to computers can be similar to reaction to human collaborators" (p. 51). Thus, designing automation to have the characteristics of a good, trusted team member is viewed as a pathway to overcoming the potentially negative effects on human performance of poorly designed automation.

In human teams, there must be shared situation awareness (SA) to support the coordination of the team's goal-directed activity. An important aspect of shared SA is the extent to which team members have a common understanding of what is needed to support team performance, and knowledge of the responsibilities of each individual member, the status of their activities, and an expectation of the actions they will take in the future. Multi-agent teams are no exception, and a shared SA still is needed (Kaber, Riley, Tan & Endsley, 2001). Inherent in designing automation are assumptions about the roles and responsibilities of the human operators with whom it will interact. Likewise, operators need to clearly understand the automation's role, be able to anticipate its contributions, and what it will do in the future.

Supporting team SA is the degree to which operators perform their tasks together in an "open" work environment (Hutchins 1990). Typically, they are aware of others' actions and infer their intentions, often without explicit communication. By contrast, automated systems tend to be "opaque" (Billings, 1997a) and lack the means of taking into account the operator's actions.

While designing automation to be a team player may be appealing from a design perspective, it is challenging. An example comes from a study of extended teamwork that included nuclear-plant operators and automatic agents (Skjerve, Nihlwing & Nystad, 2008). The operators found it "highly" useful when an automatic agent provided suggestions for handling a disturbance. However, in those situations where the operator's workload was high, operators did not like the fact that the agent continued giving suggestions, and did not adjust its behavior in response to the current situation. The authors quoted one operator as stating

> When the turbine operator looks at me, in a situation where I am busy, the turbine operator immediately observes that I have a lot to do, and he will start to take it easier…such as, delay the performance of certain tasks until I tell him to perform them… And I can say to him: OK, now you can start to open this valve to 10% - slowly. And he will be careful. He might start opening the valve 1% and see what happens, and if things are OK, he will open it to 2% and so on. The (automation) agent will open the valve to 10% immediately (p. 23).

Nevertheless, the potential benefits of making automation a better team player are widely acknowledged.

Within a teamwork model of automation, the first aspect to address is the issue of operator trust in automation. Parasuraman and Riley (1997) offered some general guidance: Fostering an understanding of automation; minimizing overreliance on automation; and establishing confidence in automation. Each is summarized below.

Fostering an understanding of automation - When operators understand the purpose and the functioning of the automation, they can use it more effectively. Therefore, the conditions for using the automation, and conditions under which it should not be used, should be explicitly included in procedures and operator training. The decision to use automation should not be influenced by the effort involved in managing it; automation should not be difficult or time-consuming to engage.

Minimizing over-reliance on automation - Operators may rely on automation excessively unless they encounter factors favoring overreliance. For example, when monitoring the performance of automation, common behavioral biases make operators unlikely to detect an automated process going wrong, especially when they are busy. One way to counter this tendency is to reduce workload, especially that involved in monitoring. Feedback about the automation's states, actions, and intentions can be presented such that they direct the operators' attention appropriately without imposing an undue burden. In making decisions under uncertainty, operators may overvalue the data provided by the automation and fail to seek out independent information; again, training can help operators to recognize and counter decision biases that may lead to overreliance.

Establishing confidence in automation - Operators will not use automation if they lack confidence in it. Ensuring sufficient reliability is the automatic system will promote confidence.

Klein et al. (2004, 2005) proposed six requirements for automation to be a team player:

- Basic compact
- Predictability
- Goal negotiation

- Coordination phases
- Attention management
- Controlling costs of coordination activity

Basic compact refers to engaging in a joint activity where the activities and status are communicated to operators and other agents; i.e., maintaining common ground. This includes having the automated agent alert personnel in the event it can no longer perform any of its functions.

Predictability refers to the predictability of the actions of human and machine agents. To participate in joint activity, automation should be predictable; and, therefore, must signal its status and "intentions," including its goals, state of knowledge, and impending actions. To the extent possible, automation should be able to interpret the status and intentions of human operators. Klein noted the basic asymmetry between humans and machine agents in their abilities to interpret the activities of other agents. Thus, people can infer the goals of other operators by observing their actions, and can detect whether another operator is having difficulty by observing their behavior. Developing corresponding capabilities for machine agents is problematic; one approach may be inferring the operator's intent from the manipulations made on plant systems and equipment, such as starting and stopping pumps. Another approach is using machine sensing of physiological indicators to infer the operator's cognitive state of operators. However, broad application of such technology may be years away. The authors cautioned that such machine-initiated adaptability to the inferred state of the human participants in the activity could make automation less predictable or cause the human agents to alter their behavior.

Goal negotiation requires that automation convey its goals to the operator and allow operators easily to revise them.

Coordination phases refer to the collaboration between operators and automation and involve handoffs from one to the other. As the automation becomes more capable, it is necessary to exchange more information to maintain common ground.

Attention management suggests that automation must strike a balance between working silently (leading to the well-documented automation-related problems), or burdening operators by requiring them to frequently pay attention to it.

Controlling costs of coordination activity refers to the fact that time and effort is necessary to maintain common ground among operators and automation. The design should account for coordination activities, aiming to adequately support collaborative activity while controlling its costs (e.g., fewer interruptions of ongoing tasks).

Woods and colleagues identified some of the generic characteristics that support teamwork involving machine agents (Christoffersen & Woods, 2002; Dekker & Woods, 2002; Ranson & Woods, 1996b; Woods, 2005; Woods & Hollnagel, 2006; Woods & Sarter, 2000). Below is a list of the characteristics including examples of each.

Coordination - the ability to coordinate and synchronize activity across agents, e.g.,

- Share frame of reference or common ground
- Synchronize activities
- Ensure the predictability of all agents

Resilience - the ability to anticipate and adapt to potential for surprise and error, e.g.,

- Availability of failure-sensitive strategies
- Explore outside current boundaries, set, focus, or priority
- Overcome automation's brittleness
- Revise focus

Observability - feedback giving insight into a process, e.g.,

- Highlight current changes and events
- Show how and why the automation arrived at its current state
- Provide a historical context
- Indicate data sources, processing, synthesis, and effects
- Delineate the relations between the current process state, the control means, and the target state
- Show what the automation will do next and why, including conditional logic to support attention direction, predictions, planning, and coordination
- Signal when the automation is progressing towards its limits and having trouble
- Support the anticipation of failure, and show the reasons for automation failures that have occurred or are about to occur
- Disambiguate automation effects and process effects, showing how automation contributes to process evolution
- Reveal contributions of multiple players and illustrate side effects of each agent's actions

Directability - ability to direct/re-direct resources, activities, and priorities as situations change and escalate, e.g.,

- Assign specific sub-activities to the automation (or permit the automation to do entire tasks)
- Specify the strategies to be used (or permit the automation to suggest or select one)
- Take detailed control over an activity (or make only high level inputs as needed)

Directing Attention - ability to reorient focus in a changing world

- Track the focus of attention of other agents
- Judge the interruptability of others

Shifting Perspectives - contrasting points of view, e.g.:

- Suggest other possibilities as activity progresses
- Point out alternatives as activities approach completion

Based on a review of studies of multi-agent teamwork, including those discussed above, O'Hara and Higgins (2010) derived several general principles for human-automation interaction. These principles are presented in Table 5.

Table 5. General Principles for Supporting Teamwork with Machine Agents (from O'Hara & Higgins, 2010).

Principle	Definition
Define the purpose of automation	Automation should have a clear purpose, meet an operational need, be well integrated into overall work practices, and be sufficiently flexible to handle anticipated situational variations and adapt to changing personnel needs.
Establish locus of authority	In general, personnel should be in charge of the automation, be able to redirect, be able to stop it, and assume control, if necessary. This does not preclude the automation from initiating actions. Some actions are allocated to automation because they cannot be reliably performed by personnel within time- or performance-requirements. There may be situations where automation initiates a critical action because personnel have failed to do so. Such automatically initiated actions, e.g., SCRAM and ECCS, are needed to support the safety of personnel and equipment.
Optimize the performance of human-machine team	The allocation of responsibilities between humans and machine agents should seek to optimize overall integrated team-performance. This may involve defining various levels of automation, each with clear-cut, specific responsibilities for all agents and each with a clear rationale. It also may involve flexible allocations that change in response to situational demands. Personnel's interactions with automation should support their development of a good understanding of the automation, and the maintenance of their personal skills needed to perform tasks if automation fails. This optimization may involve exposing personnel to various levels of automation. The HSIs should support a clear mutual understanding of the roles and responsibilities for both human and machine agents.

Principle	Definition
Understand the automation	Personnel should clearly understand the automation's abilities, limitations, and goals, and be able to predict its actions within various contexts. Minimizing automation's complexity will support this objective. While operators' understanding largely will come from training and experience, the HSI should support that understanding by reinforcing the operators' appropriate mental model through the information provided in automation displays. That is, the HSI should accurately representation of how the automation functions overall, and how it interacts with the plant functions, systems, and components.
Trust the automation	Personnel should have a well-calibrated trust in automation that involves knowing the situations when the automation can be relied on, those which require increased oversight by personnel, and those which are not appropriate for automation. The HSIs should support the calibration of trust, such as providing information about the automation's reliability in its various contexts of use and specific functions.
Maintain situation awareness	The HSIs to automation should provide sufficient information for personnel to monitor and maintain awareness of automation's goals, current status, progress, processes (logic/algorithms, reasoning bases), difficulties, and the responsibilities of all agents. Special attention should be given to changing LOAs and for AA where the roles and responsibilities of all agents may alter. HSIs should support differing levels of need for information, from determining the overall status at a glance to more detailed data in support of greater interaction with automation.
Support interaction and control	Personnel interaction with automation should support the human's supervisory role: • HSIs should support personnel interaction with automation at a level commensurate with the automation's characterization, e.g., level, function, flexibility, and its reliability. • Communication functions should enable personnel to access additional information about its processes beyond that provided in monitoring displays. Automation should communicate with personnel when necessary, such as when it encounters an obstacle to meeting a goal, or when information is needed from personnel (e.g., information not accessible to automation). Communications from automation should be graded for importance, so as not to be overly intrusive. • Personnel should be able to redirect automation to achieve operational goals. Then they should be able to override automation and assume manual control of all or part of the system.
Minimize workload from secondary tasks	A minimal workload should be entailed in dealing with the automation's configuration, and in monitoring, communicating, changing allocations, and directing it.
Manage failures	Automatic systems should support error tolerance and manage failures: • Personnel should monitor the activities of automation to detect automation errors, and be adequately informed and knowledgeable to assume control if automation fails. • Automation displays should support operators in determining the locus of failures as being either the automation, or the systems with which the automation interfaces. • To the extent possible, automation should monitor personnel activities to minimize human error by informing personnel of potential error-likely situations. • Automation should degrade safely and straightforwardly when situations change sufficiently to render its performance unreliable, and should communicate this to personnel in a timely way to enable them to become more engaged in the responsibilities of the automation.

3.4 Conclusions

Based on our review of the general research on what constitutes good human teams, the researchers derived a set of general requirements or principles:

- Belief in the Concept of Team
- Effective Communication
- Team Leadership
- Monitoring Individual and Group Performance and Providing Feedback

- Coordination and Assistance
- Awareness of Internal and External Performance Shaping Factors That Affect Team Processes
- Awareness that each individual's mental model is unique and that it is difficult to create shared mental models in teams.

The researchers focused on research identifying the characteristics of commercial nuclear power plant teams and found that while their operational context provides some unique challenges, the characteristics for effective teaming, and the training operators are provided, were consistent with the general principles identified for teams in general. In nuclear plant crews, technical skills and communications were emphasized as key factors necessary to help resolve discrepancies in individual mental models and to coordinate work.

The research team anticipate that teams for new nuclear power plants, including AdvSMRs, are likely to require the same principles. However, the composition of these teams will likely change from all human teams to human-automation teams. How human-automation teams will effectively manage both their 'cognitive' processes and coordinate their efforts such that they operate these new plants optimally is an important question that needs to be addressed..

To anticipate the new demands posed by the introduction of automation agents to nuclear plant teams, the researchers examined the research on human-automation teams. The researchers conclude that this literature attempts to consider the suitability of using human team models as a basis for designing human-automation teams. However, the research results suggested several automation agent characteristics that call into question the suitability of a strict human teamwork model. These characteristics include:

- Failure to anticipate a human's intention
- Limited flexibility to handle novel situations
- Poor communication facilities
- Difficulty with mutual predictability
- Difficulty establishing shared cognition such as shared situational and mental models
- Limited ability to redirect and adapt
- Limited interaction with high levels of automation
- Failure to consider generic and specific competencies
- Failure to anticipating the needs of teammates
- Failure to manage interruptions
- No sense of responsibility

Where automation agent capabilities can mimic those of a human teammate, principles for doing so can be developed. For those teammate characteristics that agents cannot mimic, principles for designing alternative approaches for accomplishing the characteristic need to be developed if human-automated teams are going to perform well.

Perhaps the demands on human-automation interaction are greatest when the automation agent is supporting higher-level cognitive process, such as situation assessment and response planning. As mentioned earlier, in the nuclear industry, such agents are generally called COSSs. Research on COSSs has suggested the following requirement on designing operator-COSS interactions:

- The way information is processed is accessible to the operator
- The design of communication functions and features enable operators to obtain the information they need to use automation information, such as assessing the credibility of the results
- Interruptions are minimized
- The level of detail can be controlled
- Operators can provide input to the processing and direct its activities

Prior researchers have also suggested many general principles of human-automation teamwork. The principles presented in Table 5 provide a good summary of this work. While these principles provide a reasonable starting place for developing guidance for human automation collaboration, they do not comprehensively address all the human team characteristics and issues discussed in Sections 4.1 and 4.2, nor the challenges raised by human-automation teams, discussed in Sections 4.3 and 4.4. This is a gap that can be bridged with further research.

4. THE PERFORMANCE MEASURES ANALYTICAL STUDY

4.1 Introduction and Objective

The HAC research effort is focused on producing guidance on how to design the automation and human-automation collaboration in new AdvSMR plants to facilitate optimal operator and plant performance in a highly automated system. To accomplish this goal, it is necessary to be able to measure performance of the system, operator, plant, and the human-automation interaction in this specific context. As discussed in Oxstrand et al (2013a), there is a need for a more complete set of performance measures that are applicable in this domain and focus on the relationship between humans and automation, including operator awareness, trust in the automation, and use of automation, as well as measures that reflect multi-agent teamwork. Previous efforts to organize performance measurement suites (e.g., Pina, Donmez & Cummings, 2008) have been focused on identifying a set of metrics for specific domains, such as military command and control, NASA missions, or air traffic control. With a handful of exceptions (discussed in Section 4.3 below), most of these efforts were not specifically focused on nuclear power plants control rooms or the unique operating circumstances of AdvSMR control rooms. Furthermore, these efforts to identify performance measures often had different goals and may have been less systematic than what would be appropriate for AdvSMR HAC research. Additionally, existing measures of individual, system, and team performance may not adequately translate to the highly automated systems that are expected in AdvSMR designs. For these reasons, it is necessary to identify performance measures that are appropriate for AdvSMR HAC research.

The objective of this study is to develop a standardized performance measurement suite for measuring human-automation collaboration. This performance measurement suite will be used in the research experiments conducted as part of the HAC project and for the testing, evaluation, and validation of HAC designs.

The selection of appropriate performance measures is vital to establishing a technical basis upon which findings can be generalized and used to develop guidance for designers. Selection of poor or insensitive measures can result in misleading findings and the possibility of missing important HAC relationships altogether.

The availability of a standardized HAC performance measurement suite helps to ensure our studies evaluate all important aspects of performance, using the best measures available. The use of a standard suite of metrics also supports the comparison of results across studies and, therefore, generalization of findings. A standardized suite can also help ensure that performance measures used are appropriate for high-automated systems being operated in a commercial nuclear power and AdvSMR domain.

After use of the suite in project studies, the suite of metrics can be developed as part of an assessment framework for vendor use in evaluating their unique HAC designs.

