Some Challenges in the Design of Human-Automation Interaction for Safety-Critical Systems

Michael Feary [a] *and Emilie Roth* [b]

[a]*National Aeronautics and Space Administration*
Ames Research Center
Moffett Field, CA, 94035, USA

[b]*Roth Cognitive Engineering*
2 Oliver Court
Menlo Park, CA, 94025

ABSTRACT

Increasing amounts of automation are being introduced to safety-critical domains. While the introduction of automation has led to an overall increase in reliability and improved safety, it has also introduced a class of failure modes, and new challenges in risk assessment for the new systems, particularly in the assessment of rare events resulting from complex inter-related factors. Designing successful human-automation systems is challenging, and the challenges go beyond good interface development (e.g., Roth, Malin, & Schreckenghost 1997; Christoffersen & Woods, 2002). Human-automation design is particularly challenging when the underlying automation technology generates behavior that is difficult for the user to anticipate or understand. These challenges have been recognized in several safety-critical domains, and have resulted in increased efforts to develop training, procedures, regulations and guidance material (CAST, 2008, IAEA, 2001, FAA, 2013, ICAO, 2012). This paper points to the continuing need for new methods to describe and characterize the operational environment within which new automation concepts are being presented. We will describe challenges to the successful development and evaluation of human–automation systems in safety-critical domains, and describe some approaches that could be used to address these challenges. We will draw from experience with the aviation, spaceflight and nuclear power domains.

Keywords: Human-Automation Interaction, Systems Engineering, Risk Assessment

INTRODUCTION

Increasing amounts of automation are being introduced to safety-critical domains. Designing successful human-automation systems is challenging, and the challenges go beyond good interface development (e.g., Roth, Malin, & Schreckenghost 1997; Christoffersen & Woods, 2002). These challenges have been recognized in several safety-critical domains, and have resulted in increased efforts to develop training, procedures, regulations and guidance material (CAST, 2008, IAEA, 2001, FAA, 2013, ICAO, 2012). Evaluation of human-automation interaction remains an area of particular challenge. While there are many existing methods for evaluating human–automation interaction, many of these methods have limitations when attempting evaluation early in the design process, and fewer still are adequate for use in safety-critical domains (Leveson, 2012). There remains a continuing need for new methods to

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2097-8

describe and characterize the operational environment within which new automation concepts are being presented to enable early identification of factors that are likely to contribute to human performance problems and safety risks.

The incident and accident trend in human-automation interaction in safety-critical domains points to an inability to predict rare events resulting from complex inter-related factors. In particular, there is a need to place greater emphasis on searching vs. screening for more comprehensive identification of sources of risk to avoid prematurely screening out important sources of risk (Siu et al., 2013). This is particularly true when considering human-automation interaction, where risk is likely to be greatest when multiple factors combine to create unanticipated circumstances (e.g., interacting faults that cause the automation to behave in an unpredicted manner) that challenge the ability of people to appropriately recognize and respond to the situation. In this paper we will describe challenges to the successful development and evaluation of human–automation systems in safety-critical domains, and describe some approaches that could be used to address these challenges. We will draw from experience with the aviation, spaceflight and nuclear power domains.

A Rise in "Normal Accidents"

As the systems become safer overall, there is a trend towards a relative increase in the rate of "normal accidents" (Perrow, 1984). These accidents are predicated on interactive complexity and tight coupling in dynamic systems, rather than direct links between any single component failure and an accident. As the environment for which these systems are being designed is complex, the design and evaluation processes will need to improve. In fact, recent trends in aviation incidents and accidents have pointed to misspecification in design as a large and increasing contributor to risk (Sarsfield et al, 2000; FAA, 2013). These trends are highlighted by a few examples from the aviation and nuclear power industries.

In the nuclear power industry, the recent Fukushima Dai-ichi accident highlights the importance of more accurately modeling the situational challenges that arise in real-world accidents (e.g., unavailability or misleading sensor indications; lack of relevant procedural guidance) and the individual, team, and organizational decision-making processes that are likely to influence performance under time-pressured, high-stress conditions.

The Fukushima accident occurred on March 11, 2011, as a result of a large earthquake and tsunami that caused a total loss of AC and DC power across multiple units at the plant (Investigation Committee Interim Report, 2011). This resulted in a loss of all ability to monitor or control key safety cooling functions. As a consequence, operators were faced with a very challenging situation where the preplanned procedures were not applicable, and operators had to fall back on knowledge-based behavior to diagnose plant state and develop and execute ad hoc plans of action to mitigate the accident. They came up with a variety of creative solutions including use of car batteries to drive sensors to get information about plant state and controllers in order to re-establish safety systems.

The aviation industry also illustrates a need for more accurately modeling the situational complexity. A report of accidents involving the worldwide commercial jet fleet showed that of 75 fatal accidents between 2003 and 2012, the majority of the accidents occurred with no major component failures, and the systems on the aircraft working as designed (Boeing, 2013).

In 2013, the FAA and U.S. aviation industry released a report addressing safety issues with aircraft flight deck automation. The report mentions that "the highly integrated nature of current flight decks, and additional 'add-on' features and retrofits in older aircraft, have increased flightcrew knowledge requirements and introduced complexity that sometimes results in pilot confusion and errors in flight deck operations." An example of the complexities of these interactions has appeared as a contributing factor in several recent aircraft accidents.

