Design for Safety: A cognitive engineering approach to the control and management of nuclear power plant

Guy A. Boy and Kara A. Schmitt

ABSTRACT

This paper presents an analytical approach to design for safety that is based on 30 years of experience in the field of Human-centered design. This field is often qualified as governing safety–critical systems where risk management is a crucial issue. We need to better understand what the main facets of safety are that should be taken into account during the design and development processes. There are many factors that contribute to design for safety. We propose some of these factors and an articulation of them from requirement gathering and synthesis to formative evaluations to summative evaluations. Among these factors, we analyze complexity, flexibility, stability, redundancy, support, training, experience and testing. However, we cannot design a safe and reliable product in one shot; design is incremental. A product and its various uses become progressively mature. When we deal with new products, issues come from the fact that practice features emerge from the use of the product and are difficult, even impossible, to predict ahead of time. The automation within is an important portion of this maturity, and must be understood well. This is why design for safety is not possible without anticipatory simulations and a period of tests in the real world, such as operational testing in nuclear power plants. In addition, designing for safety is not finished when the product is delivered; experience feedback, or human-in-the-loop simulation (HITLS) is an important part of the overall global design process. The AUTOS pyramid approach can assist in simplifying the understanding, and improving the design of a complex system by describing and relating Artifacts, Users, Tasks, Organizations, and Situations.

KEYWORDS


1 Human—Centered Design Institute, Florida Institute of Technology, 150 W. University Blvd., Melbourne, FL 32901, USA. Phone: 1 (321) 674-7631. Fax: 1 (321) 674-7175, gboy@fit.edu — schmittk@fit.edu
1. Introduction

Design for safety has become a key process in industry producing systems that involve risks. These systems are denoted “safety–critical”, e.g., nuclear systems, aircraft, spacecraft, medical systems and automobile. What do we mean by safety? The field of system safety, reliability, availability and dependability is very broad and deep; it is investigated for a long time, both in research and industry (Johnson and Malek, 1988; Laprie, 1992, 1994; Prasad et al., 1996; Nilsen and Aven, 2003). Methods were developed for the assessment of human reliability as extension of probabilistic methods addressing system reliability such as the Technique for Human Error Rate Prediction (THERP) (Swain and Guttman, 1983) or Standardized Plant Analysis Risk (SPAR-H) (Gertman et al., 2005). Unfortunately, probabilistic models do not accurately account for predicting human errors and more generally human behavior. What do we mean by the “probability of a human operator to be incapacitated or inoperative in a nuclear power plant control room”? First of all, is it a meaningful question? Answers to these kinds of questions fostered the need for starting deeper investigations on the human side of safety–critical human–machine systems. Consequently, Human Reliability Analyses (HRA) were developed from various perspectives (Byers et al., 2000). HRA is based on the likelihood of human errors or erroneous actions where experience feedback is the source of knowledge. However, resulting databases enable us to explain the genesis of an incident or accident after the fact, but does not provide insight into accident prevention in the future. In other words, we can easily explain after the fact, but we cannot predict the future. We then need to have an approach that addresses deeper knowledge of human–machine systems, and more specifically their socio-cognitive complexity.

Cognition-induced problems motivated numerous research efforts during the last three decades to the point of creating a new field of research called cognitive engineering. Physical restraints within the workplace have now become cognitive constraints. Even if these constraints may seem softer, they are not less implacable; we moved from physiological and mechanical exhaustion of the worker to mental exhaustion that may cause more pernicious effects. The cognitive community promoted “human reliability” through the investigation of human errors, associated risks and recovery strategies. Everybody knows that “errare human est”, i.e., to err is human, but we often recover almost immediately. Unfortunately, there are errors that may lead to undesirable and even catastrophic situations. For that matter, cognitive engineering introduced a new set of conceptual tools such as the Contextual Control Model (COCOM) and the Cognitive Reliability and Error Analysis Method (CREAM) (Hollnagel, 1993, 1998). The concept of human error dominates this kind of cognitive approach (Reason, 1990; Hollnagel, 1991). Reason emphasizes the systemic approach of human error management that concentrates on the conditions under which individuals work and tries to build defenses to avert errors or mitigate their effects, instead of blaming these individuals for forgetfulness, inattention, or moral weakness. But cognition is not derived independently; human operators evolve in a social environment where they have to comply with established socio-technical safety culture and principles. The main principle in the nuclear industry is “defense in depth” which establishes a series of barriers to stop failures and avoid their propagation (Guillermain and Salazar-Ferrer, 1999). Socio-technical reliability (that includes the integration of both human reliability and technical reliability) is related to the distribution of appropriate roles or functions among agents (Boy, 1998). This is related to authority sharing among agents (Boy and Grote, 2009). Early developments were based on Fitts’s approach to “function allocation” that attempts to systematically characterize the general strengths and weaknesses of humans and machines (Fitts, 1951). Principles and methods, referred as MABA–MABA (Men-Are-Better-At/Machines-Are-Better-At), were developed to determine which system-level functions should be carried out by humans, and which by machines.

The U.S. Nuclear Regulatory Commission (NRC) outlines function allocation techniques to be applied to satisfy plant safety objectives prior to obtaining an operating and control license in the United States. This includes identification of functions, specification of requirements, analysis of allocation, automation justification, design development and modification, and function verification (Nuclear Regulatory Commission, 2004).