4.2 Method and Technical Approach

To accomplish the objectives of this research (i.e., to create a recommended standard suite of metrics for use in HAC research and design), the project team conducted a review of methodological and experimental research on performance measures and metrics, including the automation, nuclear power plants, and human factors domains, as well as any other research that involves automation or human performance in complex, technical systems. The project team focused on identifying generally recommended approaches to performance assessment in human-in-the-loop and HAC research studies. This literature review was conducted with the following goals:

1. To identify performance variables measured in relevant psychological, human factors, and HAC research studies

2. To identify specific performance metrics utilized in relevant psychological, human factors, and HAC research studies
3. To evaluate the identified variables and metrics for applicability to AdvSMR HAC research and design based on a standard set of criteria (e.g., validity, reliability, sensitivity, practicality, context-dependence, etc.)
4. To evaluate whether existing metrics adequately assess known HAC failure modes and performance issues
5. To identify any variables where appropriate metrics do not exist or existing metrics are not sufficient or applicable to AdvSMR research
6. To develop new metrics for important HAC performance variables for which there are no existing metrics (or the existing metrics are inadequate)
7. To create a recommended standard suite of metrics for use in HAC research, and to provide information about the qualities of and usage guidance for each metric.

This report documents work conducted to date and outlines work to be continued at a later stage of the project. As of this report, significant progress has been made on items 1-3 above and partial progress on item 8.

4.3 Identification of Variables Important for HAC Research

In an earlier milestone report on the HAC project, Oxstrand et al. (2013) conducted a detailed literature review of the research related to automation and human performance. Based on this review, the project team identified three general categories of variables of interest for HAC research and design: measurements of human performance, measures of system performance, and measures of HAC performance.

Human performance variables that have been repeatedly demonstrated to be of interest include situation awareness, or the operator's understanding of what is happening in the system (Endsley, 2000; Endsley, 1996, 1997; Endsley & Kaber, 1999; Endsley & Kiris, 1995; Jou et al., 2009; Kaber & Endsley, 2004; Lin et al., 2009, 2010a, 2010b; van de Merwe et al., 2012), workload, or the amount of physical or cognitive tasks that an operator must accomplish.

In addition to measuring human performance, it is important to measure how well the entire human-system performs. The research team refers to this as system performance.

Finally, there are some variables that are particularly important to identify the quality of human-automation interaction in highly automated systems. The additional HAC performance variables are trust and interface management.

4.4 Criteria for Evaluating Metrics

Each of the metrics was evaluated according to standard measures for assessing metric quality, including validity, reliability, sensitivity, diagnosticity, and intrusiveness. Additionally, the metrics were evaluated based on other criteria related to ease of use and appropriateness for AdvSMR HAC research, including practicality and scenario dependence. Each of the characteristics used in this study is described inTable 6.

Table 6. Criteria for Evaluating Metrics.

Characteristic	Meaning
Construct Validity	The extent to which the measure represents accurately the aspect of performance it is intended to measure.

Reliability	A measure should be repeatable; i.e., same behavior measured in exactly the same way under identical circumstances should yield the same results.
Sensitivity	A measure's range (scale) and its frequency (how often data are collected) should be appropriate to that aspect of performance being assessed.
Diagnosticity	The extent to which the measure distinguishes between different levels of performance, or produce qualitatively different diagnoses. Specifically, the extent to which the results produced by the metric allow one to evaluate the strengths of different research hypotheses.
Objectivity	The extent to which the measure is based on easily observed phenomena. Subjective measures are often based on participant self-report, whereas objective measures are based on actual performance data.
Intrusiveness	A measure should minimally alter the psychological or physical processes that are being investigated.
Practicality	Refers to ease of use in experimental or research settings.
Scenario Dependence	The extent to which the measure is customized to be applicable only to the specific context under investigation. Context-independent measures are applicable regardless of the specific scenario used in the research.
Summary	An overall summary of the metric.
Applicability to AdvSMR domain	An evaluation of the metric's applicability or relevance to AdvSMR HAC research and design.

4.5 Identification and Evaluation of Metrics for Each Variable

This section presents reviews conducted to date on metrics in each of the three performance measure categories: human performance, HAC performance, and system performance.

4.5.1 Human Performance Metrics

4.5.1.1 Metrics for Primary Task Performance

When evaluating human performance, one of the most obvious questions is "how does the human perform on the task he is assigned?" This seemingly straightforward question leads to several metrics for evaluating human performance. Human performance can be evaluated based on accuracy or efficiency, both of which comprise a variety of measurement methods.

Primary Task accuracy

Human performance is often evaluated based on the human operator's accuracy when executing his primary task. How accuracy is defined varies considerably depending on contextual factors including: the specific domain, the specific scenario, and the operator's task.

Table 7. Review of Primary Task Accuracy

Metric Characteristic	Discussion
Validity	The validity of this measure depends on the validity of the analysis that defines performance criteria used to measure accuracy.
Reliability	Once criteria for accuracy have been established the results are typically reliable.
Sensitivity	The sensitivity of this measure is highly-context dependent and is subject to ceiling and floor effects if not designed properly.
Diagnosticity	Accuracy can be highly diagnostic, but it varies depending on the specific context.

Intrusiveness	Measures of accuracy can be collected in simulator logs or automatically in other ways, so this metric has low intrusiveness.
Objectivity	This measure is generally the most objective, however bias can enter in defining the criteria for accuracy.
Practicality	It is usually practical to collect accuracy data, however criteria must be established before a study.
Scenario Dependence	Accuracy is highly scenario dependent.
Summary	Primary task accuracy is an important performance metric; however it is highly scenario dependent.
Applicability to AdvSMR domain	Primary task accuracy is applicable to the AdvSMR and is a recommended metric. Researchers should be careful to ensure that the criteria for accuracy are reviewed to ensure that they represent meaningful performance criteria.

Primary Task Performance assessed by expert opinion or observation

Many researchers have measured individual operator and crew performance by expert observation.

Table 8. Review of Primary Task Performance assessed by expert opinion or observation

Metric Characteristic	Discussion
Validity	The validity of this measure depends on the validity of the analysis that defines performance criteria.
Reliability	Observational data are susceptible to observer bias. Reliability can be enhanced by providing structured guidance for observation.
Sensitivity	Variable, depending on the context.
Diagnosticity	Variable, depending on the context.
Intrusiveness	Mildly intrusive. Operator behavior may change when he knows he is being observed
Objectivity	This measure is subjective, but objectivity can be enhanced by providing observers with structured guidance and criteria for evaluation.
Practicality	It is usually practical to collect accuracy data, however criteria must be established before a study so that the data can be collected automatically.
Scenario Dependence	This measure is highly scenario dependent.
Summary	Primary task performance as assessed by expert opinion is useful in large scale simulator studies where discrete performance criteria are difficult to define. However, observation may impact the operator's performance, and evaluation is subject to observer biases.
Applicability to AdvSMR domain	Performance assessed via observation is applicable to the AdvSMR domain, however it is recommended to be used in conjunction with more objective measures of performance.

Primary Task Time to Initiate/Time to Complete

Performance is often evaluated based on the time it takes to initiate an action once an event has occurred, or the time it takes to complete a scenario.

Table 9. Review of Primary Task Time to Initiate/Time to Complete

Metric Characteristic	Discussion

Validity	This is a valid measure of efficiency, but may not reflect overall performance.
Reliability	Variable, depending on the context.
Sensitivity	Not Reported.
Diagnosticity	Not Reported.
Intrusiveness	Low intrusiveness.
Objectivity	Objective performance measure.
Practicality	Times can be automatically recorded with simulator software, so it is highly practical to collect.
Scenario Dependence	The collection of this metric is not scenario-dependent. However, the interpretation of the results is highly scenario dependent.
Summary	This is an objective performance measure that is easy to collect. Differences in performance times may not be meaningful, and researchers should be cautious when interpreting results.
Applicability to AdvSMR domain	Applicable to the AdvSMR domain and recommended in conjunction with other methods.

Response Time

Response time can be used to measure efficiency for performing small-scale, discrete tasks. Rather than a measure of how long it takes to complete an entire scenario, response time refers to the amount of time it takes to respond to a particular stimulus.

Table 10. Review of Response Time

Metric Characteristic	Discussion
Validity	Response time is valid as a measure of efficiency in responding to a stimulus.
Reliability	If the study is designed with proper experimental control, results are repeatable
Sensitivity	Response time is usually sensitive
Diagnosticity	Response time is usually not diagnostic
Intrusiveness	Low intrusiveness. It can be collected automatically.
Objectivity	High Objectivity
Practicality	In small scale experimental studies, it is practical to measure response time. In larger scale simulator studies, correlating a stimulus with a response can be more challenging.
Scenario Dependence	The practical conclusions that can be drawn from response time are highly scenario-dependent
Summary	Response time can be used to test performance using specific aspects of a display of human-system interface, for example, but should not be used to evaluate an entire display.
Applicability to AdvSMR domain	Reponses time may be applicable in design phases and in research of AdvSMR HAC.

4.5.1.2 Metrics for Situation Awareness (SA)

Operator awareness of what is happening in the plant and with the automation has been shown repeatedly to be an important factor in operator and system performance. In highly automated systems, the operator is typically assigned to monitor the system rather than directly control the process him or herself. This can lead to performance problems when the operator must intervene in the system, because the operator may not be as aware of what is happening. This is known as the out-of-the-loop performance problem. Because operators may err if they must take action on an automated system, it is important to assess operator awareness in any study of HAC. See Oxstrand et al. (2013) for a thorough review of the situation awareness (SA) literature.

Situation Awareness Global Assessment Technique (SAGAT)

Endsley (2000) developed a measurement technique to assess operator awareness called the Situation Awareness Global Assessment Technique (SAGAT). SAGAT was developed to be integrated with air traffic control or flight simulation scenarios (though it can be and has been used in other types of scenarios and other domains), and requires that the simulation be suspended at certain points in order to ask the participants questions about what is happening and what they expect to occur next. These questions correspond to the three levels of SA: Level 1—perception of data in the situation, Level 2—understanding the meaning of the data, and Level 3—projection of the near future of the situation. SAGAT is widely used (e.g., Endsley, Selcon, Hardiman, & Croft, 1998), and is considered reliable and corresponds well to operator performance, thus demonstrating validity of the measure. Our evaluation of SAGAT is shown in Table 9.

Table 11. Review of SAGAT.

Metric Characteristic	Discussion
Validity	SAGAT is well validated; the individual queries predict performance and correlates with other objective measures of awareness, including expert ratings (Endsley & Garlund, 2000; Endsley, Selcon, Hardiman, & Croft, 1998)
Reliability	SAGAT is reliable.
Sensitivity	SAGAT is sensitive to differences in display designs and information presented.
Diagnosticity	SAGAT is diagnostic of the three levels of SA, and suitable for identifying changes in SA due to display design.
Objectivity	Objective measure
Intrusiveness	SAGAT is integrated into the simulation, and at certain points the simulation is paused and questions are administered to assess operator awareness. While this may disrupt the natural progression of the crew activities, it has been shown that operators are able to resume the simulation after the interruption with little difficulty (Endsley & Garlund, 2000; Endsley, 1998).
Practicality	SAGAT is designed to be administered in conjunction with a simulation; it is software itself. As long as one has the software and can tailor it to the specific simulation of interest, SAGAT is useful and practical. It may be expensive (I was not able to obtain pricing information)
Scenario Dependence	SAGAT has been used across many different types of scenarios, but is often modified to be specific to each scenario.
Summary	SAGAT is one of the most commonly used and accepted measures of SA, with strong empirical support. It is well validated, reliable, and objective.
Applicability to AdvSMR domain	SAGAT is certainly the most widely used measure of SA, but it would need to be tailored to the AdvSMR domain.

Situation Awareness control room Inventory (SACRI)

Hogg, Follesø, Strand-Volden, & Torralba (1995) developed a measure of SA based on SAGAT but tailored for nuclear power plant control rooms. Like SAGAT, SACRI requires pausing the simulation at certain points throughout the exercise to ask questions about the state of the process.

Table 12. Review of SACRI.

Metric Characteristic	Discussion
Validity	Adapted from SAGAT, a widely validated and reliable measure of SA. Has been used in several studies in the NPP domain (i.e., Sebok, 2000; Hallbert & Sebok, 2000). The validity of this measure depends on the validity of the process parameters selected, which are scenario specific.
Reliability	Adapted from the very reliable SAGAT; while SACRI is not yet widely used, the several studies that have used it do appear to support reliability of the measure (Hogg et al., 1995, Hallbert & Sebok, 2000, Sebok, 2000).
Sensitivity	SACRI has been shown to be sensitive to differences in operator competence, to differences in SA due to changes in the process state, and to changes in the HSI (Hogg et al, 1995).
Diagnosticity	SACRI is sensitive to changes in the HSI, and differences in operator competence, which allows diagnosis of differences in SA.
Intrusiveness	Operators have reported no effect of freezing the simulation to ask questions.
Objectivity	Objective measure of SA in the NPP domain.
Practicality	Practical, given that one has the simulator and the SACRI software.
Scenario Dependence	Specific to the NPP domain, may be tailored to the specific scenario.
Summary	SACRI is specifically designed for the nuclear domain. It was developed for traditional NPP plants, but it is suited for advanced digital control room displays (it was designed at Halden HAMMLAB to enable them to test SA on their advanced control room HSI).
Applicability to AdvSMR domain	Applicable, and highly relevant. Recommended metric.

The Situational Awareness Rating Technique (SART)

The Situational Awareness Rating Technique (SART; Taylor, 1990) is a questionnaire method for measuring SA that focuses on measuring operator knowledge in three areas (on a 7-point ordinal scale from low to high): demands on attention, supply of attentional resources, and understanding of the situation (Gawron, 2000). SART is administered after the task.

Endsley, Selcon, Hardiman, & Croft (1998) compared SAGAT and SART (situation awareness rating technique) on a display evaluation study. SAGAT is an objective measure of situation awareness; SART is a subjective rating of SA by operators. They found that SAGAT and SART both provided sensitivity and diagnosticity related to the evaluation of the display, and SART correlated to subjective operator self-assessments of confidence levels, situation awareness, and performance . SAGAT was highly useful for identifying actual changes in SA due to the display concept. SAGAT and SART were not correlated with each other, however, which showed that the operators' perception of their SA did not always match their actual SA.

Endsley, Sollenberger, & Stein (2000) conducted a study to compare several SA measures, including SAGAT, SART, on-line probe questions similar to the situation present assessment method (SPAM), and a subjective SA assessment, as well as the NASA-TLX to assess operator workload. Participants conducted an air traffic control simulation that used either a traditional or enhanced display, and answered questions for the various SA measures throughout. SART and the on-line probe questions did not show a difference in SA based on the type of display; SAGAT was sensitive to the display type and did show differences, particularly showing how the enhanced display was supportive of operator SA. The various SA measures were poorly correlated with each other, however.

Table 13. Review of SART.

Metric Characteristic	Discussion
Validity	Is not correlated with SAGAT, but it does correlate with simple subjective SA ratings. May be more of a measure of operator confidence than actual SA (Endsley et al., 1998; Endsley, Sollenberger, & Stein, 2000). SART also correlates well with workload measures (Gawron, 2000). Evidence to support validity is weak (Salmon, Stanton, Walker, Baber, Jenkins, McMaster & Young, 2008; SkyBrary, 2013).
Reliability	SART suffers from the inherent reliability problems of all subjective measures (Gawron, 2000).
Sensitivity	Some studies have shown sensitivity to evaluating a display (Endsley et al., 1998; others found that SART was not sensitive to changes in displays (Endsley et al., 2000).
Diagnosticity	Some indication of diagnostic ability related to evaluation of displays, but inconsistent results (Endsley et al., 2000).
Intrusiveness	Unobtrusive, easy to administer and rank the nine questions.
Objectivity	Subjective self-assessment.
Practicality	Practical, quick to administer.
Scenario Dependence	SART can be administered in a wide variety of task types and does not require customization for different domains. It can be used in real-world tasks as well as simulations (Endsley et al., 1998; Pina, 2008).
Summary	Simple and easy to use; may be a better indication of operator confidence or perceived knowledge than actual SA.
Applicability to AdvSMR domain	Only recommended if used in addition to objective measures.