An example of interaction complexity within a single aircraft is highlighted by an accident that occurred February 25, 2009. A Boeing 737-800 on landing approach into Amsterdam, was flying with a failed radar altimeter on the Captain's (left) side. To account for this risk, this type of aircraft had a second radar altimeter installed, feeding information to the First Officer's (right) instruments. The crew was aware of the failure of the radar altimeter, and the impact on manual flight of the aircraft was not impacted. However the crew did not recognize the implications of the erroneous data that was being sent to the aircraft's autoflight system, (specifically the autothrottles), which led to less thrust than required for the approach, an eventual stall of the aircraft, and a failed recovery. The accident investigation board concluded that one of the contributing factors was "convergence of circumstances", referring to a spe-

cific set of circumstances occurring which interacted with each other to lead to the accident. The circumstances included a First Officer in the latter stages of operational training, and a lack of salient alerts. (The autothrottle annunciation was shared for two different aircraft behaviors; one of the behaviors is what the crew would have expected to see, but the other was in effect during the accident scenario).

The sensor failure was well known at the airline (and the manufacturer), but not well understood, and (according to the accident report) not replicable on the ground. Boeing had written procedures addressing the case of a failed radar altimeter prior to the flight, which allowed the flight but recommended that "the associated (right or left) autopilot or autothrottle should not be used for approach or landing" (Boeing, 2004) However, Boeing considered that it was not a safety hazard and therefore did not address the risk via procedures, training or increased alerts.

An example illustrating interacting system and situational complexity in the aviation industry occurred on July 1, 2002 near Lake Constance, Germany. A Boeing 757 and Tupolev T-154 collided over the towns of Üeberlingen and Owengin due to a combination of ambiguities in the procedures for the automated Traffic Collision Avoidance System (TCAS), and non-normal Air Traffic Control procedures. (BFU, 2002)**.**

The Üeberlingen accident showed the importance of operational context with human–automation interaction (i.e., TCAS). In the accident, the flight crew was given instructions by the TCAS system, followed almost immediately by different instructions from a human Air Traffic Controller. Although there were procedures written at the time for the flight crew, the nature of the sequence of events and operational experience of the crews led the crews to follow what turned out to be erroneous information from the Air Traffic Controller. Ten years after the accident, studies have shown that, depending on the context, flight crews may still disregard TCAS instructions in favor of human controller instructions (Pritchett et al., 2012). In some cases the behavior of the flight crews was appropriate given the context, and it is unclear how often TCAS instructions are appropriately disregarded in actual operations.

These examples reinforce the point that while the systems were performing as designed, actual accidents often involve a confluence of interacting faults resulting in situations that have not been previously anticipated, placing a premium on the ingenuity and adaptability of the humans on the scene. This is a point that has repeatedly been made in the human factors and sociotechnical systems communities (e.g., Perrow, 1984; Leveson, 2012; Hollnagel et al., 2011). The accidents also illustrate a need for a greater appreciation of people as a source of resilience and recovery. Too often humans are treated as the "weak link" in systems. In fact there is growing evidence that people are a source of system resilience because of their ability to adapt creatively in response to unforeseen circumstances (Hollnagel, Woods & Leveson, 2006; Reason, 2008; Paries, 2011). As automated systems are being designed to handle increasingly complex situations, the design and risk assessment processes will have to improve to ensure that these joint systems are capable of sufficient flexibility, resilience, and the ability to recover.

These accident examples also highlight the challenges in being able to predict risk. From a human performance perspective, there is a need to ensure that the types of situations that arise in real accidents, and impose challenges to human performance (e.g., multiple interacting faults; loss or degradation of sensor information; situations that go beyond "pre-planned" procedures or create goal-conflicts) are explicitly searched for and considered as part of human reliability and risk analyses. Methods for systematic search of plausible complicating scenarios exist that can provide a foundation to build upon (e.g., Whaley et al./NUREG-2114, 2012). As a counter-example, joint human-automation systems in safety-critical domains where the environment has been well described, controlled and anticipated have shown considerable success. Examples include driverless trains, use of robots in the home, and process control. (UITP, 2013; Heber, 2013)

DESIGN AND SAFETY RISK ASSESSMENT FOR EVALUATION OF HUMAN-AUTOMATION

The successful implementation of Safety Management Systems and risk assessment methods has led to a significant improvement in safety in many safety-critical domains (Bayuk, 2008). Safety-critical domains, including the aviation, nuclear regulatory, and manned spaceflight safety communities use variations of these SMS processes to track,

mitigate and control risks, and these systems have contributed to significant increases in safety. Modern Safety Management Systems typically consist of 4 major pieces: Safety Policy, Safety Assurance, Safety Risk Management (sometimes referred to as Continuous Risk Management or CRM), and Safety Promotion. This paper will focus on the Safety Risk Management (SRM) component of Safety Management Systems. SRM started when engineering systems (e.g., mechanical, electrical, etc.), became reliable enough to devote effort in the design process to predicting which failures were most likely and try to mitigate those risks. SRM/CRM typically consists of 5 stages:

- Identification of Hazards, which may include a description of the mission context and system analysis
- Analysis of Risks, generated from the intersection of the hazards and the expected operation of the system
- Risk Assessment, which assigns a likelihood and magnitude of severity to the identified risks
- Planning for disposition and tracking of the risks
- Control of the risks through design, procedures, training, maintenance, etc. mechanisms

In some domains the entire SRM/CRM process may be referred to as Probabilistic Risk Assessment. For our purposes, we will refer to PRA as the first three stages (Hazard Identification, Risk Analysis, and Risk Assessment). PRA is the most popular method in use for assessing human - automation risks in the aviation, nuclear power and manned spaceflight domains (DOE, 2013, NASA, 2011).

The trend of incidents and accidents occurring due to complex inter-related factors, with independent system components working as designed, points to a need for methods to improve the characterization of the operating environment and context. It is important to note that a significant difference across different processes is the starting point. It is only recently that an explicit description of the system has been recognized as an important first step. SRM/CRM/PRA processes start with identification of risks immediately without including the crucial step of explicitly defining objectives and constraints. We will consider the explicit definition of strategic goals, objectives and constraints as a first step, and focus on methods expanding the use of contextual information.