In this paper, we claim that the balance of automation between humans and machines can be an excellent resource when it is designed correctly. More importantly, people are not the problem but are the solution to
proactively maintain global safety when they are competent and a good balance of automation exists. Indeed, before human operators can effectively conduct safety–critical systems, they need to be well trained and have developed the appropriate experience. In addition, despite all possible training and experience, people are always subject to failure, i.e., they commit errors. Communication, cooperation and coordination among team members may fail. Human errors can be patent (e.g., erroneous knowledge and knowhow, slips, and use of wrong mental models), or latent (e.g., design flaws of user interfaces, operational documentation and organizational setups). This is why constant human operator involvement and crosschecking is mandatory in safety–critical systems, e.g., Nuclear Power Plant (NPP) control rooms are staffed by two operators and a supervisor for redundancy. Finally, when the command and control system itself is correctly automated, it is a useful and effective barrier to most disturbances whether they are nuclear system failure, human errors or external threats.

Typically, people involved in NPP control and management set up adaptive mechanisms to cope with normal, abnormal and emergency situations. In incidental situations, several cases may arise. First, if a human operator faces a simple problem (e.g., a sub-system failure), he or she takes care of the situation directly and reports to his or her manager, who will acknowledge or continue the investigation. Second, if a more complex problem arises, the manager will have to take the overall problem into account and use his or her team to solve the problem. More generally, simple problems induce event-driven responses, and more complex problems induce goal-driven problem solving. In any case, control and management of a NPP is a team process, which is now extended to artificial agents (i.e., software-based automation). Teams require that their members know about knowledge, knowhow and attitudes of their colleagues. If the humans and machines are to work together as team players (Klein et al., 2004), more efforts are required in both automation design, and training of human operators using the automation. We will see in this paper that there are three kinds of interaction models in multi-agent systems: supervision, mediation and cooperation by mutual understanding (Boy, 2002). Another important characteristic of nuclear systems, especially in incidental situations, is the speed of evaluation of the situation. Finally, even if a multi-agent interaction model has been chosen for each generic operational context, articulation work is always necessary, i.e., communication rules should be clearly understood. This aspect involves trust and personal involvement.

Therefore, even if we would like to rationalize safety and safety–critical technology and organization in a systemic sense, the field remains a matter of people. They are of course human operators (i.e., users of safety–critical technology), but also designers, manufacturers, maintainers, certifiers, trainers and (not to forget) managers. Most Human Factors approaches attempt to correct and adapt engineering systems to users after these systems are developed; whereas Human-Centered Design (HCD) takes into account people from the very beginning of design (Boy, 2011). HCD must consider not only the system components, but also the inter-actions. A proper analysis will holistically take into account five entities and their interrelations:

- The Artifact being designed (i.e., technology that is being designed).
- Possible Users or human operators (i.e., a categorization of user profiles is necessary).
- The various Tasks that are anticipated (i.e., inputs of the various cognitive functions that the various agents will have to use).
- The Organization in which users will perform tasks using the artifact (i.e., typically a set of human and machine agents).
- The various Situations (i.e., various kinds of context patterns that characterize the environment).

The AUTOS \(^2\) pyramid framework describes the various interrelations among these five entities (Boy, 2011). It is a means of addressing the operational basis of resilience engineering (Hollnagel et al., 2006) as well as technology and practices’ complexity, by a combined study of complementary perspectives of

\(^2\) AUTOS means Artifact, User, Task, Organization and Situation.
human–system integration. These interrelated perspectives are intended as the framework to support completeness in Human–Machine System (HMS) analysis, design and evaluation. This approach has been successfully applied in aviation in order to determine that airline pilots rely on skills and knowledge over procedures (Boy and de Brito, 2000), and also to determine that Cognitive Function Analysis (CFA) and AUTOS were successfully used as an appropriate approach to organize experience feedback while designing cockpits that improve situational awareness (Boy and Ferro, 2003). Applications range as diverse as improving to comfort (Dumur et al., 2004) for passengers in the cabin.

AUTOS is not an alternative to HRA or other human reliability approaches. It is intended to guide designers to systematically consider crucial human-centered properties. It will be used in this paper to support the definition of a safe HMS and the design of new NPP Control and Management Systems (CMSs) by looking at examples of the past. It is based on a 30 years of experience in safety–critical systems. It also provides a rationalizing distinction between incidental/accidental and routine use. Design for safety is a matter of both expertise and shared knowledge and practices, and properly applied operational feedback. The knowledge of the NPP operational population is very important and should be involved in the participatory design of new NPP CMSs. In addition, an ideal CMS cannot be designed in one shot, it has to be matured and improved throughout the entire life cycle, i.e., “constant” evaluations should be performed.

2. What is a safe human–machine system?

It is crucial to understand and master the attributes of design with regards to safety in order to improve them. Safety can be seen as the resilience of a HMS to failure, i.e., the result of HMS adaptations necessary to cope with the complexity of the real world (Hollnagel et al., 2006). We will use the term “failure” for both humans and machines, even if we usually talk about error in the case of people. We would like to be able to avoid, detect, remove, tolerate or resist failures in human–machine systems. However, it is never possible to anticipate all probable failures at design time. This is due to ignorance, uncertainty and imprecision. In particular, we cannot anticipate all possible practices. Even if designers try to anticipate unforeseen potential uses, users always find peculiar ways of using a system that cannot be anticipated. Disciplines such as reliability engineering or quality management have developed methods of insuring a system would work as well as possible, and most industries use the resulting methods. Standards such as ISO 14001 are now routinely used in industries that produce safety–critical systems. Procedural application of such standards is supposed to guarantee the safety of products that will be delivered. In practice, nuclear systems are certified with extremely high level of safety, such as probability of failure per hour of activity of $10^{-9}$. Ultra-safe systems, to take the terminology used by Amalberti (2001), are the result of the highest level of expertise, experience feedback and a systematic habit of testing.