Eye Tracking

Eye tracking can be used to measure workload (see discussion of eye blink rate below), but it has also been shown to be a useful implicit, continuous indicator of situation awareness (Pina et al., 2008). From fixation times, visual scan patterns, and eye point of gaze, it is possible to infer the information being processed at each moment and understand the interaction between the display and the operator (Pina et al., 2008). It is useful for understanding how people develop SA and process information in complex environments. When the research study has identified the important aspects of the display for any given time, increased eye gaze on the important display elements (i.e., fixation time) is indicative of increased SA (Pina et al., 2008). It can also document operator anticipation, when operators look at a piece of information on a display before a change. Limitations include noise-to-signal ratios, possible non-correlation between eye gaze and what the operator is thinking about, and intensive data analysis that requires expertise.

Eye tracking requires expensive scientific equipment, calibration to the display, and is limited to the technical capabilities of the hardware. In other words, it is only possible to include as many displays as the hardware can measure in one session.

Table 14. Review of Eye Gaze Tracking.

Metric Characteristic	Discussion

Validity	Good metric for understanding development of SA and operator attention allocation (Pina et al., 2008).
Reliability	Unknown.
Sensitivity	Unknown.
Diagnosticity	Useful in diagnosing interface problems (Pina et al., 2008).
Intrusiveness	Depends on the hardware. With familiarity, may be unintrusive. Does not intrude on conducting the task.
Objectivity	Objective, based on performance data
Practicality	Requires expensive equipment and integration with software, intensive data analysis (Pina et al., 2008).
Scenario Dependence	Must be tailored and calibrated to the specific scenario.
Summary	
Applicability to AdvSMR domain	Limitations of the display size may limit the applicability to the AdvSMR domain. May be useful in evaluations of single displays, but may be challenging or impossible to use to evaluate an entire control room at once.

Summary of Situation Awareness Metrics

Based on the review of the above SA metrics, the HAC research team recommends SAGAT or SACRI, customized to the AdvSMR domain, and SART for operator self-report. Eye tracking is recommended if budgets, hardware, and analytical capabilities allow.

Table 15. Overall Review of Situation Awareness Measures.

Metric	Advantages	Disadvantages	Important Considerations
SAGAT	• Objective • Widely used • Extensively validated	• Requires integration with the simulation • Requires customization to specific domain	
SACRI	• Designed specifically for use in NPP domain • Based on SAGAT • Objective	• Requires integration with the simulation	
SART	• Ease of use, low cost	• Subjective • Evidence to support validity of SART is weak at best (SkyBrary, 2013)	• Best used in combination with objective metrics
Eye tracking	• Objective	• Requires expensive equipment • Requires calibration to a limited-size display	• Useful to corroborate other SA metrics • Appropriate when task requires a display that is within the capabilities of the hardware to calibrate to • Eye trackers must have a wide enough field of view to capture eye movement across all possible displays

4.5.1.3 Metrics for Workload

Secondary Task Metrics

Secondary tasks are tasks that people must perform in addition to the primary task. The reason secondary tasks are used is that they allow measurement of the spare attentional capacity that remains from the primary task. As workload on the primary task increases, secondary task performance degrades (Miller, 2001; Gawron, 2000), particularly if the secondary task requires the same cognitive resources as the primary task (Hwang et al., 2008; Miller, 2001; Le Blanc et al., 2010).

There are a wide variety of possible secondary tasks (Gawron [2000] lists 29 separate secondary tasks), including but not limited to tracking, monitoring, detection, choice reaction time, mental mathematics, classification, and memory scanning tasks (Gawron, 2000). Common secondary tasks used in many studies in the nuclear power domain use some variety of a target or signal detection task (e.g., Jou et al., 2009; Lin, Yenn, & Yang, 2010).

Table 16. Review Of Secondary Task Metrics

Metric Characteristic	Discussion
Validity	Validity of secondary task metrics for workload is variable.
Reliability	Unknown.
Sensitivity	Objective performance measures of workload (such as performance on a secondary task) may be sensitive for high workload, but not for situations of low workload.
Diagnosticity	Unknown
Intrusiveness	High intrusiveness. In order to be valid, the secondary task must be designed such that the operator must use the same cognitive resources as his primary task. Therefore, it may interfere with the operator's primary task.
Objectivity	Objective, performance-based measure.
Practicality	Can be very practical, particularly if the secondary task is embedded in the primary task.
Scenario Dependence	Selection of the secondary task depends largely on the primary task, and is therefore scenario dependent.
Summary	Secondary tasks are a good metric for workload, but should not be the only measure. It is important to carefully measure primary task performance as well, and to use other indicators of workload.
Applicability to AdvSMR domain	Applicable; care must be chosen in selecting an appropriate secondary task to ensure that it uses the same cognitive resources as the primary task. Only recommended for laboratory tasks which require continuous assessment of workload.

Physiological Metrics of Workload

Heart Rate

Heart rate is easy to record, and has been used in a variety of studies. Heart rate is an indicator of arousal: with increases in boredom, heart rate drops; with increased stress, heart rate increases (Ha et al., 2007; Huang et al., 2006; Hwang et al., 2008). Its primary limitation is that heart rate can be affected by factors other than workload, such as anxiety or physical activity (NATO, 2003; Pina et al., 2008).

Table 17. Review of Heart Rate as a workload metric.

Metric	Discussion

Characteristic	
Validity	Heart rate has been shown to have systematic relationships with information-processing activities and a number of studies (NATO, 2003; Ha et al., 2007; Huang et al., 2006; Miller, 2001). Heart rate varies with workload, but it also is affected by other factors, which means that its validity depends largely on the specific context of the study and whether other factors are controlled or accounted for.
Reliability	Effects are repeatable.
Sensitivity	Sensitive to variations in task demand. Sensitive to effects from physical effort, emotions, stress, has a large signal-to-noise ratio (Kramer, 1991, in NATO, 2003; Miller, 2001).
Diagnosticity	Measures relative levels of workload, not absolute levels (Roscoe, 1992, 1993, in Miller 2001).
Intrusiveness	Low intrusiveness.
Objectivity	Low subjectivity, but prone to contamination by effects of other variables than workload.
Practicality	Does require equipment to record heart rate data, with electrodes or other means of measuring heart rate. Portable recording equipment is available, and does not require many electrodes or precise location. Data does require sophisticated analysis and relationship with performance may be complex (NATO, 2003).
Scenario Dependence	Not scenario dependent.
Summary	NATO (2003) recommends that heart rate data be collected if recording heart rate variability data.
Applicability to AdvSMR domain	Depending on the equipment availability and the ability to analyze the data, heart rate data could be a valuable metric for AdvSMR research.

Heart Rate Variability

Heart rate variability measures the intervals between beats over time (Miller, 2001; NATO, 2003). It has been shown to be a good measure of mental effort when examining the 0.1Hz spectral band (Miller, 2001; NATO, 2003). As workload (or stress) increases, the variability of the heart rate decreases; as workload drops (or as boredom increases), heart rate variability increases (Ha et al., 2007; Huang et al., 2006; Hwang et al., 2008).

Table 18. Review of Heart Rate Variability.

Metric Characteristic	Discussion
Validity	Generally valid, but it depends largely on the analysis technique (NATO, 2003)
Reliability	Reliable, successfully applied in a variety of environments.
Sensitivity	High sensitivity to changes in workload, but vulnerable to confounding effects from stress or other factors in the environment. May have better accuracy than heart rate alone (Miller, 2001).
Diagnosticity	Good as a global measure of effort (NATO, 2003)
Intrusiveness	Depends on the equipment, but with modern cardiac equipment can be relatively unobtrusive (NATO, 2003)
Objectivity	Low subjectivity, but prone to contamination by effects of other variables besides workload.
Practicality	Calibration and rest measurements are required between test conditions, and at least two minutes of data must be collected to determine HRV.

Scenario Dependence	Not scenario specific.
Summary	NATO (2003) recommends HRV if there is interest in obtaining an objective measure of effort for periods of at least a few minutes.
Applicability to AdvSMR domain	Depending on the equipment availability and the ability to analyze the data, heart rate variability data could be a valuable metric for AdvSMR research.

Eye Blink Rate

Eye blink rate is generally accepted as a reasonable measure of visual workload (Miller, 2001; NATO, 2003; Hwang et al., 2008; Ha et al., 2007). With increasing workload, blink frequency and duration decrease; in other words, people blink less often and for shorter lengths of time as their workload increases. One advantage of using blink rate to assess workload is that it can measure changes in workload within the task.

Table 19. Review of [insert metric here].

Metric Characteristic	Discussion
Validity	Valid measure of visual workload; not valid for measuring other types of workload.
Reliability	Consistent relationships between task loads and blink rate have been found across a variety of populations and tasks (Kramer, 1991, in NATO, 2003).
Sensitivity	Sensitive to changes in visual task demands.
Diagnosticity	Diagnostic of visual workload.
Intrusiveness	Requires expensive equipment that involves electrodes placed above and below the eyes; nevertheless studies have found high operator acceptance and low intrusiveness (NATO, 2003).
Objectivity	Objective measure of visual workload.
Practicality	Requires equipment, calibration, and electrodes, but has been successfully applied in multiple environments.
Scenario Dependence	Scenario independent.
Summary	NATO (2003) notes that while specialized equipment is required, eye blink data is valuable if there is particular interest in the visual processing demands of the primary task.
Applicability to AdvSMR domain	Applicable; may be more relevant to HSI or system design evaluation/validation.

Subjective Metrics of Workload

Modified Cooper-Harper (MCH) Scale

The original Cooper-Harper scale was developed for pilots to rate the handling qualities of aircraft (NATO, 2003). It has since been modified to be a 10-point uni-dimensional rating scale that produces a global measure of workload (Hill et al, 1992, in Miller, 2001). The MCH can be used to measure perceptual, cognitive, and communications workload (Wierwille, 1983; Gawron, 2000). It has variable reliability and validity, however. Some studies have found it to be a good estimator of overall workload, but others have found it to be of little value (Miller, 2001). It apparently depends on how well the operators accept the measure, which uses a decision tree to assign a value to the workload of 1-10.

Table 20. Review of [insert metric here].

Metric Characteristic	Discussion
Validity	While correlated with other measures of task performance and workload (NATO, 2003), some researchers (e.g., Hill et al, 1992) have found it to have a poor description of workload (Miller, 2001).
Reliability	Depends largely on operator acceptance; generally high reliability (NATO, 2003; Gawron, 2000).
Sensitivity	Sensitive to variations in task difficulty, but not as sensitive to differences in task load (NATO, 2003).
Diagnosticity	Not diagnostic.
Intrusiveness	Low intrusiveness; operator navigates a decision tree after the task.
Objectivity	Subjective self assessment.
Practicality	Easily administered.
Scenario Dependence	Questions in the MCH are not specific to any domain or scenario (Gawron, 2000), but Miller (2001) suggests the MCH may work better in some environments than others.
Summary	NATO (2003) considers it a useful simple metric for overall workload, but it is not diagnostic. Miller (2001), on the other hand, considers the scale to have contradictory support. It may therefore be context-dependent and work better in some environments than others.
Applicability to AdvSMR domain	May be applicable, but not yet recommended.

Instantaneous Self Assessment (ISA)

The Instantaneous Self Assessment technique is a simple tool for obtaining real-time perceived workload during a task. The operator gives a rating of perceived workload or busyness on a scale from very low to very high (NATO, 2003). The advantage of this technique is that it is possible to obtain perceived workload estimates mid-task, and therefore is somewhat sensitive to changes in workload across a task.

Table 21. Review of ISA.

Metric Characteristic	Discussion
Validity	Highly correlated with the NASA-TLX and other measures of workload (NATO, 2003).
Reliability	Depends on operator report; reliable if operators are accurate in their self-reports.
Sensitivity	The five-point scale limits the sensitivity somewhat.
Diagnosticity	Not diagnostic.
Intrusiveness	Lowest intrusiveness of all subjective measures reviewed in NATO (2003).
Objectivity	Subjective self-assessment.
Practicality	Very practical, minimal equipment required (NATO, 2003). Useful in dynamic environments.
Scenario Dependence	Not scenario specific.
Summary	Useful as a rough estimation of workload.
Applicability to AdvSMR	Potentially relevant to AdvSMR, depending on the specific task in question. Recommended for intra-task estimates of workload.

domain	

NASA-TLX

The NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988) is a widely used and validated scale for measuring workload after a task (Ha et el., 2007; Sebok, 2000; Huang et al.,2006; Converse, 1995; Hwang et al., 2008; Hallbert & Sebok, 2000; Le Blanc et al, 2010; Miller, 2001; NATO, 2003). It is a six-item scale that was developed specifically for the aviation industry, though it has been used in hundreds of studies in a wide variety of fields, including many nuclear power plant control room studies (Le Blanc, 2010; Hart, 2006). It has been shown to be a reliable measure of differences in workloads between tasks in many different conditions (Hart, 2006).

Table 22. Review of NASA-TLX (drawn from NATO, 2007).

Metric Characteristic	Discussion
Validity	Extensively validated, correlates well with other measures of workload.
Reliability	High reliability, widely used in numerous domains.
Sensitivity	High sensitivity to changes in task load.
Diagnosticity	Provides scores on six dimensions (Mental, Physical, Temporal, Effort, Performance, Frustration), each of which is differentially diagnostic, as well as a global score.
Intrusiveness	Assessed after the task, off-line, takes less than 5 minutes to complete.
Objectivity	Subjective self-assessment of workload.
Practicality	Consists of event scoring and a paired comparison weighting process to determine the importance of each factor of the task in question.
Summary	A very well-established, validated, and well-accepted metric that has a sound technical basis.
Applicability to AdvSMR domain	Applicable, and recommended.

Subjective Workload Assessment Technique (SWAT)

SWAT was developed for aircrew studies, and has scales for time load, mental effort load, and psychological stress load, each with three levels (low, medium, high; NATO, 2003). Participants rank the 27 possible combinations of these three scales according to their perception of increasing workload, then use those rankings to assess the workload of particular tasks. This scale takes more time and effort to complete, and while the overall results can provide useful data, the underlying dimensions are not orthogonal (Miller, 2001; NATO, 2003).

Table 23. Review of SWAT.

Metric Characteristic	Discussion
Validity	Underlying dimensions not empirically validated, nor are they orthogonal (Miller, 2001; NATO, 2003; Gawron, 2000). Gawron (2000), on the other hand, states that SWAT has been found to be valid.
Reliability	Reliable, even if reporting is delayed by 30 minutes (NATO, 2003; Gawron, 2000). SWAT ratings are less variable then MCH (Gawron, 2000).
Sensitivity	Less sensitive than the NASA-TLX, but still high and widely tested (NATO, 2003; Gawron, 2000).
Diagnosticity	The three scales have been found to be differentially diagnostic (NATO, 2003).

Intrusiveness	More demanding than other subjective measures, can take over an hour to complete and has lower user acceptance than NASA-TLX; it takes time for the operators to learn to do the initial ranking (Gawron, 2000; NATO, 2003; Miller, 2001).
Objectivity	Subjective self assessment.
Practicality	Less practical than NASA-TLX.
Scenario Dependence	Has been used in diverse environments; not scenario dependent (Gawron, 2000).
Summary	More effort required and more subjective than NASA-TLX.
Applicability to AdvSMR domain	Not recommended.

Summary of Workload Metrics

Given the high number of possible measures of workload, it was important to focus on identifying the metrics that had the largest potential to be relevant for AdvSMR HAC research. The research team recommends a combination of objective and subjective metrics, including secondary task performance, NASA-TLX, ISA, and physiological metrics of heart rate, heart rate variability, and blink rate, if equipment allows.

Table 24. Overall Review of Workload Measures.