An example of the context information in the aviation community (FAA Order 8040.4A) includes:

- definition and documentation of the scope of the system objectives
- development of a safety risk acceptance plan, including safety risk acceptance criteria, designation of authority for risk decisions, and assignment of decision –makers
- description and model of the system and operation in sufficient detail for safety analysts to understand and identify the hazards that can exist in the system, including potential interaction with other systems
- examination of the system as a sub-component in the context of a larger system
- appropriate consideration of :
    - function and purpose
    - operating environment
    - outline of system’s process

The next step is to identify the risks associated with the identified hazards. The risks associated with hazards in many safety-critical domains could be too large to reasonably assess, and therefore examples based on historic data or analyses, active study, experiment or investigation into relevant hazards, or prognostic analyses based on expected changes to the operating environment. An example of risk identification in commercial aviation is to examine incident and accident data, and where possible use operational data collected from aircraft or air traffic system.to inform the team of experts estimating risk. Examples include focused studies, such as modeling, simulations and empirical studies. There are also some methods to predictively identify risks based on expected changes in the future operating environment. An example of this in commercial aviation is the Future Aviation Safety Team (FAST, 2012).

The results of these analyses are used to inform assignments of likelihood and magnitude of severity ratings to each of the identified risks. While the use of Probabilistic Risk Analysis has been a significant contributor to increases in safety, there have been many critiques of the PRA approach since it’s first use in 1975 (WASH-1400/NUREG75/014, Hubbard, 2009). Mauro and Barshi (2009) argue that while the cells in the risk matrix shown in Table 1 (and table 3) appear to be equidistant from each other, in fact some catastrophic events may have non-linear risk relationships. Examples include destruction of an entire ecosystem, epidemic level health events, and failures of companies due to accidents.

Minimal 5	Minor 4	Major 3	Hazardous 2	Catastrophic 1
Negligible safety effect	Physical discomfort to personsSlight damage to aircraft/ vehicle	Physical distress or injuries to persons Substantial damage to aircraft/vehicle	Multiple serious injuries; fatal injury to a relatively small number of persons (one or two); or a hull loss without fatalities	Multiple fatalities (or fatality to all on board) usually with the loss of aircraft/ vehicle

Table 1. Example severity magnitudes from commercial aviation in the U.S. (FAA, 2012)

Likelihood estimates for human contributions to risk often suffer from bias in analysis. An example is optimism bias (Helweg-Larsen et al., 2001). These biases can be very difficult to control without accurate data about the nature of the source variability and representative data from actual operations. Sherry et al. (2014) additionally make the case that the calculations used to produce the risk probabilities are additive of individual component risks, instead of the level of the worst outcome for any of the interacting system components. An example of likelihoods of events in the U.S. National Airspace System are shown in Table 2.

Frequent A	Probable B	Remote C	Extremely Remote D	Extremely Improbable E
Probability of occurrence per operation/operational hour is equal to or greater than $1x10^{-3}$	Probability of occurrence per operation/ operational hour is less than or equal $1x10^{-3}$ but equal to or greater than $1x10^{-5}$	Probability of occurrence per operation/operational hour is less than or equal $1x10^{-5}$ but equal to or greater than $1x10^{-7}$	Probability of occurrence per operation/ operational hour is less than or equal $1x10^{-7}$ but equal to or greater than $1x10^{-9}$	Probability of occurrence per operation/operational hour is less than $1x10^{-9}$

Table 2 Example likelihood frequencies for different events in the U.S. National Airspace Systems (Falteisek, 2010)

An overall critique of PRA is that the frequency and severity of occurrence of any event has a large amount of variance, based on the availability of good data, and the poor resolution of the matrix (i.e. making an estimation of a complex set of variables into simply defined cells)(Cox, 2008). An example risk matrix is shown in Table 3.

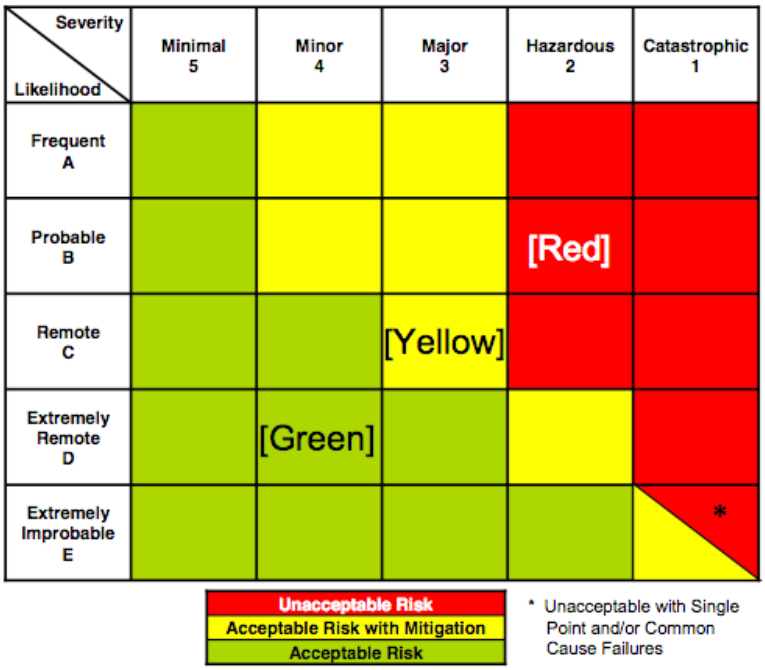

Table 3. Example Risk Matrix

Examples of the types of probabilistic risk analyses conducted for nuclear power plants include: Facility operating States Analysis (FOS), Initiating Events Analysis (IE), Event Sequence Analysis (ES), Success Criteria Development (SC), High Winds PRA (W), External Flooding PRA (XF), Systems Analysis (SA) and Human Reliability Analysis (HR)

It is the last two of the listed examples, System Analysis (SA) and Human Reliability Analysis (HRA) that are the most relevant for the characterization of context in human-automation systems. These methods need improved information about the context and operating environment human – automation system.