Our worldwide economy has cut industry into pieces, which are then disseminated all over the world, and assumes that well-designed procedures and organization can lead to the production of quality products. A piece of equipment may be designed and built by people who do not have the knowledge of how it will be used and maintained, and conversely the people who will use it may not have enough knowledge of the way it was designed and built. For safety–critical systems, it is important to master the integration of these pieces. When technological and practice maturity are low, strong and effective connectivity among the various actors is crucial. Safety of new critical systems is therefore a matter of participatory design, development, maintenance, use and evolution until technological and practice maturities are reached. In this paper, we will focus on three main factors: complexity, cognitive stability and flexibility.

---

2.1. Complexity

A distinction is made between internal complexity of an artifact and perceived complexity induced by the use of it (Boy, 2008). Of course, the former is likely to induce the latter, but even if internal complexity is mastered, people may perceive some complexity in the interaction. Systems that we develop are more complex because they are becoming more interconnected. This increasing interconnectivity causes current software-intensive systems to look like biological systems where we need to identify emerging properties. Therefore, reliability, availability and maturity of these artifacts need to be tested in all possible critical situations. In practice, this is difficult, and sometimes impossible, because human operators use them in many ways that were not anticipated by designers. We usually talk about surprises. The design of a safety–critical system requires that such surprises are anticipated before they are delivered.

Work is already underway to identify the sources of complexity. Some of the identified sources include environmental, organizational, interface and cognitive complexity (Sasangohar, 2010). The difficulty in managing these complexities, such as control room size, number of procedures or real time update rates, is not only identifying all of the sources, but also understanding the dynamic interactions between them.

Safety–critical systems design and engineering are typically based on system functional analyses. However, prototyping and formative evaluations based on Human-In-The-Loop Simulation (HITLS) enable activity analysis. This is why modeling and simulation are key processes in design for safety. Human operators must understand the complexity and dynamics of not only the reactor, but also the layered processes that control them. This is often an issue of trust. Therefore, providing the right information at the right time using the right format (to be understood in context) is a key issue in HCD. Visualization techniques can be used to achieve this kind of objective. The main problem is always to provide the necessary and sufficient information at the right time: too much information is likely to overload human operators, and too little information can create situation awareness issues. Design for safety can be seen as design for simplicity, thereby reducing the complicated nature of the system by rationalizing their internal complexity. Therefore, it is important to assess the complexity of a safety–critical system being designed. There were several attempts to assess artifact complexity. In cognitive engineering, Rasmussen (1983) proposed the SRK (i.e., Skills, Rules and Knowledge) model to capture three types of behavior. He also developed an ecological approach based on five levels of abstraction hierarchy. Vicente (1999) used this approach to develop the Cognitive Work Analysis (CWA) approach. Norman (1986) proposed a generic model that takes into account human actions, learning, usability and possibility of errors. Amalberti (1996) related interaction complexity to action reversibility and effect predictability, the dynamics of underlying processes, time pressure, the number of systems to be managed at the same time, resource management when the execution of a task requires several actors, artifacts representation, risk, factors coming from the insertion of safety–critical systems into cooperative macro-systems and factors related to the human–machine interface, users’ expertise and situation awareness.

2.2. Cognitive stability

Design for safety can also be seen as design for stability. The concept of stability has been studied in mechanics and physics. There are systems that are stable by themselves and others that are unstable. The pendulum for example is a stable system. When one pulls the mass at the bottom of a pendulum, it returns to its stable state by itself. This is referred to as passive stability. This model works for a Human–Machine Cognitive System (HMCS) where a human agent interacts with a machine agent. Whenever a human agent or a machine agent fails, the overall system returns to a stable state by itself (passive cognitive stability) or requires additional actions (active cognitive stability). Consider a HMCS where there are human operators managing and controlling a NPP.

There are various types of system reliability philosophies. Fail-operational systems continue to operate when their command and control sub-systems fail, e.g., some nuclear reactors are passively safe. Note that this kind of strategy may be unsafe. Fail-safe systems remain safe when they cannot operate, e.g., nuclear weapons belong to this class. Fail-secure systems maintain maximum security when they cannot operate, e.g., some kind of automation that takes control. Note that this kind of strategy may be dangerous if the
automation is clumsy (Wiener, 1989). Fail-passive systems continue to operate in the event of a subsystem failure. Fault-tolerant systems avoid service failure when fault are introduced to the system, e.g., current nuclear reactor control systems.

It is useful to extend the concept of cognitive stability, i.e. a human agent interacting with a machine agent, to a multi-agent environment. Therefore, it would be more appropriate to talk about *Socio-Cognitive Stability* (SCS) even if we will use the term “cognitive stability” for both individual agents and groups of agents. Consequently, interaction between human and software agents can be modeled as a network of cognitive functions (Boy, 1998) where cognition is distributed among humans and machines (Hutchins, 1995).

In addition, there are unexpected or unplanned events that oblige the revision of agents’ tasks in order to ensure an acceptable level of safety. The concept of socio-cognitive stability is very close to the concept of the resilience (Hollnagel et al., 2006) of the socio-cognitive system, e.g., defense in depth. Passive SCS refers to a multi-agent system that returns by itself to a stable state after a disturbance. Active SCS refers to a multi-agent system that requires external intervention to make it return to a stable state. Obviously, there is a continuum between passive and active socio-cognitive stability where several levels of difficulty can be defined to stabilize a multi-agent system, as well as several levels of resilience of that system. We proposed that this difficulty be assessed using three kinds of metrics: time pressure criticality; complexity; and flexibility (Boy and Grote, 2009).