Metric	Advantages	Disadvantages	Important Considerations
NASA-TLX	• Widely used • Validated • Applicable across scenarios and domains	• Must be administered after the task • Cannot measure within-task changes in workload	• Can only be used to measure workload across an entire trial • Subjective measure
Performance on a Secondary Task	• Can be used to measure within-task changes in workload	• Secondary task performance may not be sensitive to all changes in primary task workload.	• In order to gain maximum sensitivity, make sure that secondary task uses the same mental resources as the primary one. • May reduce ecological validity if the secondary task is artificial (i.e. not a realistic secondary task)
Eye Blink Data	• Continuous measurement • Non-intrusive (provided that modern equipment is used) • Can measure changes in workload within a single task	• Requires expensive equipment • Requires extensive data analysis	
Heart Rate Data (rate and variability)	• Objective • Provides continuous, real-time measure of ongoing operator workload (Pina et al., 2008)	• Requires extensive data analyses • Data collection is subject to noise; can be influenced by factors aside from workload	• Make sure that equipment is synched with other data collection methods to ensure accurate timing
SWAT	• Used in diverse environments, not scenario specific	• Takes time and effort for operators to complete • Underlying dimensions	• Has lower user acceptance than NASA-TLX due to the time and effort to

Metric	Advantages	Disadvantages	Important Considerations
	• Reliable ratings	are not orthogonal	complete
MCH	• Practical, easily administered • Simple	• Not diagnostic • Depends largely on how well operators accept the metric	• Has variable and contradictory empirical support
ISA	• Possible to obtain perceived workload rating mid-task • Easily administered, very practical • Useful in dynamic situations • Correlates well with NASA-TLX	• Not diagnostic, subjective self-assessment • 5-point scale limits the sensitivity somewhat	• Useful as a rough estimate of workload

4.5.2 Metrics for HAC performance

4.5.2.1 Metrics for Trust

Operator trust in automation has long been identified as a key issue for development and performance of automated systems (Khasawneh, Bowling, Jiang, Gramopadhye, & Melloy, 2003; Lee & Moray, 1992; Lee & See, 2004; Madsen & Gregor, 2000; MORE REFERENCES). Specifically, trust plays an important role in personnel use of automation (Lee & See, 2004; Merritt, Heimbaugh, LaChapell, & Lee, 2013). The benefits of automation can be offset when operators do not trust it. If operators do not trust automation, they may not use it; or if they do, their workload may be significantly increased by overly verifying the automation's behavior. Similarly, failures of automation can remain undetected if operators trust it too much and hence, become complacent. Trust is closely related to the reliability of the automation; with higher reliability, operator trust in the automation is likely to be high. Reliability is an objective variable that can be manipulated by researchers; trust is often one of the dependent variables evaluated to determine performance (REFERENCES). Trust is therefore an important variable to measure for the purpose of HAC research and design.

There are a number of models of trust (e.g., Muir, 1994; Muir & Moray, 1996; Liu & Hwang, 2000), but there is a surprising dearth of specific metrics of trust given the importance of the variable. Many studies that measure trust (e.g., Muir, 1994; Muir & Moray, 1996; Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003) either do not specify how they measured trust, or they created their own study-specific set of questions. To date, the project team was able to find only three standard measures of trust that were developed with any sort of psychometric rigor: the Jian, Bisantz, & Drury (2000) Trust Scale, the Madsen & Gregor (2000) Human-Computer Trust Scale, and the Freedy et al. (2007) Objective Trust Measure.

Jian et al. (2000) Trust Scale

Jian et al. (2000) developed an empirically derived scale of trust in a three-phase study in which they used factor and cluster analysis of words related to trust to create a multidimensional measurement scale. This scale has been used in multiple studies (e.g., Seong & Bisantz, 2008; REFERNCE MORE). Their scale consists of 12 Likert-style items on a questionnaire, each of which is rated from "not at all" to "extremely." The trust scale items are shown below:

1. The system is deceptive
2. The system behaves in an underhanded manner
3. I am suspicious of the system's intent, action, or output
4. I am wary of the system

5. The system's action will have a harmful or injurious outcome
6. I am confident in the system
7. The system provides security
8. The system has integrity
9. The system is dependable
10. The system is reliable
11. I can trust the system
12. I am familiar with the system

Additionally, the researchers did not find significant differences in trust in automation, humans, or trust in general, indicating that researchers may not need to treat those types of trust differently. However, questions that assess perception of the automation's intent (e.g., "the system is deceptive") may not be relevant or appropriate for AdvSMR research. This should be evaluated further.

Table 25. Review of Jian et al. (2000) Trust Scale.

Metric Characteristic	Discussion
Validity	Validated by a factor analysis study (Spain, Bustamante, & Bliss, 2008).
Reliability	Reliability for this metric has not been reported.
Sensitivity	The scale was able to distinguish between different levels of trust based on automation performance, and to show how trust changed across trials (Bisantz & Seong, 2001).
Diagnosticity	Unknown.
Intrusiveness	Questionnaire is administered after the simulation. Not intrusive.
Objectivity	Based on participant self-report. Subjective.
Practicality	Very practical; administered easily via pen and paper or on a computer.
Scenario Dependence	Questionnaire items are not scenario specific.
Summary	This measure is potentially useful, but has not been extensively used or validated.
Applicability to AdvSMR domain	May be useful for gauging operator perceived levels of trust. Not AdvSMR specific, but is general enough that it may not matter. Given the lack of other validated trust metrics, this may suffice until a new metric can be developed.

Madsen & Gregor (2000) Human-Computer Trust Scale

Madsen and Gregor conducted a rigorous metric development process to create a five-construct, 25-item questionnaire to measure trust in computer systems, specifically decision support systems (Madsen & Gregor, 2000; Pina et al., 2008). The Human-Computer Trust (HCT) Scale's five constructs are perceived reliability, perceived technical competence, perceived understandability, faith, and personal attachment. Each construct consists of five questions that were selected for having the best predictive validity.

The items in the HCT are as follows (Pina et al., 2008):

1. Perceived Reliability
    a. The system always provides the advice I require to make my decision.
    b. The system performs reliably.
    c. The system responds the same way under the same conditions at different times.
    d. I can rely on the system to function properly.
    e. The system analyzes problems consistently.
2. Perceived Technical Competence

a. The system uses appropriate methods to reach decisions.
b. The system has sound knowledge about this type of problem built into it.
c. The advice the system produces is as good as that which a highly competent person could produce
d. The system correctly uses the information I enter.
e. The system makes use of all the knowledge and information available to it to produce its solution to the problem.

3. Perceived Understandability
   a. I know what will happen the next time I use the system because I understand how it behaves.
   b. I understand how the system will assist me with decisions I have to make.
   c. Although I may not know exactly how the system works, I know how to use it to make decisions about the problem.
   d. It is easy to follow what the system does.
   e. I recognize what I should do to get the advice I need from the system the next time I use it.
4. Faith
   a. I believe advice from the system even when I don't know for certain that it is correct.
   b. When I am uncertain about a decision I believe the system rather than myself.
   c. If I am not sure about a decision, I have faith that the system will provide the best solution.
   d. When the system gives unusual advice I am confident that the advice is correct.
   e. Even if I have no reason to expect the system will be able to solve a difficult problem, I still feel certain that it will.
5. Personal Attachment
   a. I would feel a sense of loss if the system was unavailable and I could no longer use it.
   b. I feel a sense of attachment to using the system.
   c. I find the system suitable to my style of decision making.
   d. I like using the system for decision making.
   e. I have a personal preference for making decisions with the system.

This measure was designed to evaluate trust in intelligent decision aid systems that either provide advice or make decisions subject to the user's discretion (Madsen & Gregor, 2000). It would therefore be appropriate for lower levels of automation than supervisory control. Another limitation is that the HCT assumes that the operator has some months of familiarity with the system (Pina et al., 2008).

Table 26. Review of the Madsen & Gregor Human-Computer Trust Scale.

Metric Characteristic	Discussion
Validity	Validated by Madsen & Gregor (2000).
Reliability	Madsen & Gregor (2000) provided reliability level (Chronbach's Alpha) of α = .94
Sensitivity	Unknown.
Diagnosticity	Unknown.
Objectivity	Based on operator self-report. Subjective.
Intrusiveness	Can be administered before a task if the operator has familiarity with the system, or after a task. Not intrusive during performance of the task.

Practicality	It is a short, 25-item questionnaire; very practical.
Scenario Dependence	Generic, not scenario specific.
Summary	A good measure for evaluating decision support systems.
Applicability to AdvSMR domain	Not applicable to supervisory control applications, but may be relevant in cases of adaptive automation or lower levels of automation.

Freedy, DeVisser, Weltman, & Coeyman (2007) Objective Metric of Trust

Freedy, DeVisser, Weltman, and Coeyman (2007) developed an objective measure of operator trust in human-robot collaboration in the military, specifically human supervisory control over robotics such as unmanned vehicles and other robotic systems. They set out to mathematically represent the concept of trust and measure the necessary parameters for computing a single score. They used a rational decision model and equations first provided by Sheridan and Parasuraman (2000) to establish a compound "goodness" score that transforms the observed human task allocation decision behavior (i.e., how often the operator overrides or takes control from the robot), risk, and observed robot performance into a relative expected loss score:

$$REL = \sum_{i=1}^{i=n} P_i * C_i / K$$

Where Pi*Ci is the expected loss of robot autonomous control at trial i, n is the number of trials, and K is the number of operator overrides or interventions to take control from the robot.

Freedy et al. (2007) tested this objective measure of trust in a tactical human-robot collaborative task simulation in which they varied robot competency and measured trust. Across trials, the objective measure of trust corresponded to the differences in the robot competency.

The Freedy et al. metric is only relevant in simulator studies where the system is largely autonomous, the operator is in a supervisory role, and the trust measure of interest is operator overrides of the system. This does not apply to other levels of automation, such as decision support. Given that AdvSMR systems will be highly automated and the operator primarily in a supervisory role, this metric is relevant, but it may still need to be adapted to apply better to the AdvSMR HAC domain instead of robotics.

Table 27. Review of Freedy et al. (2007) Objective Metric of Trust.

Metric Characteristic	Discussion
Validity	Unknown.
Reliability	Mathematically determined from operator performance in the simulator. Reliable.
Sensitivity	This measure shows sensitivity to different levels of automation competence (reliability).
Diagnosticity	This measure can diagnose over trust, under trust, and proper trust.
Objectivity	Objective; does not rely on participant self-report.
Intrusiveness	Not intrusive. Calculated directly from performance in the simulation.
Practicality	Very practical, as long as the simulation is designed to record the necessary information to calculate the metric.
Scenario Dependence	This metric is somewhat context-specific. It is integrated into the simulation, which would have to be designed to document operator overrides, risk/consequences of poor performance (specific to the scenario being studied).
Summary	

Applicability to AdvSMR domain	Applicable, as long as the simulations are designed to provide the information necessary to calculate the trust equation. Recommended for simulation studies.

Summary of Trust Metrics

The Jian et al. (2000) Trust Scale provides operator self-assessment of trust in the system, the HCT evaluates trust in decision-aid systems (Madsen & Gregor, 2000), and the Freedy et al. (2007) metric provides objective data on how often the operator overrides the system and mathematically converts that to a metric of trust. A summary of advantages and disadvantages of the trust metrics is shown in Table 16 below. These three measures may be adequate for AdvSMR research, but they may not be sufficient. There may be other aspects of trust that these metrics do not assess (EXAMPLE(S) NEEDED).

Table 28. Overall Review of Trust Performance Measures.

Metric	Advantages	Disadvantages	Important Considerations
Jian et al. Trust Scale	• Unintrusive • Practical • Generic	• Subjective	• Some of the questions ask about the automation's intent (e.g., "the system behaves in an underhanded manner"), which may not be appropriate.
Human-Computer Trust Scale	• Validated and reliable • Developed through a rigorous process • Recommended for evaluating trust in decision support systems • Generic	• Not appropriate for high LOAs such as supervisory control.	• Assumes operator familiarity with the system.
Freedy et al. Objective Metric	• Objective • Unintrusive • Recommended for evaluating trust in highly automated, autonomous systems	• Requires integration with a simulation that is programmed to collect information necessary to calculate the trust score	• Is only relevant when the system is largely autonomous and the aspect of trust of interest is operator overrides of the system.

It is surprising, given the importance of trust in human-automation collaboration, that there are not more metrics of trust than the Jian et al. (2000) questionnaire, the Madsen & Gregor (2000) HCT, and Freedy et al. (2007) quantitative metric. The HAC research team will evaluate whether these three metrics are sufficient for HAC research and design; it may be necessary to develop additional metrics to assess aspects of trust that are important for AdvSMR HAC. Each of the above three metrics are recommended for use in AdvSMR HAC research, in their appropriate applications.

4.5.2.2 Metrics for Interface Management

To perform their primary tasks successfully, personnel must successfully perform interface management tasks, such as navigating or accessing information at workstations and arranging various pieces of information on the screen. In part, these tasks are necessitated by the fact that operators view only a small amount of information at any one time through the workstation displays. Therefore, they must perform interface management tasks to retrieve and arrange the information.

In many ways, interface management tasks are similar to secondary tasks used to gauge operator workload. However, for HAC research is less interested in interface management as a metric of workload, for which there are a number of other metrics, but instead as a means of evaluating the effectiveness and efficiency of the human-automation collaboration. Interface management tasks are not directly associated with monitoring and controlling the plant; however, they can lead to error, delay performance, create high workload, and may divert attention away from primary tasks and make them difficult to perform (O'Hara & Brown, 2002). Additionally, the user interfaces to automation can influence an operator's decision whether or not to use an automated system. Thus, interface management tasks are important and need to be carefully addressed.

Interface management is a scenario dependent measure and must be identified for each scenario, based on the mission objectives of the scenario. However, traditional human factors methods for evaluating system interfaces may be appropriate for assessing interface management. The HAC research team will review these types of methods for applicability to AdvSMR HAC research and design later in the project.

4.5.3 Metrics for System Performance

4.5.3.1 Function Performance

The most direct measure of system performance is how well the human-automation system performs the intended functions. Function performance can be measured by monitoring process values and comparing them to required or ideal values.

Table 29. Review of Function Performance

Metric Characteristic	Discussion
Validity	The validity depends on the quality of the analysis that defined the ideal performance parameters.
Reliability	Variable, depending on the context.
Sensitivity	Variable, depending on the context.
Diagnosticity	Unknown.
Intrusiveness	Measure process variable can be done automatically via system logs, so it has low intrusiveness.
Objectivity	This is an objective measure.
Practicality	Most simulators have the ability to record important process values built-in, so this metric is practical.
Scenario Dependence	This metric is highly scenario dependent.
Summary	
Applicability to AdvSMR domain	This measure is applicable to the AdvSMR domain. It is recommended that it is used in conjunction with human performance measures for validation of systems and in large scale simulator studies.

Discrepancy Scores

Ha, Seong, Lee, and Hong (2007) developed a mathematical formula to calculate the overall deviation in plant parameters across an entire simulated scenario. This is a good objective measure of overall human-system performance.

Table 30. Review of Discrepancy scores

Metric Characteristic	Discussion

Validity	The validity of this measure depends on the quality of the analysis that defines the ideal parameters for the scenario.
Reliability	Not Reported.
Sensitivity	Not Reported.
Diagnosticity	Useful for comparing performance between test scenarios (Ha, Seong, Lee, and Hong, 2007). May also be useful for comparing performance between competing interface designs.
Intrusiveness	Parameters are recorded automatically, so this metric has low intrusiveness.
Objectivity	This is an objective performance measure.
Practicality	Once the analysis for a scenario has been established, it is practical. However, this metric requires lengthy analysis prior to running test scenarios.
Scenario Dependence	Highly scenario –dependent.
Summary	This is an objective, quantitative method for measuring plant performance.
Applicability to AdvSMR domain	Applicable to the AdvSMR domain. Recommend to be used in conjunction with other performance measures.

4.5.3.2 Summary of System Performance Metrics

Table 31. Overall Review of Task Performance Measures.

Metric	Advantages	Disadvantages	Important Considerations
Function Performance	• Objective • Practical	• Scenario dependent	
Discrepancy scores	• Objective • Quantitative	• Scenario dependent • Requires extensive analysis	

4.5.4 Metrics for Crew Performance

Team and crew performance are important aspects of HAC performance. The research team has not reviewed metrics for team performance, but will in future efforts.

4.6 Conclusions

Objective performance measures are the most direct measure of human performance and system performance; however they are highly context-dependent. Therefore, researchers recommend that objective measures of performance (such as accuracy and function and system performance) be used in conjunction with other, less context-dependent measures. Objective measure should be used in conjunction with subjective measures to validate the objective measures, to gain more sensitivity in situations where performance may be subject to ceiling effects, and to gain qualitative insights into system design that may lead to better design in the future.

Although the aim is to design a standardized suite of measures, different metrics may be better in different situations. Researchers should use the criteria tables to aid in making decisions about which metrics to use when the recommended metrics are not appropriate for the research or situation.

4.6.1 Recommended Metrics for HAC Research

The research team recommends the following metrics be used for HAC research.