Risk Assessment in Automation/Software System Development

Software requirements specification is part of the iterative software design process, which is similar to the SRM process. It consists of a complete description of the behavior of the software, including characteristics of the intended users, and potentially use cases. In SRM/CRM/PRA terms, the requirements specification step focuses on understanding and describing the system in enough detail to identify any associated risks

The problem of context description is recognized in software design. Curtis et al. (1988) pointed out that “writing code isn’t the problem, understanding the problem is the problem”. Sherry and Ward (1995) found that the majority of aviation software code was dedicated to determining the correct automation situation and executing the correct automation behavior, which requires domain expertise.

The mission/context definition step is the most vulnerable to the lack of appropriate methods for the evaluation of risk for human-automation interaction because the mis-specification of context can result in gross errors in the risks associated with inappropriate automation support (Lim and Long, 1994). Despite the potential impact of mis-specification there are a lack of methods to help with the specification of context. Looking at the upper left corner of Figure 1 shows a lack of methods for description and context taken from a recent NASA Software Safety Guidebook.

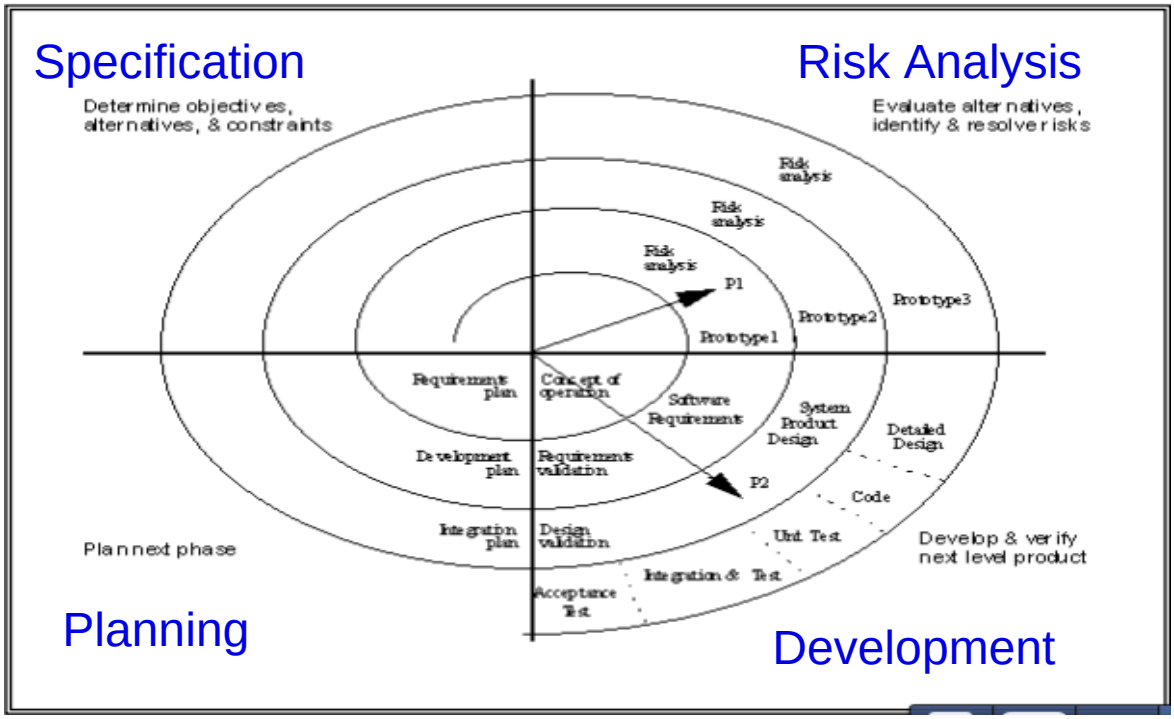

Figure 1. Iterative Design NASA Software Safety Guidebook-8719.13 (NASA, 2004)

Just as in evaluation, design of automation works well in well-understood or controlled environments (e.g. autonomous trains at airports, robots in homes, etc.)

Support for Human Reliability Analysis

Human Reliability Analysis (HRA) methods are used to model the human contribution to the overall reliability of complex systems (Kirwan, 1994). This includes generating estimates of the likelihood and severity of human errors as well as modeling and quantifying the likelihood that a person will take recovery action that will reduce the likelihood or mitigate the consequences of an accident. Typical methods associated with this step include task analysis methods, (including hierarchical and cognitive task analysis), as well as work domain description methods. Additional methods include Human-in-the-Loop (HITL) simulation and detailed cognitive modeling, however, these methods may have limited applicability, either because they require too many resources, skills, or expertise, or are not usable in the allotted timeframe.