### 2.3. Flexibility

Flexibility needs to be considered in relation to cognitive stability and complexity. Safety nets, barriers and protections will not remove the possibility of failure either from humans, or from safety systems themselves. This is why expertise and experience are still crucial assets in the management of safety–critical systems. Consequently, human operators need to have appropriate skills and tools to handle unpredictable or infrequent events. Therefore, design for safety can also be seen as design for flexibility. When we refer to flexibility, we are speaking of adaptation and resilience, and the capacity to appropriately act in response to a disturbance. Flexibility deals with the ability of the operating agents to diagnose, solve problems and act in the right context. Conflict arises from the fact that we develop (rigid) procedures to help people in abnormal and emergency situations. People do not use operational procedures when they do not fully understand their effects in context. This is specifically true in highly safety–critical situations. In these cases, airline pilots, for example, rely on their skills and knowledge, and require enough flexibility to diagnose, solve problems and ultimately act (Boy and de Brito, 2000).

When an unpredicted event occurs, agents in charge should be able to make the system return to its stable state. Such agents could be either automata or human operators. The latter require redundancy in the control of a safety–critical system. A human operator requires cognitive stability to act safely and therefore appropriate cognitive support that could be provided in the form of redundant states useful for crosschecking. Safety assurance is also a matter of intimate connectivity between the various appropriate agents; this is the source of complexity. This assumes that each agent has sufficient awareness of the relevant states of the other agents in the system. The more awareness, the better in safety–critical situations, but such awareness should be complemented by appropriate situated action patterns. In order to insure situation awareness, there are two requirements: appropriate human operator’s attention and vigilance; and appropriate affordances of the software agent. Attention and vigilance depends on training, fatigue and other physiological, psychological and social factors (Wickens and Hollands, 2000). Natural relationships between people and machines, called affordances (Gibson, 1979), emerge from various kinds of interactions among them. These affordances have to be identified in order to reduce the difficult problem of unknowns. This is why HITLS are necessary to induce the various interactions between people and machines.

A human–machine system is safer when it is both passively and actively stable, i.e., passive and active stability phenomena must be clearly identified, understood and mastered. Though the nature of design is iterative, beginning correctly is crucial.
3. How can we characterize human-centered automation?

3.1. Automation as cognitive function transfer

The concept of cognitive function was developed in order to capture both concepts of task, what is prescribed to be done, and activity, what is actually done (Boy, 1998, 2011). Activity can also be called situated work or performance. A cognitive function transforms a task into an activity (Fig. 1). The main problem is the availability of appropriate end-users. In the following of this paper, several examples coming from operational experience will be presented to illustrate needs for the development of an integrated approach to a human-centered design for safety. Automation can be characterized as a transfer of cognitive functions from human to machine. Therefore, designing a new artifact is somehow automating some cognitive functions that were used by people to achieve goals, and encapsulating the resulting machine cognitive functions into a machine or an artifact. Douglas Engelbart said, “We extend human cognition by providing cognitive tools to people”. Cognitive functions are typically defined by their role, context of validity, and resources that enable their operationalization (Boy, 1998). The difficult part is not to design a new artifact that will encapsulate a previously-human cognitive function; but to understand the resulting emerging cognitive function(s) induced by the use of the new artifact.

![Fig. 1. A cognitive function as a transformation of a task into an activity.](image)

Automation can be characterized as a transfer of cognitive functions from human to machine. Therefore, designing a new artifact is somehow automating some cognitive functions that were used by people to achieve goals, and encapsulating the resulting machine cognitive functions into a machine or an artifact. Douglas Engelbart said, “We extend human cognition by providing cognitive tools to people”. Cognitive functions are typically defined by their role, context of validity, and resources that enable their operationalization (Boy, 1998). The difficult part is not to design a new artifact that will encapsulate a previously-human cognitive function; but to understand the resulting emerging cognitive function(s) induced by the use of the new artifact.

3.2. Satisfying product and practice maturity requirements

Human-in-the-loop simulations (HITLs) have to be developed in order to develop realistic user experiences, which create the conditions for emergent cognitive functions and behaviors. This allows us to design these functions into the system prior to delivery time. The certification process is typically during this period. The main problem comes from the fact that industry wants to shorten the maturity period of a system while keeping costs competitive. Maturity takes time and appropriate efforts! This is why competent leaders are required for maturity management.

Curtis and his team at Carnegie Mellon University have taken a serious interest in maturity through the development of the Capacity Maturity Model (CMM) for software development (Paulk et al., 1995). This technique is currently used in industry worldwide. However, it emphasizes the maturity of development processes, and it assumes that if these processes are mature then the product will be mature; this is an assumption that only works in certain contexts. We developed an alternative approach to product maturity that takes into account a human-centered definition of high-level requirements, as well as usability and usefulness tests of the product being developed (Boy, 2005). Product maturity is a matter of time and extensive tests that involve end-users. It can be seen as incorporating user experience in the definition and
implementation of the product. In fact, user experience is crucially needed during the entire design and development processes.

Today, industry is producing technology using a multitude of agents worldwide. Consequently, integration of the product and coordination of the agents are two crucial processes that need to be better understood and managed. Prior to modern techniques, products were designed and manufactured within integrated organizations where people knew each other. However, products were much simpler than they are today. Quality was insured by the technological glue that this integrated way of working was naturally providing. Today, work is performed in isolated places where the turnover of people is high. We are in a very isolated, yet inter-connected set of networks with people who are not familiar with each other. Parts are described on paper or electronic media, but their description is not as deep as integration requires. This is why new types of integrators and coordinators are required. They often are not there. This is why research is highly needed in design for safety.