Table 32. Recommended Metrics for HAC Research

Measure	Recommended Metric(s)	Comments
Primary Task Performance	• Accuracy (or errors) • Time-to-complete • Time-to-initiate • Expert observation	• Expert observation should be used in earlier stages of research and more objective measures should be used in experiments and in validation of designs.
SA	• SACRI • Eye gaze	• SACRI is the preferred method, however eye tracking can be used in addition when it is practical
Workload	• NASA-TLX	
Trust	• Jian et al. (2000) Trust Scale	• It is recommended that a trust measure be developed specifically for AdvSMR HAC research, however the Jian et al. (2000) trust scale is sufficient until a different method can be developed.
System Performance	• Function performance • Discrepancy scores	• Both metrics should be used together and results should be compared with objective and subjective measures of human performance.

5. INITIATORS AND TRIGGERING CONDITIONS FOR ADAPTIVE AUTOMATION IN ADVANCED SMALL MODULAR REACTORS

5.1 Introduction

It is anticipated that Advanced Small Modular Reactors (AdvSMRs) will employ high degrees of automation. High levels of automation can enhance system performance, but often at the cost of reduced human performance. Automation can lead to human out-of the loop issues, unbalanced workload, complacency, and other problems if it is not designed properly. Researchers have proposed adaptive automation (defined as dynamic or flexible allocation of functions) as a way to get the benefits of higher levels of automation without the human performance costs. Adaptive automation has the potential to balance operator workload and enhance operator situation awareness by allocating functions to the operators in a way that is sensitive to overall workload and capabilities at the time of operation. However, there still a number of questions regarding how to effectively design adaptive automation to achieve that potential. One of those questions is related to how to initiate (or trigger) a shift in automation in order to provide maximal sensitivity to operator needs without introducing undesirable consequences (such as unpredictable mode changes). Several triggering mechanisms for shifts in adaptive automation have been proposed including: operator initiated, critical events, performance based, physiological measurement, and model-based methods (Parasuraman et al., 1992).

Additional research is required to identify the appropriate triggering mechanisms for automation changes, and how they should be implemented to minimize any disruptions to the operator's performance when the change occurs.

The purpose of this research is to identify the triggering mechanisms that:

1. Provide the best support for operator and system performance
2. Are likely to be achievable in the advSMR control room context

5.2 Method

For each of the possible initiators listed above, the research team evaluated:

- What influence the method has on HAC performance. Are there any tradeoffs associated with using the method (e.g.., reduced workload under some conditions but increased workload under others)?
- The feasibility of using the method in the context of an advSMR control room

This evaluation was based on a review and synthesis of the existing literature on adaptive automation and a critical evaluation of the contextual factors that may reduce the feasibility of some of the methods in the advSMR control room. The research team reviewed psychological and human factors literature related to adaptive automation. The researchers used the following list of questions as a guide to summarize each research article that was reviewed.

- What is the main objective of the article (e.g., compare static automation to adaptive automation, review methods of adaptive automation, compare levels of automation, etc.)?
- What domain is the article focused on (process control, combat, aviation, etc.)?
- Is this article an experimental evaluation of adaptive automation?
- Does the study measure human performance and/or human system performance of adaptive automation?
- What triggering conditions were used in the implementation of adaptive automation? If multiple triggering conditions were used or compared, describe them all.

- Is there any basis for determining how the triggering condition affected performance in this study?
- What performance measures were used? (e.g., workload, SA, errors), and how did the adaptive automation method affect the results?
- Are there any important discussions or conclusions related to adaptive automation?

5.3 Results

5.3.1 Operator-initiated

Operator initiated adaptive automation (also known as adaptable automation), gives the operator the decision authority to switch to a higher or lower level of automation. Operators may prefer this type of adaptive automation, because it gives them a greater degree of control (Sauer, Nickel, & Wastell, 2013). However, there is evidence that performance using this triggering condition is not optimal. Sauer, Nickel, and Wastell (2013) compared performance using adaptable automation on a simulated process control task to three static levels of automation (manual, decision aiding, and operation by consent). They found that the adaptable automation condition produced the worst performance in every situation except when the automation failed (in which case, performance was roughly equal to that of operation by consent). The authors concluded that under conditions of adaptable automation, operators chose to use lower levels of automation, thus compromising overall system performance.

Similarly, Kaber and Riley (1999) evaluated performance using adaptive automation that was triggered based on performance on a secondary task. In one condition, the computer generated a suggestion to change the level of automation (up or down), but the operators were not required to follow the suggestion. In the other condition, the computer-generated switch was mandatory. While both adaptive automation conditions enhanced performance compared to manual control, participants in the mandated adaptive automation condition performed better than those on the non-mandated condition. Further, secondary task performance was also lower in the non-mandated conditions, which the authors interpreted as an effect of increased workload caused by the need to evaluate the computer-generated suggestions. This study did not compare non-mandated adaptive automation to static automation, so it is difficult to draw general conclusions on how non-mandatory automation influences performance compared to other levels of static automation.

Conversely, Kidwell et al. (2012) found that operator-initiated adaptive automation enhanced change detection performance compared to a performance-based adaptive automation. The adaptable automation also enhanced operator confidence in decision making. However, the adaptable automation also increased workload. This study did not compare adaptive and adaptable automation to static conditions, so the conclusions that can be drawn from are limited.

While traditional approaches to operator-initiated adaptive automation typically define adaptability as shifting the level of automation (i.e., how much the automation does), several researchers have advocated for designing adaptive automation based on a hierarchical task delegation method (Miller & Parasuraman, 2007; Parasuraman et al., 2005; Miller et al., 2011; Shaw et al., 2010). In this adaptive automation scheme, the operator still has final decision authority over shifts in automation; however the functions are broken into predetermined tasks that the operator can choose to delegate to automation. Researchers argue that this automation scheme is analogous to the way human supervisors of teams delegate and that this type of adaptable automation may be ideal for managing tradeoffs between operator workload and system unpredictability (Miller & Parasuraman, 2007).

Several studies that have investigated adaptable automation based on task-delegation interfaces have found that this type of operator initiated adaptive automation enhances performance (Parasuraman et al., 2005; Miller et al., 2011; Shaw et al., 2010). Studies have also found that performance is highest when the delegation interface includes the highest level of abstraction (analogous to level of automation in this scheme), even when it is not appropriate for the current context. In two separate studies, participants executed scenarios in which the operators had to revert to fully manual performance (the lowest level of abstraction in the task delegation interface), however their performance was still enhanced by having the higher level of abstraction available (Miller et al., 2011; Shaw et al., 2010). The researchers argued that this adaptive automation scheme allows operators to use high levels of automation (thus, reducing workload) without becoming complacent (Miller et al., 2011).

One of the benefits of operator-initiated adaptive automation is that it is relatively straightforward to implement. The decision of when to shift to automation is made by the operator in real-time; therefore it does not need to be made in the design phase. Unfortunately, the fact that the operator needs to make the decision of when to automate may contribute to increased workload ((Miller & Parasuraman, 2007). Additionally, when given the choice, operators may prefer to use lower levels of automation even when it hinders system performance (Sauer, Nickel, & Wastell, 2013). Recent research on adaptable automation using hierarchical task delegation is promising, and may provide a method to effectively implement adaptive automation that leaves authority in the operator's hands without compromising system performance. Additional research is needed to understand how the task delegation interface would be designed in the context of AdvSMR, and how the functions and tasks would be abstracted into differing levels.

5.3.2 Critical events

Critical event triggering may be the most straightforward way to initiate adaptive automation (Inagaki, 2003). The use of critical events as a triggering condition for adaptive automation has received some criticism because this method is not necessarily sensitive to overall operator workload (Inagaki, 2003). However some researchers argue that in time-critical fault situations, automation should be triggered because the human does not have the time to respond (Moray,Inagaki, & Itoh, 2000).In a study investigating human operator's ability to respond to time-critical faults, Moray, Inagaki, and Itoh (2000) found that adaptive automation triggered by critical events enhanced performance during severe accidents.

Although there hasn't been much research on the use of critical events as a triggering condition for adaptive automation, it should be used in time-critical situations (Moray,Inagaki, & Itoh, 2000). Critical events can be implemented along with other adaptive automation schemes in AdvSMRs.

5.3.3 Operator performance measurement

Several studies have found that adaptive automation triggered by operator performance (on a primary or secondary task) enhance performance compared to manual control (Kaber, Wright, Prinzel & Clamann, 2005; Kaber, Perry, Segall, McClernon, & Prinzel, 2006; Kaber, & Riley, 1999).The fact that performance-based adaptive automation enhances performance compared to manual control does not make a strong case for performance-based adaptive automation. Automation, in general, enhances system performance unless there is an automation failure. Fortunately, studies have also found that performance-based adaptive automation can enhance performance compared to static automation (Calhoun, Ruff, Spriggs, & Murray, 2012; Parasuraman, Cosenzo, & De Visser, 2009). Finally, one study found that performance based adaptive automation using a change detection task also enhanced SA compared to static automation (Parasuraman, Cosenzo, & De Visser, 2009).

Though several studies have found that performance-based adaptive automation enhance performance compared to manual performance and static automation, two studies have compared performance-based initiators with other triggering conditions. One study compared two methods for triggering adaptive automation: performance and heart rate variability (HRV). The authors found that performance was best under adaptive automation triggered by HRV (Lagu, Landry, & Yoo, 2013).Another study compared performance-based adaptive automation with operator-initiated adaptive automation, and found that participants under operator-initiated adaptive automation performed better than under performance-based adaptive automation (Kidwell, Calhoun, Ruff, & Parasuraman, 2012).

The results of studies investigating performance-based adaptive automation are somewhat inconsistent. Furthermore, adaptive automation based on performance measurement requires that the operator have a frequent or routine task for which performance can be measured. This is not necessarily feasible in a supervisory control environment where the operator has little or no overt tasks assigned to him. If AdvSMR are to be highly automated it is unlikely that performance-based adaptive automation will be a feasible triggering condition.

5.3.4 Operator physiological assessment

There are a variety of ways to assess operator state (e.g., workload, engagement, and fatigue) using physiological measurements. The most common physiological methods used to assess operator state for the purposes of initiating adaptive automation are Electroencepholagram (EEG) and Electrocardiogram (ECG). EEG is typically used as a way to measure operator engagement (Prinzel et al., 2003; Freeman, Mikulka, Scerbo, & Scott, 2003).ECG is typically used as a way to measure workload, and the most common method is using heart rate variability (HRV).

Some studies have found that adaptive automation imitated by EEG-based measures of engagement enhance performance compared to yoked-control groups (Prinzel et al., 2003; Freeman, Mikulka, Scerbo, & Scott, 2003).Interestingly, these studies found that negative feedback (defined as an decrease in the level of automation when engagement was low) was better than positive feedback (defined as an increase in the level of automation when engagement was high). This implies that for supervisory control applications, it may be effective to manage operator engagement and performance by using EEG-based adaptive automation.

Another study used EEG to measure workload and compared performance on the adaptive automation that was triggered when workload was high to fully manual performance. The researchers found that performance was better with the EEG-based adaptive automation, but a t the cost of temporarily reduced situation awareness (Dorneich et al., 2006). This indicates that EEG-based measures of workload may not be as good as EEG-based measures of engagement in balancing performance and situation awareness.
A study conducted by Lagu, Landry, and Yoo (2013) compared performance using HRV as a triggering condition to performance-based triggering of adaptive automation. They found that performance was best using HRV.

The advantage of physiological assessment is that it can be used measure operator state continuously without any intervention form the operator. Therefore, physiological assessment may be better than performance based-measures for supervisory control applications because it doesn't require frequent actions to be taken by the operator (which is unlikely in a supervisory control context). One disadvantage of physiological assessment is that is highly intrusive.

5.3.5 Modeling

Another way to initiate adaptive automation is to use models. Models are typically defined during the design phase, meaning that certain scenarios, scenarios, evolutions or sets of conditions would have set levels of automation. One study investigated adaptive automation based on a predefined model of workload and compared performance with clumsy automation (increasing automation during the low workload conditions) and manual control. The researchers found that performance was best under adaptive automation with matched workload, i.e., increasing automation as workload increased

(Parasuraman, Mouloua, & Hilburn, 1999). Another study, conducted by Kaber, Wright, and Sheik-Nainar (2006), used a similar method based on a model of expected workload and compared two different schemes of adaptive automation. One used primarily manual control and the other used primarily supervisory control. The researchers compared performance and situation awareness in each adaptive automation condition to fully manual and supervisory control conditions. Situation awareness was best in manual control, followed by the adaptive automation condition with primarily manual control. The conditions with primarily supervisory control produced the lowest situation awareness. Further, the researchers found that there was an immediate drop in situation awareness following a shift in SA.

5.4 Conclusions

Most studies that investigate adaptive automation have found that adaptive automation, regardless of the triggering condition, enhances performance. This makes it difficult to assess the differential effectiveness of triggering condition because each method has a least some empirical evidence indicating that it is effective. Further, there are very few studies that directly compare different triggering conditions. Most of the studies select one triggering condition and compare adaptive automation to static levels. The majority of studies that compare adaptive automation performance to performance using static automation typically use fully manual and fully automatic levels, but do not compare adaptive automation to intermediate levels of automation. In order to fully assess the utility of adaptive automation compared to static automation, future research needs to compare adaptive automation across a variety of levels of automation.

The use of physiological measures, particularly EEG, to trigger adaptive automation is promising. In several studies, superior performance was reported for adaptive automation using these methods. However, there is still some question as to whether it will be practical to employ these methods in a control room due to the intrusiveness of the technology.
Another promising method for triggering adaptive automation is adaptable automation using hierarchical task delegation. Research needs to address how to define the levels of abstraction for AdvSMR in the task delegation interface.

Many of the studies reviewed did not measure all of the important aspects of performance in order to fully assess the effectiveness of adaptive automation. Adaptive automation is typically regarded as a way to achieve enhanced system performance of increased automation while also maintaining operator situation awareness. Many studies measure operator performance, but not situation awareness. Furthermore, in order to assess the effect adaptive automation may have on an operator’s ability to regain manual control, measuring situation awareness may not be sufficient. Future research should investigate the effect of adaptive automation on an operator’s ability to resume manual control in the event of an automation failure (or the ability to detect an automation failure).

6. GENERAL CONCLUSIONS AND PATH FORWARD

The purpose of the HAC research effort is to do the foundational research needed to enable AdvSMRs to achieve safety goals and economic viability. It is anticipated meeting these goals will likely be achieved through a greater use of automation. A large body of research suggests that higher levels of automation can lead to desirable performance outcomes, but often at the cost of reduced operator monitoring (often referred to the human-out-of the-loop phenomenon). This reduced monitoring can have negative consequences in the event of an automation failure, leading to the operator having difficulty in regaining manual control.

Research activities conducted by the HAC research team identified a broad spectrum of gaps related to the understanding of how collaboration between automated systems and human operators should be defined, and how it should be implemented to support performance and safety goals. Based on the findings from the previously conducted activities the HAC research team selected a couple of initial gaps to address through analytical studies and field studies. The result from these studies provides valuable information regarding short-term results that are useful to both the research field and the industry. The selected topics for the analytical studies were; models of teamwork, performance measures, and triggering conditions for adaptive automation.

The field study revealed that operators of AdvSMR may have similar roles to those at IF power. According to NUREG/CR-1368 the human operator's primary functions would be to monitor and verify performance of safety systems, maintain communication with appropriate onsite and offsite personnel, and initiate recovery actions following an event (NRC, 1987). This operator role is similar to the operators at IF Power that participated in the field study. The filed study also indicated that human can interact effectively with near-autonomous systems. The IF Power operators stated that they managed to stay engaged in the process due to their well-defined and specific roles. The automatic system is not equipped to do the tasks that the operators do manually, therefore the operators serve an essential function other than their supervisory control roles. The IF Power operators reported that they fully trusted the automation to correctly execute the tasks assigned to the automation and that they could not recall a time when an operator overrode the automatic system.