While HRA provides a useful method for identifying the human contribution to risk, it has well-recognized limitations. In particular traditional HRA methods are best suited for modeling performance in well-understood and practiced routine tasks, while experience with actual accident conditions makes clear that the greatest source of risk is likely to come from non-routine situations that challenge higher-level cognitive and collaborative performance (e.g., situation awareness, situation assessment, planning and decision making, communication and collaboration). To more accurately gauge risk, HRA/PRA methods need to more realistically model contextual, complicating situational factors that may arise in actual events, such as missing or misleading sensor indications, conflicting guidance, or a mismatch to expectations based on training or prior experience, that can challenge cognitive and collaborative performance (Patterson, Roth and Woods, 2010). There is growing consensus on this point in the risk assessment community and several research and development thrusts have been initiated to improve HRA methods along this front (e.g., Bye, Lois, Dang et al., 2011; Lois, Dang, Forester et al., 2009; Roth, Mosleh, Chang, et al., 2012; Chang, Bley, Criscione et al., 2013; Whaley, Xing, Boring et. al., 2012).

As highlighted earlier, many recent accidents are attributed to system degradation of multiple components or due to an unforeseen set of circumstances, however many of the frequency and severity numbers are predicated on the frequency of independent, low probability events. Some factors that are inadequately used in risk assessment of human–automation systems include examination of the impact of:

- Degraded cognition, including detecting, sensemaking, planning, executing, deciding
- Task limitations, including overconstrained tasks (can't do it all), doublebind situations (dilemmas), goal conflicts (impossible task), or workload (time pressure)
- Individual performance differences and the impact of physiological stressors on performance, including stress, fatigue, illness, medication/alcohol, emotion
- Coordination, and communication issues, including weak leadership (in-fighting), unreliable/poor communications, decreased access to team members (remote teams), interdependencies among roles (coordination bottlenecks)
- Leading indicators that are usually relied upon in diagnosis are unavailable (late change in plan, interruptions, unexpected change of role)

Task analysis, decomposition and modeling methods can help with this need. Unfortunately, existing methods and techniques are time and labor intensive and the design community does not perceive high value in their results. Part of this perception on the part of designers is that while an implicit task decomposition is an intrinsic part of the design process, there is no inherent part of the design process that requires an explicit task description. Additionally, existing task analysis techniques may not accurately represent tasks that are complex, context-dependent, non-linear or have alternatives available (Rasmussen, 1986; Hoffman et al, 2002; Feltovich et al, 2004). The methods need to be more integrated and less invasive in the design process, and produce more applicable results. It should also be noted that there is a requirement for presentation of the contextual information in the right form and scale to support the users of the contextual information who may have limited human factors or cognitive science expertise.

Now that some of the missing factors that need to be considered have been defined, let's examine some of the potential methods available for including the evaluation of these factors in the risk assessment process.

Work Analysis Methods

Work domain modeling refers to methods that provide a functional description of the objectives, means, and constraints of a work domain that can be used to define the work goals, activities and associated cognitive and collaborative challenges. A variety of functional analysis methods that have been developed to characterize work demands. These methods typically generate function goal–means decompositions, i.e., they specify the major goals that need to be achieved in a given domain, the functions needed to achieve the goals, and systems available to achieve those functions (Rasmussen, 1986; Roth & Woods, 1988; Woods & Hollnagel, 2006).

Cognitive Work Analysis (CWA) is the most fully developed analysis framework that incorporates a function-based goal-means decomposition representation of a work domain (Rasmussen, 1986; Vicente, 1999). CWA is used to identify and represent the requirements, constraints, and opportunities for cognitive and collaborative work in a domain based on an analysis of the purposes of the system and the physical functions and processes available to meet those goals. The goal is to produce *formative models* that map out what it takes to do the job independent of the agent (person or machine) performing it or specific events. This contrasts with other approaches that are intended to provide *normative/prescriptive* models that specify how work *should* be accomplished; or *descriptive models* that represent how work *is actually* accomplished.

CWA includes five interlinked analyses. (1) *A work domain analysis* represents the goals, means and constraints in a domain that define the boundaries within which agents must reason and act. The results of this analysis provide the basis for identifying functions to be performed by humans (or machines) and the cognitive activities those entail. The remaining layers of the CWA build on the WDA foundation. (2) *Control task analysis* defines the work objectives and methods; (3) *Strategies analysis* defines specific strategies that can be used to accomplish identified control tasks; (4) *Social, organizational, and cooperation analysis* defines the social and organizational influences; and finally (5) *Worker competencies analysis* represents the knowledge and skill requirements for effective performance.

Work domain models provide functional decomposition representations from which one can objectively derive the demands inherent in a work domain that any agent (person or machine) would need to cope with. By providing a complete specification of the domain goals, functions, systems, and associated operator monitoring and control requirements, work domain models ensure that the cognitive demands associated with monitoring and control of a complex system are comprehensively covered, and that the information and decision-support that is needed to enable operators to effectively monitor and control all system functions and processes are comprehensively identified. In this way function-based work domain models ensure that the resultant displays and decision-support systems will not only support operators in performing well-defined, pre-analyzed tasks, but will also enable operators to effectively monitor and control system functions even under unanticipated conditions where the preferred means for achieving a system function may be unavailable and operators need to rely on knowledge-based reasoning to identify alternative means for achieving the goal.

CWA, and work domain modeling in particular, can also be used to support search for situations that are likely to be cognitively or collaboratively demanding. For example Bisantz, Roth, Brickman et al., (2003) used work domain analyses to uncover situations that were likely to be cognitively demanding (e.g., goal-conflict situations) as well as situations that would place a premium on close communication and collaboration across distributed personnel, as part of analysis and design of a new Navy ship. Similarly, work domain analysis methods can be used to model human automation design so as to uncover sources of cognitive and collaborative challenges and points of vulnerability (Mazaeva & Bisantz, 2007).