Management is intended to support technical work. However, the complexity and uncertainty of economical industrial environments have led to the necessity to justify almost all activities and control costs. Therefore, industrial organization and management setups have been financially mechanized. In particular, structured reporting has been installed. The main problem that arises from the related emergent human factors induced from the implementation of poor quality processes. Even if quality engineered processes are implemented, product quality and maturity are remaining issues because attributes of the production process must be mastered by people. We have quality assurance processes that contribute to the generation of huge quantity of cost information, but the processes fail at providing meaningful socio-technical information on product usability and usefulness. Related authority has been transferred from previous technical leadership to managers whose main task is to decrease costs and increase benefits. Issues are not based on long-term visions, but on short-term benefits. Consequently, even if human factors are taken into account, this is done in a way that barely involves end-users, simply because it is too costly for a project with short-term goals. However, in the long term, involving users in the design process can save significant amounts of money, and increase safety tremendously.

3.3. Inventing human-centered design solutions and the importance of domain expertise

Perceived over-regulation has led to new considerations that operators may be reaching a position where they are unwilling to step out of a procedure to perform the correct action, even if the correct procedure does not exist. While witnessing simulator training, operators began to take proper action, only to stop to locate the procedure associated with that action, losing valuable seconds. An operating plant has hundreds of procedures, with even more enclosures for context-management, e.g. a procedure for a component and enclosures for start-up, shut-down, normal operations, etc. Categories of procedures are plant specific, but can be generally classified as Emergency Operating Procedures, Normal Operating Procedures, Abnormal Operating Procedures, Surveillance Instructions (driven by technical specifications), Performance Instructions, and Admin Procedures. All of these procedures are built in attempts to help clarify context for the users, but such a large number of variable situations can be dreamt that we’re in a position of information overload. In the event of a “beyond design basis” event, there are also Severe Accident Mitigation Guidelines (SAMGs). The intent of SAMGs are to prevent the escalation of an event into a severe accident, mitigate the consequences of a severe accident, and achieve a long term safe stable state (IAEA, 2009). Research is currently being performed in order to determine where the proper balance of human operators, problem solving abilities and automation lies (Schmitt and Boy, 2012).

The nuclear industry is probably the safest industry. One main problem with such ultra-safe industries (Amalberti, 2001) is that these industries can often be overregulated to the point that they produce procedures and rules that pile up and constrain operations within a very narrow domain (Fig. 2). Consequently, human operators need to comply with these procedures and rules, and have much difficulty extrapolating out of this very constrained operational domain when it is needed. A study carried out in the aviation domain showed that 65% of airline pilots do not use procedures that are not judged relevant in context (De Brito, 1998). In addition, they typically do better by using their airmanship. People are unique problem solvers when unanticipated situations occur, but they can make errors. Human factors research has
focused so much on human errors during the last three decades that it seems that we have forgotten that people can be unique resources during operations. This is why domain expertise and experience are crucial in problem solving. Problem solving knowledge and skills are necessary individually and collectively. Individually, human operators need to have appropriate domain knowledge depth and skills to solve operational problems, including unexpected event. For that matter, preparation is key, and can be typically supported by adapted simulation. Collectively, operations teams should be organized to mix complementary as well as redundant domain knowledge and skills. Communication, cooperation and coordination (the 3Cs) are organizational processes that need to be mastered to enhance collective problem solving.

There is a basic contradiction in the job of human operators of safety–critical systems, that is they are asked to very strictly follow procedures, and be creative and deviate from these procedures when unexpected events occur (Rosay, 2011). Therefore, procedure following and creative problem solving have to be learned and experienced constantly in safety–critical environments. For that matter, designing systems and procedures that go with them tends to rigidify human operators’ possibilities of actions, which is good when situations are well identified and known in advance, but also need to be flexible when human operators have to take the lead to solve unexpected problems.

![Fig. 2. The homeostatic positive cycle of the accumulation of procedures.](image)

4. Designing automation in a NPP

Applying what we already presented in the first part of the paper, this section proposes an analysis and an approach for the design of automation in a NPP. The following human-system integration approach is intended to support the development of an architecture based on appropriate levels of automation and supported by the analytical AUTOS pyramid framework (Boy, 2011).

4.1. Levels of automation

To consider automation issues, it is necessary to provide a common frame of reference. There are a few taxonomies that support the classification of automation (Sheridan and Verplank, 1978; Krobusek et al., 1988; Billings, 1991, 1996; Endsley, 1995). Sheridan and Verplank proposed 10 levels of automation that are: (1) The computer offers no assistance: human must take all decision and actions; (2) The computer offers a complete set of decision/action alternatives, or; (3) narrows the selection down to a few, or; (4) suggests one alternative, and; (5) executes that suggestion if the human approves, or; (6) allows the human a restricted time to veto before automatic execution, or; (7) executes automatically, then necessarily
informs humans, and; (8) informs the human only if asked, or; (9) informs the human only if it, the computer, decides to; (10) the computer decides everything and acts autonomously, ignoring the human.