In the case of the division of roles and responsibilities at IF Power one can argue that the operators and the automated systems were working in parallel rather than together, a model that seem to work very well for IF Power. However, in the context of AdvSMRs designers and researchers are hoping to design a more dynamic collaboration between the different agents. Therefore, it is important to study existing models of teamwork, both for human-human teams and human-automation teams. The research team also investigated the specific constraints and requirements related to teamwork in the nuclear industry. The analytical study of teamwork suggests that the demands on human-automation interaction might be the greatest when the automation agent is supporting higher-level cognitive process, such as situation assessment and response planning. The result of the teamwork study suggests the following design requirements for designing human-automation interaction:

- The way information is processed is accessible to the operator
- The design of communication functions and features enable operators to obtain the information they need to use automation information, such as assessing the credibility of the results
- Interruptions are minimized
- The level of detail can be controlled
- Operators can provide input to the processing and direct its activities

In order to facilitate optimal operator and plant performance in a highly automated system it is necessary to be able to measure performance of the system, operator, plant, and the human-automation interaction in this specific context. The performance measures analytical study aimed to develop an initial set of performance measures that are applicable in this domain and focus on the relationship between humans and automation, including operator awareness, trust in the automation, and use of automation, as well as measures that reflect multi-agent teamwork.

The main conclusions from the performance measures study are that objective performance measures are the most direct measures of human and system performance, however, they should be used in conjunction with subjective measures to validate the objective measures and to better detect ceiling effects. The subjective measures can also be used to gain qualitative insights useful for future system design. As described in section 4.6.1, the researchers recommend a set of metrics to be used to measure primary task performance, situational awareness, workload, trust, and system performance.

There are very few studies that directly compare triggering mechanisms for adaptive automation, making it difficult to assess which mechanism is the most effective. A review of the adaptive automation literature indicates that there are tradeoffs associated with each triggering mechanism, but that some are more applicable to the AdvSMR domain. The two mechanisms that consistently improve performance in laboratory studies are operator-initiated adaptive automation based on hierarchical task delegation and the Electroencephalogram (EEG) –based measure of engagement. Current EEG methods are intrusive and require extensive analysis; therefore it is not recommended for an AdvSMR control rooms at this time. Researchers recommend that future research focus on comparing performance across multiple triggering conditions and levels of static automation, and investigating the effect of adaptive automation on an operator's ability to resume manual control in the event of an automation failure.

During FY14 a series of studies will be conducted to extensively and empirically investigate gaps, which need to be addressed in order to provide guidance for a long-term solution to the HAC related issues the AdvSMR concept designers will be facing. The overarching research question for the project is "How do we get the most benefits out of using automation while keeping the operators engaged in the process at the same time".

The HAC research team plans to conduct two experimental studies and a series of small scale studies in FY14. Some of the studies will be conducted using collage students as participants. One argument against conducting this type of studies is that it most likely has been done before. However, the research team have found that almost all studies reviewed (regardless if they use student as participants or not) have focused on very simple and non-representative tasks for the contexts of nuclear and/or AdvSMRs. The studies reviewed investigate the effect of one or two variables. When the results of existing studies are compared and synthesized, the studies do not provide consensus or guidance on specific issues important to the AdvSMR HAC research.

One approach to achieving the balance between automation benefits and operator engagement, as mentioned above, is to trade tasks between automation and the operator. The literature review and the analytical studies conducted to date revealed that this trade of tasks is usually thought of in two ways;

1. Through adaptive automation, or
2. By designing the HSI in a manner that keep the operator informed of the status of the process. (This second approach does not solve the issue of operator engagement).

The research team believes that there are two main approached to achieve operator engagement, which both requires further research: 1) modify the HSI, or 2) change the tasks. The latter includes the question regarding the feasibility of assigning tasks that are better suited to be performed by an automated system to the operator in order to keep the operator more engaged in the process.

The HAC research team plans to conduct experimental studies where college students are asked to conduct tasks that are more complex than tasks used in previous studies. Compared to the existing studies, the researchers plan to look at all relevant variables. The goal is to collect as much data as possible to be able to provide insights relevant to the bigger picture rather than focusing on a very specific context. The team will conduct an adaptive automation study using performance based triggering condition. Performance, situation awareness, and workload will be measured and performance between the following variables will be compared:

3. Adaptive automation system and Non-adaptive system
4. Manual and Automated tasks (LOA)
5. Perfect automation and Automation with inserted faults
6. Expected scenario and Outside design basis scenario

The research question and design of the second experimental study depends on the outcome of the first study. The second experimental study will be conducted in the same manner as the first one, i.e. in collaboration with the ISU and using a simulated process system to investigate aspects related to human-automation collaboration in highly automated and highly complex systems.

Some knowledge gaps are better suited to be explored in smaller studies. The researchers envision addressing discrete research gaps and issues in studies that can be conducted (from planning to data analysis) in a shorter time frame than the two larger experimental studies described above. Each individual issue explored and addressed in a small scale study provides valuable insights to the overall knowledge and understanding related to how to best design an effective and reliable collaboration between humans and automated systems. These studies will explore topics especially relevant to AdvSMRs and the nuclear power generating industry. The intent is to involve people from the commercial nuclear industry, navy nuclear operators, and the nuclear operators at INL as participants in the studies. The researchers will conduct a couple of smaller scale studies with the intent to explore items such as the relationship between automation use and trust using more subtle differences in reliability than is typically used in existing psychological research, situation awareness during monitoring of an autonomous system during a manual task: comparison of novices and experts, adaptive automation, and different HSI design concepts to support effective HAC.

One of the objectives with the framework for human-automation collaboration is to specify how the human-automation collaboration should be evaluated to ensure that integrated human-automation system performance acceptably meets design performance requirements for both production and safety. A methodology to do so will be formalized in subsequent research. In addition, as a part of the overall research effort a tool to present the framework for human-automation collaboration for AdvSMRs will be developed. The purpose of the tool is to present the framework in a manner that is more useable and comprehensible to the concept designers than a paper-copy of a report would be. The fourth activity to be conducted in FY14 will be to identify the requirements for the tool to establish a foundation for its development.

7. REFERENCES

7.1 Field Study

Miller, C.A., Parasuraman, R., 2007. Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. Human Factors, 49 (1), 57e75.

Bainbridge, L., 1983. Ironies of automation. Automatica 19 (6), 775e779.

Vicente, K.J., 2003. The Human Factor. Routledge, New York.

Carelli, M. D., Garrone, P., Locatelli, G., Mancini, M., Mycoff, C., Trucco, P., & Ricotti, M. E. (2010). Economic features of integral, modular, small-to-medium size reactors. *Progress in Nuclear Energy*, *52*(4), 403-414.

IAEA (2005). Innovative Small and Medium Sized Reactors: Design Features, Safety Approaches and R&D trends (IAEA-TECDOC-1451). Vienna, Austria: International Atomic Energy Agency.

IAEA (2006). Status of Innovative Small and Medium Sized Reactor Designs 2005: Reactors with Conventional Refueling Schemes (IAEA-TECDOC-1485). Vienna, Austria: International Atomic Energy Agency

de Visser, E. & Parasuraman, R. (2007). Effects of imperfect automation and task load on human supervision of multiple uninhibited vehicles. In Proceedings of the Human Factors and Ergonomics Society 51st Annual Meeting. Santa Monica, CA: Human Factors and Ergonomics Society. (pp. 1081-1085).

Dixon, S. & Wickens, C. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. Human Factors, 48 (3), 474-486

Ruff, H., Calhoun, G., Draper, M., Fontejon, J., and Guilfoos, B. (2004). Exploring automation issues in supervisory control of multiple UAVs. Proceedings of the Human Performance, Situation Awareness and Automation Conference, March 22-25, Daytona Beach, FL, 408-412.

Wickens, C., Li, H., Santamaria, A., Sebok, A., & Sarter, N. (2010). Stages and Levels of Automation: An Integrated Meta-analysis. In Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, Santa Monica: CA: HFES. Wiener, E., & Curry, R. (1980). Flight-deck Automation:

NUREG-1368 (1994). Preapplication Safety Evaluation Report for the Power Reactor Innovative Small Module (PRISM) Liquid-Metal Reactor: Final Report. U.S. Nuclear Regulatory Commission, Washington, DC.

Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications*, 163-181. Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. R. (1997). Level of automation: Integrating humans and automated systems. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*, Santa Monica, CA: Human Factors and Ergonomics Society, 200-204.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42,* 462-492.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381-394.

7.2 Team Work

Bernard, J. & Washio, T. (1989a). *Expert Systems Applications Within the Nuclear Industry*. LaGrange Park, Illinois, American Nuclear Society.

Billings, C. (1991). *Human-Centered Aircraft Automation: A Concept and Guidelines* (NASA Tech. Memo No. 103885). Washington, DC: NASA.

Billings, C. (1997a). *Aviation Automation: The Search for a Human-centered Approach*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Billings, C. (1997b). Some Questions About Advanced Automation. In M. Mouloua & J. Koonce (Eds.), *Human-automation Interaction: Research and Practice* (pp. 314-320). Mahwah, NJ: Lawrence Erlbaum Associates.

Bowers, C., Salas, E., & Jentsch, F. (2006). Creating high-tech teams: Practical guidance on work performance and technology. Washington, D.C.: American Psychological Association.

Carvalho, P. V. R., dos Santos, I. L., & Vidal, M. C. R. (2006). Safety implications of cultural and cognitive issues in nuclear power plant operation. Applied Ergonomics, 37(2), 211-223.

Carvalho, P. V. R., Vidal, M. C. R., & de Carvalho, E. F. (2007). Nuclear power plant communications in normative and actual practice: A field study of control room operators' communications. Human Factors and Ergonomics in Manufacturing, 17(1), 43-78.

Christoffersen, K., & Woods, D. (2002). How to Make Automated Systems Team Players. In E. Salas (Ed.) *Advances in Human Performance and Cognitive Engineering Research* (Vol. 2, pp. 1-12): NY: Elsevier.

Chung, Y. H., Yoon, W. C., & Min, D. (2009). A model-based framework for the analysis of team communication in nuclear power plants. Reliability Engineering & System Safety, 94(6), 1030-1040.

Dekker, S., & Woods, D. (2002). MABA-MABA or abracadabra? Progress on human-automation co-ordination. *Cognition, Technology & Work*, *4*(4), 240 - 244.

Dickinson, T. L., & McIntyre, R. M. (1997). A conceptual framework for teamwork measurement. Team performance assessment and measurement. Team performance assessment and measurement. Mahwah, NJ: Lawrence Erlbaum Associates, 19-43.

Dien, Y., & Montmayeul, R. (1995). Taking Account of Human Factors in Control-Room Design. In *Proceedings of the Topical Meeting on Computer-Based Human Support Systems: Technology, Methods and Future*. LaGrange Park, IL: American Nuclear Society.

Evans, C. R., & Dion, K. L. (1991). Group cohesion and performance a meta-analysis. Small group research, 22(2), 175-186.

Fan, X. & Yen, J. (2011). Modeling Cognitive Loads for Evolving Shared Mental Models in Human–Agent Collaboration. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 41 (2) 354-377.

Feigh, K. & Pritchett, A. (in press). Requirements for Effective Function Allocation: A Critical Review. *Journal of Cognitive Engineering and Decision Making*.

Fiore, S., Jentsch, F., Becerra-Fernandez, I., Salas, E., & Finkelstein, N. (2005). Integrating Field Data with Laboratory Training Research to Improve the Understanding of Expert Human-Agent Teamwork. In *Proceedings of the 38th Hawaii International Conference on System Sciences*. NY: IEEE.

Gertman, D. I., Haney, L. N., Jenkins, J. P., & Blackman, H. S. (1985). Operational decisionmaking and action selection under psychological stress in nuclear power plants. (NUREG/CR-4040). Washington, DC, U.S. Nuclear Regulatory Commission.

Helmreich, R. L., & Foushee, H. C. (1993). Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In E. L. Wiener, B. G. Kanki & R. L. Helmreich (Eds.), Cockpit resource management. (pp. 3-45). San Diego, CA US: Academic Press.

Hutchins, E. (1990). The Technology of Team Navigation. In J. Galegher, R. Kraut, & C. Egido (Eds.), *Intellectual Teamwork: Social and Technical Bases of Cooperative Work*. Hillsdale, NJ: Erlbaum.

IAEA (1995a). *Computerization of Operation and Maintenance for Nuclear Power Plants*. (IAEA-TECDOC-808). Vienna: International Atomic Energy Agency.

INPO 88-003 (1988). Guideline for Teamwork and Diagnostic Skill Development. Atlanta, GA. Institute for Nuclear Power Operations.

INPO SOER 96-1 (1996). Control Room Supervision, Operational Decision-Making, and Teamwork. Atlanta, GA. Institute for Nuclear Power Operations.

INPO 10-001 (2010). Guideline for Training and Qualification of Licensed Operators. Atlanta, GA. Institute for Nuclear Power Operations.

Itoh, J, Yoshimura, S., Ohtsuka, T., & Matsuda, F. (1990). Cognitive task analysis of nuclear power plant operators for man-machine interface design. In Proceedings of the Topical Meeting on Advances in Human Factors Research on Man Machine Interactions, 96-102. Nashville, TN. American Nuclear Society.

Johnson, M., Bradshaw, J., Feltovich, P., Jonker C., Riemsdijk, B., & Sierhuis, M. (2012). Autonomy and Interdependence in Human-Agent- Robot Teams. *Intelligent Systems*, *27*, 43-51.

Kaber, D., Riley, J., Tan, K., & Endsley, M. (2001). On the design of adaptive automation for complex systems. *International Journal of Cognitive Ergonomics*, 5(1), 37-57.

Kim, S. K., Park, J. Y., & Byun, S. N. (2009). Crew resource management training for improving team performance of operators in Korean advanced nuclear power plant. In IEEE International Conference on Industrial Engineering and Engineering Management (IEEE IEEM 2009), 2055-2059.

Klein, G., Feltovich, P., Bradshaw, J., & Woods, D. (2005). Common Ground and Coordination in Joint Activity. In W. Rouse & K. Boff (Eds.), *Organizational simulation*. New York: Wiley.

Klein, G., Wiggins, S., & Dominguez, C. (2010). Team sensemaking. Theoretical Issues in Ergonomics Science, 11(4), 304-320.

Klein, G., Woods, D., Bradshaw, J., Hoffman, R., & Feltovich, P. (2004). Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Intelligent Systems*, 91-95.

Land, S., Malin, J., Thronesberry, C., & Schreckenghost, D. (1995). *Making Intelligent Systems Team Players: A Guide to Developing Intelligent Monitoring Systems* (NASA Technical Memorandum 104807). Houston, Texas: National Aeronautics and Space Administration.

Larson, C., & LaFasto, F. (1989). *Teamwork: What must go right/what can go wrong* (Vol. 10)., London, UK: Sage.

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. Journal of Personality and Social Psychology, 37(6), 822-832.

Lee, J., & See, K. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors,* 46 (1), 50-80.

Lencioni, P. (2002). *The Five Dysfunctions of a Team*. San Francisco, CA: Josey-Bass.

Lenox, T., Lewis, M., Roth, E., Shern, R., Roberts, L., Rafalski, T. & Jacobson. J. (1998). Support of Teamwork in Human-Agent Teams. In the *1998 IEEE International Conference on Systems, Man, and Cybernetics* (Volume 2), 1341- 1347.

Letsky, M., Warner, N., Fiore, S. M., Rosen, M. A., & Salas, E. (2007). Macrocognition in complex team problem solving. Paper presented at the Twelfth International Command and Control Research and Technology Symposium (12th ICCRTS), Newport, RI.

Letsky, M., Warner, N., Fiore, S. M., & Smith, C. A. P. (Eds.). (2008). Macrocognition in Teams: Theories and Methodologies. Hampshire, England: Ashgate Publishing Limited.

Madhavan, P. & Wiegmann, D. (2007). Similarities and differences between human-human and human-automation trust: an integrative review. Theoretical Issues in Ergonomics Science. Vol. 8, No. 4, July-August 2007, pp.277-301.

Malin, J. & Schreckenghost, D. (1992). *Making Intelligent Systems Team Players: Overview for Desigers*. (NASA Technical Memorandum 104751). Houston, Texas: National Aeronautics and Space Administration.

Malin, J., Schreckenghost, D., Woods, D., Potter, S., Johannesen, L., Holloway, M., & Forbus, K., (1991a). *Making Intelligent Systems Team Players: Case Studies and Design Issues. Volume 1: Human-Computer Interaction Design*. (NASA Technical Memorandum 104738). Houston, Texas: National Aeronautics and Space Administration.