The *Task Specification Language* (TSL) (Sherry et al., 2010) is an approach to documenting the cognitive operations required by the users to perform mission tasks providing a framework for a more structured Cognitive Walkthrough (Lewis et al., 1990) and can be enhanced to use recently available "affordable models of human performance" to emulate simulated user testing. TSL provides a task structure that can be applied to a wide variety of domains for detailed evaluation. This method maps traditional task analysis information into a more usable format, integrates contextual information, and responds to the need for methods and tools that do not require extensive expertise to implement and interpret. The goal is to provide a framework for developers and evaluators to think about the work activity (task), how the task is triggered, and the cues provided to the user to enable task completion and monitoring of task completion. The method may be used independently to identify issues in development, or used to provide input to computational models.

Work-Technology-Alignment, evaluates how well technology aligns with the structure of the work it is intended to support (Billman et al., 2010, 2011). Technology that is better aligned with a domain of work activity should support more effective performance in that domain. Assessment of alignment depends on discovering the elements and organization of the work domain, and on assessing how well the entities and organization of the technology corresponds with that needed for the work domain. The method uses needs analysis to identify the elements and structure of the work and integrates proposals from several research traditions in human-automation interaction, human-computer interaction (HCI), Work Domain Analysis (WDA), and related disciplines to form the analysis. The goal of the analysis is to help identify where work and the functionality used to accomplish the work are not aligned, and to help provide insight into how to provide better alignment, and therefore improve human-technology performance. Work-Technology Alignment evaluates fitness for purpose against the stated intended function of the human-automation system at a higher abstraction level than a task analysis, and is more generic than a Cognitive Work Analysis (Billman, 2011).

Computational Methods

In addition to the need for more contextual information, there is a need for structured analysis methods (e.g., computational modeling) that provide outputs to help inform designers and evaluators about which risks require more in-depth evaluation (e.g., Human in the Loop simulations). The structured analysis methods need to present the results in a form that can be understood by domain experts, to enable determination of which risk estimates may require more scrutiny rather than having the analysis methods produce risk estimate values. The aviation industry, as is the case with many transportation industries have varying levels of acceptable risk, depending on the type of operation and the perceived risk by the public. These assessments are dynamic in the sense that they are specific to the system and application, and therefore need the ability of flexible assessments by humans with the support of improved data from which to make the decisions.

Examples of Interconnected task, system and environment modeling tool include:

Work Models that Compute (WMC) (Pritchett & Feigh, 2011) is a simulation framework for describing how human agents accomplish tasks within an operational context including models of physical systems and environment. In WMC, the environment is composed of resources which includes agents representing both human actors and automation.

Hamsters (Martinie et al., 2012) and *Petshop* (Palanque and Bastide, 1996) which provide a framework for describing the task (Hamsters) and connecting the executable description to a Petri-net representation (petshop) of the automation system.

Optimal control modeling can be used to predict the strategies that people will adopt given specifications of (1) human information processing architecture, (2) the subjective utility functions that people adopt, and (3) the person's experience of the task environment. This approach uses cognitive architecture in context, and generates strategies for interaction with automation (Howes, et al., 2009; Lewis et al., (2013). This work utilizes the contextual information in the form of cognitive architecture constraints, and fits well with the specific characteristics that safety-critical domains tend to provide, such as a population of expert users as the basis for evaluation.

Tools for Collecting Operational Data

One advantage of the increased use of automation, is the increase in sensor data that is available from developed systems. While this data may be limited, it can be used to provide objective data for the operational environment. Tools for data mining and knowledge discovery are in rapid development in many domains, most notably information technology, but further development is needed to enable analysis of emergent properties rather than providing data to support answers to questions that are framed specifically for the data available. (Matthews et al., 2013).

DISCUSSION

It is an exciting time for the development of joint human-automation systems. When developed successfully these can improve efficiency, remove humans from dangerous work environments and improve overall safety. However, as joint human-automation systems are developed for increasingly complex environments, new methods to support the collection of additional contextual information about the environment in which the proposed system environment will be required. This paper provided some illustrations of the need for additional contextual information, described some areas of improvement regarding how the processes are implemented and provided examples of methods that could be used to improve the automation design and risk assessment processes.

The continuing development of new evaluation methods and tools, as well as the success of automation implementation in less complex environments shows that these challenges can be overcome with a concerted effort in the research and evaluation communities, but further development of the methods and tools are needed to be usable and useful in the design and risk assessment processes in use today.

REFERENCES

Bayuk, A.J. (2008) Aviation Safety Management Systems as a Template For Aligning Safety with Business Strategy in Other Industries. American Society of Safety Engineers - The Business of Safety: A Matter of Success Symposium. Baltimore, Maryland, march 13-14, 2008.

Billman, D., Feary, M., Schreckengost, D. and Sherry, L. (2010) Needs analysis: the case of flexible constraints and mutable boundaries. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '10). ACM, New York, NY, USA, 4597-4612. DOI=10.1145/1753846.1754201 http://doi.acm.org/10.1145/1753846.1754201

Billman, D., Arsintescucu, Lucia , Feary, M., Lee, J., Smith, A., and Tiwary, R. (2011) Benefits of matching domain structure for planning software: the right stuff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Sys-*

tems (CHI '11). ACM, New York, NY, USA, 2521-2530. DOI=10.1145/1978942.1979311 http://doi.acm.org/10.1145/1978942.1979311 (2011)

Bisantz, A. M., Roth, E. M., Brickman, B., Gosbee, L., Hettinger, L., & McKinney, J. (2003). Integrating cognitive analyses in a large-scale system design process. *International Journal of Human-Computer Studies, 58,* 177–206

Boeing (2013) Statistical Summary of Commercial Jet Airplane Accidents Worldwide Operations 1959 – 2012. Boeing Commercial Airplanes.