<table>
<thead>
<tr>
<th>Table 1: Billings’s levels of automation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level</strong></td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Autonomous Operations</td>
</tr>
<tr>
<td>Management by Exception</td>
</tr>
<tr>
<td>Management by Consent</td>
</tr>
<tr>
<td>Management by Delegation</td>
</tr>
<tr>
<td>Shared Control</td>
</tr>
<tr>
<td>Assisted Manual Control</td>
</tr>
<tr>
<td>Direct Manual Control</td>
</tr>
</tbody>
</table>
Table 2: O’Hara and Higgins (2010) Levels of automation for a NPP

<table>
<thead>
<tr>
<th>Title</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1: Manual operation</td>
<td>Involves no automation, and human operators manually perform all functions and tasks.</td>
<td>We anticipate that new NPPs will have several manual systems, such as the demineralized water system.</td>
</tr>
<tr>
<td>Level 2: Shared operation</td>
<td>Involves automatic performance of some functions and tasks, and human operators control some functions and tasks manually.</td>
<td>The Service Water System is mostly operator-controlled, but pump starts automatically on trip of running pump. Also, the human operator lines up and starts the residual heat removal, then the Suppression Poll Cooling Mode controls temperature automatically.</td>
</tr>
<tr>
<td>Level 3: Operation by consent and delegation</td>
<td>Involves machine cognitive functions that automatically perform when directed by a human operator, i.e., the human operator start them, monitor them and supervise them.</td>
<td>Advanced boiling water reactors and the reactor coolant system of an economic pressurized reactor have such functions, and consequently work at this automation level. Also at level 3 is the steam generator water level control system of US NPPs for example. Note that the EPR RCS can be level 3 or level 4.</td>
</tr>
<tr>
<td>Level 4: Operation by exception</td>
<td>Involves machine cognitive functions that operate autonomously unless specific situations or circumstances are encountered. Human operators must approve critical decisions and may intervene.</td>
<td>Examples include current boiling water reactor high pressure cooling injection system and automatic depressurization system, and the Advanced Plant 1000 passive containment cooling system (see page 14 of O’Hara and Higgins, 2010; Westinghouse, 2008).</td>
</tr>
<tr>
<td>Level 5: Autonomous operation</td>
<td>Involves machine cognitive functions that operate fully autonomously and not normally able to be disabled, but may be manually started. Human operators monitor performance and perform backup if necessary, feasible and permitted.</td>
<td>Current NPP safety systems are fully automated, such as the reactor protection system, the engineered safety feature automation system and emergency diesel generators on a loss of offsite power.</td>
</tr>
</tbody>
</table>

Later on, Billings’s contribution in human-centered automation introduced several other useful human factors issues such as vigilance, complacency and monitoring. Billings proposed seven levels of automation that are presented in Table 1. More recently, based on Billings’s contribution, O’Hara and Higgins (2010) proposed a five-level categorization for the automation of a NPP, see Table 2.

Nowadays, what is usually called automation is nothing more than software. Machines are increasingly equipped with software that takes the burden of tasks previously executed by people. It is interesting to observe that proceduralization is a kind of automation; it is automation of people (Fig. 3). The primary issue is often flexibility because neither automation nor procedures can be fully context-sensitive. It is also important to recognize that both automation and procedures are very useful in most situations, and are good for routine activities. When both automation and procedures become mature (i.e., technology and practice
maturity), they are typically used as safeguards by human operators, i.e., they are used as cognitive support to assure cognitive stability. The industry application of contextualization provided by electronic procedures is very promising. Contextualized electronic procedures can be considered as machine agents that support human interaction with complex systems.

![Diagram](image)

**Fig. 3. Automation of people and machines.**

### 4.2. Artifact-driven design for safety

Automation tends to promote management skills over direct control of the system. The main reason is that previous active stability maintenance is replaced by a more passive stability assurance that needs to be managed. Human operators go from a high-stress problem-solving job to a hypo-vigilance monitoring work. There are two main problems that emerge from automation regarding the change in the nature of socio-cognitive stability. The *maturity* of the new automated system, (i.e., reliability, availability and robustness) the *maturity of use* of the automated system (i.e., adaptation of current practice to a new one).

When researching design, it is important to look at real world examples, and attempt to learn what we can from the mistakes of our past. Within the nuclear power domain, we can provide many examples of where HCD, and specifically the artifacts were not considered. In 1986, the world’s worst nuclear disaster took place at Chernobyl. While in an abnormal situation, performing an unapproved test which disabled many of the safety systems which were in place, the reactor vessel ruptured, causing a series of explosions and significant radiation release. Though operator error was initially cited as the root cause, further investigation showed that operational instructions, design deficiencies and the safety culture through the design and operating process were at fault. The accident raised concerns about the overall safety culture of the nuclear power industry.

In 1979, *Three Mile Island* (TMI) near Harrisburg, Pennsylvania had a level 5 incident on the *International Nuclear Event Scale* (INES) including a partial meltdown of the core. The TMI disaster caused the nuclear power industry to undergo a multitude of changes to ensure the future safety of NPPs (Walker, 2004). That

---

4 The INES scale, derived by the IAEA is a methodology of identifying and responding to nuclear incidents and accidents. It ranges from 0: Deviation to 7: Major accident. Chernobyl, and Fukushima are the only level 7 events in the industry’s history.
morning the plant was running at 97% of full power when the reactor scrammed. The control rods dropped into the core and the nuclear reaction was stopped, however, the decay heat still needed to be removed from the system. Generally a simple and automatic process, it had become greatly complicated when a Pilot-Operated-Relief-Valve (PORV) became stuck open. This valve, in addition to organizational factors and human error, are credited as the primary failures that led to establishing TMI as a household name, and a centerpiece for champions against nuclear power. The accident at TMI occurred as a result of a series of human, institutional, and mechanical failures (Kemeny, 1979). These three conclusions can be correlated with the User, Organization, and Artifacts, in the AUTOS pyramid.