Malin, J., Schreckenghost, D., Woods, D., Potter, S., Johannesen, L., & Holloway, M. (1991b). *Making Intelligent Systems Team Players: Case Studies and Design Issues. Volume 2: Fault Management System Cases* (NASA Technical Memorandum 104738). Houston, Texas: National Aeronautics and Space Administration.

Morgan, B., Glickman, A., Woodard, E., Blaiwes, A., & Salas, E. (1986). Measurement of team behaviors in a Navy environment. (Technical Report No. NTSA TR-86-014). Norfolk, VA: Old Dominion University, Center for Applied Psychological Studies.

Mumaw, R. J., Roth, E. M., Vicente, K. J., & Burns, C. M. (2000). There is more to monitoring a nuclear power plant than meets the eye. Human Factors: The Journal of the Human Factors and Ergonomics Society, 42(1), 36-55.

O'Hara, J, & Higgins, J. (2010). *Human-System Interfaces to Automatic Systems: Review Guidance and Technical Basis* (BNL Technical Report 91017-2010). Upton, NY: Brookhaven National Laboratory.

O'Hara, J., & Roth, E. (2005). Operational Concepts, Teamwork, and Technology in Commercial Nuclear Power Stations. In C. Bowers, E. Salas, & F. Jentsch (Eds.) *Creating High-Tech Teams: Practical Guidance on Work Performance and Technology*. Washington, DC: American Psychological Association.

O'Hara, J., Stubler, W., & Higgins, J. (1996). *Hybrid Human-system Interfaces: Human Factors Considerations* (BNL Report J6012-T1-4/96). Upton, New York: Brookhaven National Laboratory.

Oxstrand, J., O'Hara, J., LeBlanc, K., Whaley, A, Joe, J, & Medema, H. (2013). *Development of an Initial Model of Human-Automation Collaboration: Results from a Needs Analysis* (TR-2013/01). Idaho Falls: Idaho National Laboratory.

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, *39* (2), 230-253.

Park, J., Jung, W., & Yang, J. E. (2012). Investigating the effect of communication characteristics on crew performance under the simulated emergency condition of nuclear power plants. Reliability Engineering & System Safety, 101, 1-13.

Pritchett, A. (2001). Reviewing the Role of Cockpit Alerting Systems. *Human Factors and Aerospace Safety*, 1, 5–38.

Pritchett, A., Kim, S. Y., & Feigh, K. (in press a). Measuring Human–Automation Function Allocation. *Journal of Cognitive Engineering and Decision Making*.

Pritchett, A., Kim, S. Y., & Feigh, K. (in press b). Modeling Human–Automation Function Allocation. *Journal of Cognitive Engineering and Decision Making*.

Ranson, D., & Woods, D. (1996b). Animating Computer Agents. In *Proceedings of the 3rd Symposium on Human Interaction with Complex Systems*. NY: IEEE.

Reed, J., Hogg, D., & Hallbert, B. (1995). An Evaluation of an On-Line expert System in Nuclear Process Control. *Proceedings of the Topical Meeting on Computer-Based Human Support Systems: Technology, Methods, and Future*. La Grange Park, Illinois.

Rook, F., & McDonnell, M. (1993). Human Cognition and the Expert System Interface: Mental Models and Inference Explanations. *IEEE Transactions on Systems, Man, and Cybernetics. 23* (6), 1649-1661.

Roth, E. M. (1997). Analysis of decision making in nuclear power plant emergencies: An investigation of aided decision making. In C. E. Zsambok & G. Klein (Eds.), Naturalistic decision making. (pp. 175-182). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Roth, E., Bennett, K., & Woods, D. (1987). Human interaction with an "intelligent" machine. *International Journal of Man-Machine Studies*, *27*, 479-525.

Roth, E., Hanson, M., Hopkins, C., Mancuso, V., & Zacharias, G. (2004). Human in the Loop Evaluation of a Mixed-initiative System for Planning and Control of Multiple UAV Teams. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society.

Roth, E. M., Mumaw, R. J., & Lewis, P. M. (1994). An empirical investigation of operator performance in cognitively demanding simulated emergencies, (NUREG/CR-6208). Washington, DC, U.S. Nuclear Regulatory Commission.

Roth, E. & 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.

Salas, E., Cooke, N. & Rosen, R. (2008). On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50 (3), 540-547.

Sarter, N., & Woods, D. (1992). Pilot Interaction with Cockpit Automation I: Operational Experiences with the Flight Management System. *International Journal of Aviation Psychology* 2(4), 303-321.

Sarter, N., & Woods, D. (1997). Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(4), 553-569.

Skjerve, A., Nihlwing, C., & Nystad, E. (2008). *Lessons Learned from the Extended Teamwork Study* (HWR-867). Halden, Norway: OECD Halden Reactor Project, Institute for Energy Technology.

Soraji, Y., Furuta, K., Kanno, T., Aoyama, H., Inoue, S., Karikawa, D., & Takahashi, M. (2012). Cognitive model of team cooperation in en-route air traffic control. Cognition, Technology & Work, 14(2), 93-105.

Steinberg, M. (2012). Moving from Supervisory Control of Autonomous Systems to Human-Machine Teaming. *4th Annual Human-Agent-Robot Teamwork Workshop*

Stubler, W. & O'Hara, J. (1996b). *Human-system interface design process and review criteria* (BNL Technical Report E2090-T4-5-11/95). Upton, New York: Brookhaven National Laboratory.

Stubler, W. F., O'Hara, J. M., Higgins, J. C., and Kramer, J. (1999). Human-system interface and plant modernization process: Technical basis and review guidance (NUREG/CR-6637). Washington, D.C.: U.S. Nuclear Regulatory Commission.

Suchman, L. (1990). What is human-machine interaction? In W.W. Zachary, S.P. Parasuraman & J.B. Black (Eds.), Cognition, computing, and cooperation.

Sycara, K. & Suktghankar, G. (2006). *Literature Review of Teamwork Models* (CMU-RI-TR-06-50). Pittsburgh, PA: Carnegie Mellon University.

Sycara, K. & Lewis, M. 2004). Integrating Intelligent Agents into Human Teams", In E. Salas and S. M. Fiore (Eds.) *Team Cognition: Process and Performance at the Inter and Intra-Individual Level*. Washington, DC: American Psychological Association. (pp. 203-231)

Terry, P. (1989). A Layman's Guide to Expert Systems. *Power Engineering*, September, 52-55.

Toquam, J. L., Macaulay, J. L., Westra, C. D., Fujita, Y., & Murphy, S. E. (1997). Assessment of nuclear power plant crew performance variability. Team performance assessment and measurement. Mahwah, NJ: Lawrence Erlbaum Associates, 253-287.

Vicente, K. J., Roth, E. M., & Mumaw, R. J. (2001). How do operators monitor a complex, dynamic work domain? The impact of control room technology. International Journal of Human-Computer Studies, 54(6), 831-856. doi:10.1006/ijhc.2001.0463

Warner, N., Letsky, M., and Cowen, M. (2004). Cognitive Model of Team Collaboration: Macro-Cognitive Focus. In Proceedings of the 49th Human Factors and Ergonomics Society Annual Meeting, September 26-30, 2005. Orlando, FL.

Wiener, E. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft* (NASA Report 177528). Moffett Field, CA.: National Aeronautics and Space Administration.

Wickens, C., & Hollands J. (2000). *Engineering Psychology and Human Performance* (3rd ed). Upper Saddle River, NJ: Prentice-Hall Inc.

Woods, D. (2005). Generic Support Requirements for Cognitive Work: Laws that Govern Cognitive Work in Action. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

Woods, D., & Hollnagel, E. (2006). *Joint Cognitive Systems*. Boca Raton, FL: CRC Press, Taylor & Francis.

Woods, D. ,& Sarter, N. (2000). Learning from Automation Surprises and going sour accidents. In N. Sarter & R. Amalberti, (Eds.). *Cognitive Engineering in the Aviation Domain*. Hillside, NJ: Lawrence Erlbaum.

Woods, D., Tittle, J., Feil, M., & Roesler, A. (2004). Envisioning Human–Robot Coordination in Future Operations. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, *34* (2), 210-218.

Woods, D., Tittle, J., Feil, M., & Roesler, A. (2004). Envisioning Human–Robot Coordination in Future Operations. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, 34 (2), 210-218.

Wright, M. (2002) *Effects of Automation in Team Performance and Team Coordination*. PhD Thesis submitted to North Carolina State University.

Wright, M., & Kaber, D. (2005) Effects of Automation of Information-Processing Functions on

Teamwork. *Human Factors,* 47 (1), 50-66.

7.3 Performance Measures

Pina, P. E., Donmez, B., & Cummings, M. L. (2008). *Selecting Metrics to Evaluate Human Supervisory Control Applications* (HAL20008-04). Cambridge, MA: Massachusetts Institute of Technology. Retreived from http://dspace.mit.edu/bitstream/handle/1721.1/46743/HAL2008_04.pdf.

http://www.skybrary.aero/index.php/Situation_Awareness_Rating_Technique_(SART)

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Davis, G., Lumicao, M. N., Popovic, M. & Olmstead, R. (2004). Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human-Computer Interaction, 17*(2), 151-170.

Bisantz, A., M., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics, 28*, 85-97. doi:10.1016/S0169-8141(01)00015-4

Cain, B. (2007). A review of the mental workload literature. In NATO-RTO Task Group TR-HFM-121, *Virtual Environments for Intuitive Human-System Interaction: Human Factors Considerations in the Design, Use, and Evaluation of AMVE-Technology* (RTO Technical Report [Part II], TR-HFM-121-Part-II) (pp. 4-1 – 4-34). Neuilly-sur-Seine Cedex, France: North Atlantic Treaty Organisation (NATO), Research and Technology Organisation. Retrieved from http://ftp.rta.nato.int/public//PubFullText/RTO/TR/RTO-TR-HFM-121-PART-II///TR-HFM-121-Part-II-04.pdf.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies, 58*, 697-718. doi:10.1016/S1071-5819(03)00038-7

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. ##-##). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R., Selcon, S. J., Hardiman, T. D., Croft, D. G. (1998). A comparative analysis of SAGAT and SART for evaluations of situation awareness. Paper presented at the *42nd Annual Meeting of the Human Factors & Ergonomics Society*, Chicago, IL, October, 1998.

Endsley, M. R., Sollenberger, R., & Stein, E. (2000). Situation awareness: A comparison of measures. *Proceedings of the Human Performance, Situation Awareness, and Automation: User centered Design for the New Millennium Conference*, October 2000.

Farmer, E., & Brownson, A. (2003a). *Review of Workload Measurement, Analysis and Interpretation Methods* (CARE-Integra-TRS-130-02-WP2). Brussels, Belgium: Eurocontrol, European Organization for the Safety of Air Navigation, Air Traffic Management Programme.

Farmer, E., & Brownson, A. (2003b). *Measuring Workload in Man-in-the-Loop Simulations: Worked Example* (CARE-Integra-TRS-130-02-WP4). Brussels, Belgium: Eurocontrol, European Organization for the Safety of Air Navigation, Air Traffic Management Programme.

Fink, R., Hill, D., & O'Hara, J. (2004). *Human Factors Guidance for Control Room and Digital Human-System Interface Design and Modification* (EPRI Technical Report 1008122). Palo Alto, CA: Electrical Power Research Institute. Retrieved from http://www.epri.com/abstracts/Pages/ProductAbstract.aspx?ProductId=000000000001008122.

Freedy, A., DeVisser, E., Weltman, G., & Coeyman, N. (2007). Measurement of trust in human-robot collaboration. *Proceedings of the 2007 International Conference on Collaborative Technologies and Systems.*

Guhe, M., Liao, W., Zhu, Z. Ji, Q., Gray, W. D., & Schoelles, M. J. (2005). Non-intrusive measurement of workload in real-time. Paper presented at the *Human Factors and Ergonomics Society Annual Meeting*, Orlando Florida, 2005.

Hogg, D. N., Follesø, K., Strand-Volden, F., & Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarms systems in nuclear power plant control rooms. *Ergonomics, 38*(11), 2394-2413.

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*(1), 53-71.

Khasawneh, M. T., Bowling, S. R., Jiang, X., Gramopadhye, A. K., & Melloy, B. (2003). A model for predicting human trust in automated systems. *Proceedings of the 8th Annual International Conference on Industrial Engineering – Theory, Applications and Practice,* 216-222. November 10-12, 2003, Las Vegas, NV.

Le Blanc, K., Gertman; D., & Boring, R. (2010). Review of methods related to assessing human performance in nuclear power plant control room simulations. Paper presented at the *Seventh American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies* (NPIC&HMIT), Las Vegas, NV, November 7-11, 2010. LaGrange Park, IL: American Nuclear Society.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80. doi:10.1518/hfes.46.1.50_30392

Masden, M., & Gregor, S. (2000). Measuring human-computer trust. *Proceedings of the 11th Australasian Conference on Information Systems.*

Merritt, S. M., Heimbaugh, H., LaChapell, J., & Lee, D. (2013). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors, 55*(3), 520-534. doi:10.1177/0018720812465081

Miller, S. (2001). *Literature Review: Workload Measures* (Document ID: N01-006). Iowa City, IA: University of Iowa, National Advanced Driving Simulator.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37*(11), 1905-1922.

Muir, B. M., & Moray, N. (1996). Trust in automation: Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39*(3), 429-460.

North Atlantic Treaty Organisation, Research and Technology Organisation, Task Group TR-HFM-121 (2007). *Virtual Environments for Intuitive Human-System Interaction: Human Factors Considerations in the Design, Use, and Evaluation of AMVE-Technology* (RTO Technical Report [Part II], TR-HFM-121-Part-II). Retrieved from http://ftp.rta.nato.int/public//PubFullText/RTO/TR/RTO-TR-HFM-121-PART-II///$$TR-HFM-121-Part-II-ALL.pdf.

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

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38*(4), 665-679.

Perry, C. M., Sheik-Nainar, M. A., Segall, N., Ma, R., & Kaber, D. (2008). Effects of physical workload on cognitive task performance and situation awareness. *Theoretical Issues in Ergonomics Science, 9*(2), 95-113. doi:10.1080/14639220600959237

Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J. & Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *The International Journal of Aviation Psychology, 10*(4), 393-410.

Pritchett, A. R., Kim, S. Y., & Feigh, K. M. (2013). Measuring human-automation function allocation. *Journal of Cognitive Engineering and Decision Making,* first published online June 5, 2013. doi:10.1177/1555343413490166

Ryu, K. & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics, 35*, 991-1009. doi:10.1016/j.ergon.2005.04.005.

Salmon, P., Stanton, N., Walker, G., Baber, C., Jenkins, D., McMaster, R., & Young, M. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science, 9*(4), 297-323. doi: 10.1080/14639220701561775

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Baber, C., & McMaster, R. (2008). Representing situation awareness in collaborative systems: A case study in the energy distribution domain. *Ergonomics, 51*(3), 367-384. doi:10.1080/00140130701636512

Seong, Y. & Bisantz, A. M. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics, 38*, 608-625. doi:10.1016/j.ergon.2008.01.007

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced "complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology, 3*(2), 111-122.

Smith, E., Borgvall, J., & Lif, P. (2007). Team and collective performance measurement. In NATO-RTO Task Group TR-HFM-121, *Virtual Environments for Intuitive Human-System Interaction: Human Factors Considerations in the Design, Use, and Evaluation of AMVE-Technology* (RTO Technical Report [Part II], TR-HFM-121-Part-II) (pp. 7-1 – 7-16). Neuilly-sur-Seine Cedex, France: North Atlantic Treaty Organisation (NATO), Research and Technology Organisation. Retrieved from http://ftp.rta.nato.int/public//PubFullText/RTO/TR/RTO-TR-HFM-121-PART-II///TR-HFM-121-Part-II-07.pdf

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., . . . Green, D. (2006). Distributed situation awareness in dynamic systems: Theoretical development and application of an ergonomics methodology. *Ergonomics, 49*(12-13), 1288-1311. doi: 10.1080/00140130600612762

Wickens, C. D. (2001). Workload and situation awareness. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue.* (pp. 443-450). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Feigh, K. & Pritchett, A. (in press). Requirements for Effective Function Allocation: A Critical Review. *Journal of Cognitive Engineering and Decision Making*.