Bye, A., Lois, E., Dang V.N., Parry, G., Forester, J., Massaiu, S., Boring, R., Braarud, P.O., Broberg, H., Julius, J., Mannisto, I., Nelson, P. (2011) International HRA Empirical Study – Phase 2 Report: Results from Comparing HRA Method Predictions to Simulator Data from SGTR Scenarios, NUREG/IA-0216, Vo2. 1 U.S. Nuclear Regulatory Commission electronic library (ADAMS) access number ML11250A010.

Chang, Y.J., Bley, D., Criscione, L., Kirwan, B., Mosleh, A., Madary, T., Nowell, R., Richards, R., Roth, E., Sieben, S., and Xoulis, A. (2013). The SACADA Database for Human Reliability and Human Performance. In Proceedings of the International Topical Meeting on Probabilistic Safety Assessment and Analysis, Columbia, SC, Sept. 22-26, 2013.

Christoffersen, K. and Woods, D. D. (2002). How to make automated systems team players. In E. Sasas (Eds.), *Advances in Human Performance and Cognitive Engineering Research*, Volume 2. St. Louis, MO, Elsevier Science, 1-12.

Commercial Aviation Safety Team (2008) Safety Enhancement 30 Revision-5 “Mode Awareness and Energy State Management Aspects of Flight Deck Automation” Final Report.

Cox, L. A. (2008) What’s Wrong with Risk Matrices? *Risk Analysts*, 28(2). 497-512.

Curtis, B., Krasner, H., and Iscoe, N. (1988). A Field Study of the Software Design Process for Large Systems, *Communications of the ACM*, 31(11), ACM, p. 1268-1287.

Falteisek, M. (2010) FAA ATO Safety Risk Management, Presented at the Workshop on Risk Assessment and Safety Decision Making Under Uncertainty, 21-22 Septemeber, North Bethesda, MD, USA

Federal Aviation Administration (2013) Operational Use of Flight Path Management Systems. *Report of the Performance-based operations Aviation Rulemaking Committee/Commercial Aviation Safety Team Flight Deck Automation Working Group*.

Federal Aviation Administration (2012) Safety Risk Management. United States Department of Transportation Federal Aviation Administration National Policy Order 8040.4A.

Feltovich, P. J., Hoffman, R. R., Woods, D., and Roesler, A. (2004). Keeping It Simple: How the Reductive Tendency Affects Cognitive Engineering, *IEEE Intelligent Systems*, May-June, IEEE Computer Society Publications Office, Los Alamitos, CA, p.90-95.

Future Aviation Safety Team (2006). The FAST Approach to Discovering Aviation Futures and Associated Hazards Methodology Handbook. http://www.nlr-atsi.nl/fast/FAST_Handbook_073009%20_2.pdf

German Federal Bureau of Aircraft Accidents Investigation (BFU) (2002) Investigation Report AX001-1-2/02, Accident (near) Ueberlingen/Lake of Constance/Germany

Helweg-Larsen, M. and Sheppard, J. A. (2001) Do Moderators of the Optimistic Bias Affect Personal or Target Risk Estimates? A Review of the Literature, *Personality and Social Psychology Review*, Vol. 5, 1, 74-95.

Hoffman, R. R., Klein, G., and Laughery, K. R. (2002) The State of Cognitive Systems Engineering. *IEEE Intelligent Systems*, January-February, IEEE Computer Society Publications Office, Los Alamitos, CA, p.73-75.

Hollnagel, E., Woods, D. D., and Leveson, N. (2006) *Resilience Engineering: Concepts and Precepts*. Burlington, VT: Ashgate Publishing Co.

Howes, A. , Lewis, R. and Vera, A. (2009) Rational Adaptation Under Task and Processing Constraints: Implications for Testing Theories of Cognition and Action, Psychological Review, 116(4), 717-751.

Hubbard, D. (2009) The Failure of Risk Management: Why It’s Broken and How to Fix It, ISBN 978-0470387955

International Association of Public Transport (UITP) (2013) World Atlas Report 2013.

International Atomic Energy Agency (2001). *Risk management: A tool for improving nuclear power plant performance.* IAEA-TECDOC-1209. IAEA, Vienna, Austria.

International Civil Aviation Organization (2012) Working Paper on the Development of an Aviation Automation Policy. AN-Conf/12-WP/34. Twelfth Air Navigation Conference, 19-30 November, Montréal, Canada.

Investigation Committee on the Accident at Fukushima Nuclear Power Stations of Tokyo Electric Power Company Interim Report, December, 26, 2011.

Kirwan, B. (1994) A guide to practical human reliability assessment. Bristol, PA: Taylor & Francis.

Leveson, N. (2012) Engineering a safer world: Systems thinking applied to safety (Engineering Systems). Cambridge, MA: The MIT Press.

Lewis, C., Polson, P., Wharton, C., and Rieman, J. (1990) Testing a Walkthrough Methodology for Theory-Based Design of Walk-Up-and-Use Interfaces. In *Proceedings of ACM CHI’90*, Seattle, Washington, April 1-5, ACM New York, NY, p. 235–242.

Lim, K.Y., and Long, J. (1994) *The MUSE Method for Usability Engineering*, Press Syndicate of the University of Cambridge, Cambridge, UK.

Lewis, R., Howes, A., and Signh, S. (2013) Computational Rationality: Linking Mechanism and Behavior through Bounded Utility Maximization To Appear in Topics in Cognitive Science.

Lois, E., Dang V.N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P.O., Parry, G., Julius, J., Boring, R., Mannisto, I., Bye, A. (2009) International HRA Empirical Study – Phase 1 Report: Description of Overall Approach and Pilot Phase Results from Comparing HRA Methods to Simulator Performance Data, NUREG/IA-0216, Vol. 1 U.S. Nuclear Regulatory Commission electronic library (ADAMS) access number ML093380283.