In the realm of the artifact, the control room, the heart of a NPP, is cited by the Kemeny report as one of the direct causes of the TMI incident. Operators felt as though there was no logic behind the way the control room was laid out, for example, one of the most important alarms – reactor coolant pressure – is right next to the light that tells you the elevators are stuck in the turbine building (Grey, 2003). At one point, an operator attempted to notify the plant superintendent of the design deficiencies:

“The alarm system in the control room is so poorly designed that it contributes little in the analysis of a casualty. The other operators and myself have several suggestions on how to improve our alarm system – perhaps we can discuss them sometime – preferably before the system as it is causes severe problems.” (Grey, 2003).

This letter is an example of experience feedback, which is crucial in HCD. Had operator comments been incorporated throughout the design phase, or even within the operation phase, a much different outcome may have been achieved. Take caution though, once a design has moved into the operational phase, experience feedback often leads to additional procedures and automations within the system. The issue we face today is to utilize user feedback in an integrated way. The optimal phase to perform this function is in the design phase.

Today’s control rooms have been significantly redesigned. At a minimum, the control room has allocations for a nuclear core operator, a turbine island operator, and a shift manager. The control room layouts have become more focused on supplying operatory with the information they need during emergency situations.

Under Nielsen’s Acceptability concept diagrams, usability can be defined as learnability, performance/efficiency, retention ease, human error and subjective pleasure (Nielsen, 1993). This research was further confirmed by Preece, Rogers and Sharp where they discussed the interactions beyond the technological human computer interaction (Preece et al., 2001). It’s important to understand that in the beginning, the technological limitations are what kept us from advancing, but as Moore’s law proves itself, and technology becomes cheaper and available to the masses; it has never been more important to focus on the usability for the general user. Nielsen’s acceptability of concept diagram also considers the social and practical acceptability, utility, cost, compatibility and reliability/maturity (Nielsen, 1993). Some other heuristics of usability include:

- Designing a natural and simple dialog, and speaking the language of the user.
- Using simplicity to create redundancy.
- Avoiding machine language and terminology that could be ambiguous.
- Minimizing the users memory load, considering recognition vs. recall.
- Considering default values, generic commands, and informational objects manipulation.
- Assuring internal and external consistency within the program and with surrounding interfaces.
- Providing complete informational feedback, such as good error messages.
- Anticipating human error.
- Protecting from scenarios with high-risk outcomes.

Ergonomics is another field of study to be taken into account when design of the physical artifact takes place. Ergonomics is a study of the workplace environment and the user’s interaction with the overall
system to optimize human well-being (International Ergonomics Society, 2000). The study of Ergonomics has been directly related to the field of quality, and a correlation between improving ergonomics has been identified with improvement of quality in a final product or process (Eklund, 1997). Numerous Military specifications have outlined proper metrics for ergonomics, such as MIL-HDBK-759B Human Factors Engineering Design for Army Materiel, DOD-HDBK-763 Human Engineering Procedures Guide, and specifically for the nuclear industry by the NRC under NUREG-0711, Rev. 2 Human Factors Engineering Program Review Model.

In regards to technical evolution, it is important for a system to not only be designed well, but be designed such that the ability to upgrade components is built specifically into a system. This includes understanding the functions, which may be outcomes of upgrading that hardware (Ishigami et al., 2003).

4.3. User-driven design for safety

People learn to preserve themselves from the early days of their lives and do not stop improving their defenses against external aggressions. For that matter, they try to keep an acceptable cognitive stability in their environment. Indeed, cognitive stability is context-dependent. The more human beings are familiar with their environment, the more they are cognitively stable, i.e., they know how to interact within this environment. They learn the various appropriate patterns that are mandatory to maintain high cognitive stability.

In a safety–critical environment, people learn how to anticipate, interact and recover from failures through training and experience. Anticipation is certainly one of the most important skill or cognitive function that a human operator needs to have in order to maintain a reasonable level of safety. NPP operators spend 20% of their time training on simulators. Anticipation is a matter of situation awareness that entails both perception and understanding of what is currently going on and what should happen next (Endsley, 1996).

Interaction with a safety–critical system is strongly grounded into action skills. A good operator is intimately “integrated” into his/her control room, i.e., controls and observable parameters are harmonious extensions of his/her cognition, the same as his/her arms or legs. It takes training and experience, as well as affordant devices. Unfortunately, there are failures coming from both the plant hardware and operators. Therefore, recovery is also an important asset that needs to be mastered. Failure recovery is a matter of diagnostic and problem solving. The nuclear industry learnt how to categorize failures and appropriate recovery strategies and actions. This is the basis of operator training and operating documentation.

Operators use the background and knowledge they have in order to make proper decisions; unfortunately, incorrect situational awareness leads operators to perform the wrong functions. "If the operators had not intervened in that accident at TMI and shut off the pumps, the plant would have saved itself. They [the designers] had thought of absolutely everything except what would happen if the operators intervened anyway. So the operators thought they were saving the plant by cutting off the emergency water when, in fact, they had just sealed its fate (PBS, 1999). The designers did not compensate for the emergent cognitive functions that occurred when the system is in operation.