Ha, J., Seong, P., Lee, M., & Hong, J. (2007). Development of Human Performance Measures for Human Factors Validation in the Advanced MCR of APR-1400. *IEEE Transactions on Nuclear Science, Vol. 54, No. 6.*

Hsu, S., Wu, T., & Lee, J. (2006). Development of Human Performance Evaluation Battery for Integrated System Validation of the HSI for an Advanced Control Room. In proceedings of *NPIC&HMIT 2006.* Albuquerque, NM: American Nuclear Society.

Hugo, J., Forester, J., Gertman, D., Joe, J., Medema, H., Persensky, J., & Whaley, A. (2013). Draft Function Allocation Framework and Preliminary Technical Basis for Advanced SMR Concept of Operations (SMR/ICHMI/INL/TR---2013/02). Idaho Falls, ID: Idaho National Laboratory.

Lang, A., Roth, E., Bladh, K., Hine, R. (2002). Using a Benchmark-Referenced Approach for Validating a Power Plant Control Room: Results of the Baseline Study. In *Proceeding of the 46th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors and Ergonomics Society.

Lee, M., Hong, J., Suh, J., Lee, S., & Hwang, D. (2009). Development of Human Factors Validation System for the Advanced Control Room of APR1400. *Journal of Nuclear Science and Technology, Vol. 46, No. 1, p. 90-101.*

Malcolm, S., Holford, K. & Gillard, D. (2000). HF Verification and Validation Activities: Simulator Based Operational Trials. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 2000, 3-181 to 3-184.

Norros, L., & Savioja, P. (2007). Towards a theory and method for usability evaluation of complex human-technology systems. *Activities,* 4 (2), 143-150.

Norros, L., & Savioja, P. (2004). *Usability Evaluation of Complex Systems (A Literature Review).* Radiation and Nuclear Safety Authority*: STUK-YTO-TR204. Helsinki 2004.*

Norros, L., Savioja, P., & Salo, L. (2009). Approach to Integrated System Validation of NPP Control Rooms. In Proceedings of *Sixth American Nuclear Society International Topical Meeting on*

Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies. LaGrange Park, IL: American Nuclear Society.

Ockerman, J. & Roberts, D. (2010). Opportunities for Human Factors Measures in Military Operational Test and Evaluation. In *Proceeding of the Human factors and Ergonomics Society 54th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society.

O'Hara, J., Higgins, J., Fleger, S. & Pieringer. P. (2012). *Human Factors Engineering Program Review Model* (NUREG-0711, Rev. 3). Washington, D.C.: U.S. Nuclear Regulatory Commission.

O'Hara, J., Higgins, J. & Pena, M. (2012). *Human-Performance Issues Related to the Design and Operation of Small Modular Reactors* (NUREG/CR-7126). Washington, D.C.: U. S. Nuclear Regulatory Commission

Oxstrand, J., O'Hara, J., LeBlanc, K., Whaley, A. M., Joe, J., & Medema, H. (2013a). *Development of an Initial Model of Human-Automation Collaboration—Results from a Needs Analysis* (SMR/ICHMI/INL/TR-2013/01). Idaho Falls, ID: Idaho National Laboratory.

Oxstrand, J., Joe, J., LeBlanc, K., Medema, H., Whaley, A., & O'Hara, J. (2013b). *Framework for Human-Automation Collaboration: Project Status Report* (SMR/ICHMI/INL/TR-2013/02). Idaho Falls, ID: Idaho National Laboratory.

Pritchett, A., Kim, S. Y., & Feigh, K. (in press a). Measuring Human–Automation Function Allocation. *Journal of Cognitive Engineering and Decision Making*.

Pritchett, A., Kim, S. Y., & Feigh, K. (in press b). Modeling Human–Automation Function Allocation. *Journal of Cognitive Engineering and Decision Making*.

Savioja, P., Aaltonen, I., Karvonen, H., Koskinen, H., Laarni, J. Liinasuo, M. & Norros, L. (2012). Systems Usability Concerns in Hybrid Control Rooms. In *Proceedings of the Eighth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies* (NPIC & HMIT). La Grange, IL: American Nuclear Society.

Savioja, P. & Norros, L. (in press). Systems Usability Framework for Evaluating Tools in Safety–Critical Work. *Cognition, Technology, and Work*.

Shin, Y.-C., Moon, H-K. & Jong-Hyun Kim, J.H. (2006). Human Factors Engineering Verification and Validation for APR1400 Computerized Control Room. In proceedings of *Fifth American Nuclear Society International Topical on Nuclear Plant Instrumentation, Controls and Human-machine Interface Technologies (NPIC & HMIT).* Columbus, Ohio: American Nuclear Society.

S.G. Hart, & L.E. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research", *Human Mental Workload,* Elsevier, Amsterdam, pp.139-183 (1988).

20. S.G. Hart, "NASA-Task Load Index (NASA-TLX); 20 years later", *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting,* Santa Monica, CA: Human Factors& Ergonomics Society. (2006).

Hogg, D. N., Folleso, K., Strand-Volden, F., and Torralba, B. "Development of a Situation Awareness Measure to Evaluate Advanced Alarm Systems in Nuclear Power Plant Control Rooms," *Ergonomics* (11), 1995, pp. 394-413.

M. Endsley and D. Garland, *Situation Awareness, Analysis and Measurement,* Lawrence Erlbaum Associates, Publishers, Mahway, New Jersey, 2000.

Sebok, A. (2000), 'Team Performance in Process Control: Influence of Interface Design and Staffing Levels', *Ergonomics*, vol. 43 (8), pp. 1210-1236.

Sheue-Ling Hwang; Fei-Hui Huang; Ying-Lien Lee; Tzu-Chung Yenn; Yuan-Chang Yu; Chong-Cheng Hsu; Hao-Wu Huang (2007) "Experimental evaluation of human-system interaction on alarm design" *Nuclear Engineering and Design*, v 237, n 3, Feb. 2007, 308-15

Hwang, Sheue-Ling; Yau, Yi-Jan; Lin, Yu-Ting; Chen, Jun-Hao; Huang, Tsun-Hung; Yenn, Tzu-Chung; Hsu, Chong-Cheng (2008) "Predicting work performance in nuclear power plants". *Safety Science*, v 46, n 7, August, 2008, p 1115-1124

Zhang et al., 2007 L. Zhang, X. Hec, L.C. Daia and X.R. Huang, The simulator experimental study on the operator reliability of Qinshan nuclear power plant, *Reliability Engineering and System Safety* **92** (2007), pp. 252–259.

Carvalho, P.V.R.; dos Santos, I.L.; Gomes, J.O.; Borges, M.R.S.; Guerlain, S. (2008) "Human factors approach for evaluation and redesign of human-system interfaces of a nuclear power plant simulator". *Displays*, v 29, n 3, July 2008, 273-84.

Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications*, 163-181. Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. R. (1997). Level of automation: Integrating humans and automated systems. In *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting*, Santa Monica, CA: Human Factors and Ergonomics Society, 200-204.

Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42,* 462-492.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381-394.

Jou, Y.-T., Yenn, T.-C., Lin, C. J., Yang, C.-W., & Chiang, C.-C. (2009). Evaluation of operators' mental workload of human-system interface automation in the advanced nuclear power plants. *Nuclear Engineering and Design, 239,* 2537-2542. doi:10.1016/j.nucengdes.2009.06.023

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness, and workload in a dynamic control task. *Theoretical Issues in Ergonomic Science, 5*(2), 113-153. doi: 10.1080/1463922021000054335

Lin, C. J., Yenn, T.-C., & Yang, C.-W. (2009). Evaluation of operators' performance for automation design in the fully digital control room of nuclear power plants. *Human Factors and Ergonomics in Manufacturing & Service Industries, 20*(1), 10-23. doi:10.1002/hfm.20168

Lin, C. J., Yenn, T.-C., & Yang, C.-W. (2010a). Automation design in advanced control rooms of the modernized nuclear power plants. *Safety Science, 48*, 63-71. doi:10.1016/j.ssci.2009.05.005

Lin, C. J., Yenn, T.-C., & Yang, C.-W. (2010b). Optimizing human-system interface automation design based on a skill-rule-knowledge framework. *Nuclear Engineering and Design, 240*, 1897-1905. doi:10.1016/j.nucengdes.2010.03.026

van de Merwe, K., Oprins, E., Eriksson, F., & van der Plaat, A. (2012). The influence of automation support on performance, workload, and situation awareness of air traffic controllers. *The International Journal of Aviation Psychology, 22*(2), 120-143. doi:10.1080/10508414.2012.663241

Ha, J. S., Seong, P. H., Lee, M. S., & Hong, J. H. (2007). Development of human performance measures for human factors validation in the advanced MCR of APR-1400. *Nuclear Science, IEEE Transactions on*, *54*(6), 2687-2700.

7.1 Adaptive Automation

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological psychology*, *42*(3), 249-268.

Calhoun, G. L., Ruff, H. A., Spriggs, S., & Murray, C. (2012, September). Tailored Performance-based Adaptive Levels of Automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 413-417). SAGE Publications.

Christensen, J. C., Estepp, J. R., Wilson, G. F., & Russell, C. A. (2012). The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage*, *59*(1), 57-63.

Cosenzo, K., Parasuraman, R., Pillalamarri, K., & Feng, T. (2009). *The Effect of Appropriately and Inappropriately Applied Automation for the Control of Unmanned Systems on Operator Performance* (No. ARL-TR-4933). ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE.

Dorneich, M. C., Verves, P. M., Withlow, S.D., Mathan, S., Carciofini, J., and Reusser, T. (2006). Neuro-Physiologically-Driven Adaptive Automation to Improve Decision Making Under stress. Proceedings of the Human Factors and Ergonomic Society Conference 2006, San Francisco, CA.

Freeman, F. G., Mikulka, P. J., Prinzel, L. J., & Scerbo, M. W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological Psychology*, *50*(1), 61-76.

Freeman, F. G., Mikulka, P. J., Scerbo, M. W., and Scott, L. (2004). An evaluation of an adaptive automation system using a cognitive vigilance task. Biological Psychology. Vol. 67. Pp. 283-297. doi: 10.1016/j.biopsycho.2004.01.002.

Freeman, F. G., Mikulka, P. J., Scerbo, M. W., Prinzel, L. J., & Clouatre, K. (2000). Evaluation of a psychophysiologically controlled adaptive automation system, using performance on a tracking task. *Applied Psychophysiology and Biofeedback*, *25*(2), 103-115.

Hockey, G. R. J., Nickel, P., Roberts, A. C., & Roberts, M. H. (2009). Sensitivity of candidate markers of psychophysiological strain to cyclical changes in manual control load during simulated process control. *Applied Ergonomics*, *40*(6), 1011-1018.

Inagaki, T. (2003). Adaptive Automation: Sharing and Trading of Control. Chapter 8 of the Handbook of Cognitive Task Design. Pp. 147-169. (Erik Hollnagel Ed.) LEA

Inagaki, T., Furukawa, H., & Itoh, M. (2005). Human interaction with adaptive automation: Strategies for trading of control under possibility of over-trust and complacency. *Proc. HCI-AugCog International, CD-ROM*, *10*.

Kaber, D. B., & Riley, J. M. (1999). Adaptive automation of a dynamic control task based on secondary task workload measurement. *International journal of cognitive ergonomics*, *3*(3), 169-187.

Kaber, D. B., Perry, C. M., Segall, N., McClernon, C. K., & Prinzel III, L. J. (2006). Situation awareness implications of adaptive automation for information processing in an air traffic control-related task. *International Journal of Industrial Ergonomics*, *36*(5), 447-462.

Kaber, D. B., Riley, J. M., Tan, K. W., & Endsley, M. R. (2001). On the design of adaptive automation for complex systems. *International Journal of Cognitive Ergonomics*, *5*(1), 37-57

Kaber, D. B., Wright, M. C., & Sheik-Nainar, M. A. (2006). Investigation of multi-modal interface features for adaptive automation of a human–robot system. *International journal of human-computer studies*, *64*(6), 527-540.

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47*(4), 730-741.

Lagu, A. V., Landry, S. J., and Yoo, H-S. (2013). Adaptive Function Allocation Stabilization and a Comparison of Trigger Types and Adaptation Strategies. *International Journal of Industrial Ergonomics*, pp. 1-11. doi: 10.1016/j.ergon.2013.02.006.

Langan-Fox, J., Canty, J. M., & Sankey, M. J. (2009). Human–automation teams and adaptable control for future air traffic management. *International Journal of Industrial Ergonomics*, *39*(5), 894-903.

Miller, C. A. (2005, July). Trust in adaptive automation: The role of etiquette in tuning trust via analogic and affective methods. In *Proceedings of the 1st International Conference on Augmented Cognition, Las Vegas, NV*.

Miller, C. A., & Parasuraman, R. (2003, October). Beyond levels of automation: An architecture for more flexible human-automation collaboration. In*Proceedings of the Human Factors and Ergonomics Society Annual Meeting*(Vol. 47, No. 1, pp. 182-186). SAGE Publications.

Miller, C. A., & Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control.*Human Factors: The Journal of the Human Factors and Ergonomics Society*,*49*(1), 57-75.

Miller, C., Funk, H., Wu, P., Goldman, R., Meisner, J., & Chapman, M. (2005, September). The Playbook™ Approach to Adaptive Automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, No. 1, pp. 15-19). SAGE Publications.

Moray, N., Inagaki, T., and Itoh, M. (2000). Adaptive Automation, Trust, and Self-Confidence in Fault Management of Time-Critical Tasks. Journal of Experimental Psychology: Applied 2000, Vol. 6. No. 1, pp. 44-58.

Morrison, J. G. (1993). *The Adaptive Function allocation for Intelligent Cockpits (AFAIC) Program: Interim Research and Guidelines for the Application of Adaptive Automation* (No. NAWCADWAR-93031-60). NAVAL AIR WARFARE CENTER AIRCRAFT DIV WARMINSTER PA.

Parasuraman, R., Bahri, T., Deaton, J. E., Morrison, J. G., & Barnes, M. (1992). *Theory and design of adaptive automation in aviation systems*. CATHOLIC UNIV OF AMERICA WASHINGTON DC COGNITIVE SCIENCE LAB.

Parasuraman, R., Barnes, M., and Cosenzo, K. (2007). Adaptive Automation for Human-Robot Teaming in Future Command and Control Systems. *The International C2 Journal*. Vol 1. No 2. Page 43-68.

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology*,*21*(2), 270.

Prinzel III, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., and Pope, A., T. (2003). Effects of a Psychophysiological System of Adaptive Automation on Performance, Workload, and the Event-Related Potential P300 Component. *Human Factors: The Journal of the Human Factors and Ergonomic Society*. doi: 10.1518/hfes.45.4.601.27092.

Prinzel, L. J., Parasuraman, R., Pope, A. T., Freeman, F. G., Scerbo, M. W., & Mikulka, P. J. (2003). Three experiments examining the use of electroencephalogram, event-related potentials, and heart-rate variability for real-time human-centered adaptive automation design. Langley Research Center, Hampton, VA 23681-2199. *National Aeronautics and Space Administration*.

Sauer, J., Kao, C. S., Wastell, D., & Nickel, P. (2011). Explicit control of adaptive automation under different levels of environmental stress. *Ergonomics*,*54*(8), 755-766.

Scallen, S. F., Hancock, P. A., & Duley, J. A. (1995). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Applied ergonomics*, *26*(6), 397-403.

Ting, C. H., Mahfouf, M., Nassef, A., Linkens, D. A., Panoutsos, G., Nickel, P., ... & Hockey, G. (2010). Real-time adaptive automation system based on identification of operator functional state in simulated process control operations. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, *40*(2), 251-262.

Wilson, G. F., Lambert, J. D., & Russell, C. A. (2000, July). Performance enhancement with real-time physiologically controlled adaptive aiding. In*Proceedings of the human factors and ergonomics society annual meeting* (Vol. 44, No. 13, pp. 61-64). SAGE Publications.

Sauer, J., Nickel, P., & Wastell, D. (2013). Designing automation for complex work environments under different levels of stress. *Applied ergonomics*, *44*(1), 119-127.