Martinie, C., Palanque, P., Navarre, D., Eric Barboni. A Tool-Supported Training Framework for Improving Operators. Dependability Confronted with Faults and Errors. Probabilistic Safety Assessment (PSAM11 & ESREL 2012), Helsinki, Finland, June 25-29 2012, Taylor & Francis Group.

Matthews, B., Das, S.,Bhaduri, K., Das, K., **Martin**, R., and Oza, N. (2013) Discovering Anomalous Aviation Safety Events Using Scalable Data Mining Algorithms. *Journal of Aerospace Information Systems*, 10(10),467-475.

Mauro, R. and Barshi, I. (2009) Risk Assessment in Aviation, In Proceedings of the 15th International Symposium on Aviation Psychology,April 27-30, 2009, Wright State University, Dayton, Ohio, USA

Mazaeva, N. and Bisantz, A. M. (2007). On the representation of automation using a work domain analysis. *Theoretical Issues in Ergonomics Science*, 8, 6, 509-530.

Heber, M (2013) Automation Earns Its Stripes. *Australian Mining*, 27 September.

Palanque, P., and Bastide, R. (1996) "Task Models - System Models: a Formal Bridge Over the Gap." in *Critical Issues in User Interface Systems Engineering*. David Benyon, and Philippe Palanque, editors. Springer (1996) 65-79, Chapter IV.

Paries, J. (2011) Lessons from the Hudson, In Hollnagel, E., Paries, J., Woods, D., and Wreathall, J. (Eds.) Resilience Engineering in Practice: A Guidebook. Burlington, VT: Ashgate Publishing Company

Patterson, E. S., Roth, E. M., Woods, D. D. (2010). Facets of complexity in situated work. In Patterson ES, Miller J. (Eds.) *Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams*. Ashgate Publishing. ISBN 978-0-7546-7578-5. (pp. 221- 251).

Perrow, C. (1984) *Normal Accidents: Living with High-Risk Technologies* New York: Basic Books.

Pritchett, A. and Feigh, K. (2011) Simulating First-Principles Models of Situated Human Performance, Proceedings of the 2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), Miami Beach, FL

Pritchett, A., Fleming, E., Cleveland, J., Zoetrum, Popescu, and Thakkar, D. (2012) Pilot Interaction with TCAS and Air Traffic Control, In Proceedings of ATACCS'2012 London, UK, May

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

WASH-1400/NUREG75/014 (1975). "Reactor safety study. An assessment of accident risks in U. S. commercial nuclear power plants. Executive Summary.". *WASH-1400 (NUREG-75/014)*. Rockville, MD, USA: Federal Government of the United States, U.S. Nuclear Regulatory Commission.

Roth, E. M., Malin, J. T. , & Schreckenghost, D. L. (1997). Paradigms for Intelligent Interface Design. In M. Helander, T. Landauer & P. Prabhu (Eds.) Handbook of Human-Computer Interaction (2nd edition), Amsterdam: North-Holland. (pp. 1177-1201).

Roth, E.M., Mosleh, A., Chang, Y.J., Richards, R., Bley, D., Shen, S.H., Zoulis, A (2012) "Model-based Framework for Characterizing Contextual Factors for HRA Applications" PSAM11/ESREL2012 Conference, Helsinki, Finland, June 25-29.

Roth, E. M., & Woods, D. D. (1988). Aiding Human Performance I: Cognitive Analysis. *Le Travail Humain, 51*, 39 - 64.

Sarsfield, L. P., Stanley, W. L., Lebow, C. C., Ettedgui, E., and Henning, G. (2000) *Safety in the Skies: Personnel and Parties in NTSB Accident Investigations: Master Volume*. Rand Publications, Santa Monica, CA.

Siu, N., Marksberry, D., Cooper, S., Coyne, K., and Stutzke, M. (2013) PSA Technology Challenges Revealed by the Great East Japan Earthquake, in PSAM Topical Conference in Light of the Fukushima Dai-Ichi Accident, Tokyo, Japan, April 15-17, U.S. Nuclear Regulatory Commission electronic library (ADAMS) access number: ML13038A203

Sherry, L., Edina-Mora, M., John, B., Teo, L., Polson, P., Blackmon, M., and Koch, M. (2010) System Design and Analysis: Tools for Automation Interaction, Design and Evaluation Methods. Final Report for NASA NRA NNX07A067A.

Sherry, L. and Ward, J. (1995). Formalism for the specification of operationally embedded reactive systems, In *Proceedings AIAA/IEEE Digital Avionics Systems Conference (DASC)*, 5-9 Nov., Cambridge, MA, USA, p. 416-421.

Stamatelatos , M. (2000) Probabilistic Risk Assessment:What Is It And Why Is It Worth Performing It? NASA Office of Safety and Mission Assurance.

U. S. Department of Energy (2013) Development of Probabilistic Risk Assessment Assessments for Nuclear Safety Applications, DOE-STD-1628-2013.

U.S. National Aeronautics and Space Administration (2004),NASA Software Safety Guidebook, NASA-GB-8719.13

U. S. National Aeronautics and Space Administration (2011), "Risk Management Handbook," NASA/SP2011-3422, Version 1.0.

Vicente, K. J. (1999). *Cognitive Work Analysis*. Mahwah, NJ: Erlbaum.

Whaley, A. M., Xing, J., Boring, R., Hendrickson, S., Joe, J., and Le Blanc, K. (2012) Building a Psychological Foundation for Human Reliability Analysis, NUREG-2114

Woods, D. D. and Hollnagel, E. (2006). *Joint Cognitive Systems: Patterns in Systems Engineering.* Boca Raton, FL: Taylor & Francis.