Operator training has seen significant improvements. The operators are now licensed under stringent requirements that include a basic working knowledge of the physics behind the nuclear reactor core. In order to become an operator of a current U.S. NPP, an applicant must pass a physical examination, as well as a Generic Fundamental Examination provided by the NRC, which tests the applicant’s knowledge of reactor theory, thermodynamics, and the function and operation of the mechanical components of the specific system they are to be working on, whether it be a pressurized water reactor, or a boiling water reactor. Following this, they must then pass an additional certification test, which is regional and site specific, including a plant walk through, and a simulator demonstration under emergency situations (Nuclear Regulatory Commission, 2010).

The NRC has also implemented significant human factors programs including, but not limited to, new regulations. The US regulations for the industry are housed primarily under 10CFR. The rules for human factors in design are defined specifically in 10CFR55 and they are very extensive. Among some of the
documentation for review is the process of evaluations for taking Human Engineering (HE) into design, the Human Computer Interaction (HCI) guidelines, and how to perform task/activity analyses.

Engineers tend to understand the inputs and outputs of their specific systems, but rarely the interactions between them all. Thus, knowing your users is one of the most fundamental principles of HCD. Considering that people do not come in standard sizes, programs or products should not either. In the design process, human-centered designers and writers must appropriately understand not only the task at hand, but also the industry, the culture and the language.

4.4. Task-driven design for safety

We can address difficult design problems by asking for the inputs of system experts. By focusing on the cases where there is a lack of user experience whether because no user profile exists (revolutionary case) or actual existing users are not available or not easy to access (evolutionary case). Expertise should be used in conjunction with new ideas, and creative solutions. Here is a method that may be used for successful solutions.

There are three kinds of tasks that should be carried out concurrently:

- User requirement gathering (using the solutions that are provided earlier in the paper).
- Systematic use of ergonomic rules and usability heuristics.
- Creative design and prototype development.

The result of this mixed approach should be validated by a set of experts that includes end-users, domain engineers and human factors specialists. These three tasks should be iterated several times.

Involving users is not only good for the design and development of the product, but also for the acceptability of the product when delivered. Users are more likely to accept a design that they helped build because they take personal stake in the development and release of the system.

The complexity of the prescribed task directly relates to the workload of the user (Mogford et al., 1995). Thus, we must properly understand the induced roles on the user in the design phase of the system. We must take into account the expected tasks of each user during the entire design phase of each system. As the technology improves, the role of the human within the human–machine system is moving towards a supervisory role. This allows for detection, diagnosis, decision-making and then maintenance or reconfiguration. The primary performance factors are quantity, quality and safety (Millot, 2011). The only way to properly identify these roles prior to system release is through the use of usability testing and simulation based research of the system.

4.5. Organization-driven design for safety

Emergent functions are the outcome of complex systems. In addition to the subsystems, the interactions also define the properties of the overall system. Often times, these functions are not well identified until after the system is in use, as the designers do not always foresee the functions that a system may have. Given the natural evolution of systems, the emergent functions will establish out of practice, and requirements in order to further the general goal. In the nuclear industry, “operator workarounds” have been identified within the system and are often tracked within the control rooms. These workarounds have been identified as less efficient, often ineffective or time directly pertaining to the success of the task.

---

5 Operator Workarounds are shortcuts identified by the operators to compensate for a shortcoming within the system.
Limited to normal, abnormal and emergency, but can also be broken down further. A normal scenario will
be proceduralized. Procedures become more specialized, often times attributable to the Work Planning and Practices. In 1997 they completed the longest operational run of any steam driven plant, including fossil fuels, at 617 days of uninterrupted operation. The average number of unplanned reactor shutdowns has decreased by nearly 10-fold. In FY 2007, there were about 50 unplanned scrams among 104 operating plants compared to over 500 scrams in 1985 (NRC, 2009).

4.6. Situation-driven design for safety

The HCD of a safety–critical system involves the consideration of various situations that could be normal or abnormal. In normal situations routine operations can be proceduralized. Procedures work in any situation that can be anticipated. They are built from experience and technical knowledge. There are also abnormal situations that lead to abnormal procedures because they are very well known. The same holds for emergency situations. Issues arise when situations are not persistent enough to be considered routine or well known. Issues also arise when problem solving is required to determine the proper procedure for the situation. The NRC found that over 50% of the human performance issues within the industry were attributable to the Work Planning and Practices. That is to say that over half of performance deficiencies resulted from operators using the wrong procedure (NRC, Report Annual Report of the Effectiveness of Training in the nuclear Industry for Calendar Year 2009, 2010). As we become more automated, and the procedures become more specialized, often times we lead the operator to choose the incorrect procedure for the situation.

HITLS enable studying human operator performance, workload, situation awareness, and reliability. HITLS enable searching for emerging cognitive functions (new practices), and provide new and relevant information for licensing decisions. They enable both qualitative and quantitative measurements of various kinds of human factors that would not be possible in real life environments. An efficient approach for design is to involve the end users. Recognizing the importance for human in the loop design, in 2006, the International Atomic Energy Agency (IAEA) established standards for Experience feedback. “Operating experience is a valuable source of information for learning about and improving the safety and reliability of nuclear installations.” (IAEA, 2006). The IAEA outlines the process for handling the main elements of experience feedback, the process for screening of events, and the investigation and analysis of events. It also addresses handling corrective actions, recognizing emergent problems, disseminating this information, quality assurance, and reporting events.

The importance of designing for situation and context cannot be overstated, in the nuclear power industry; these conditions are referred to as normal, abnormal and emergency. The primary purpose of looking at context is to remove ambiguity from a situation and to address situational awareness. For example, a table can be something to set things on such as a kitchen table, or a table can be a spreadsheet, such as an excel table. By defining if we are referring to a kitchen, or excel, we set the context. In a NPP, this is not only limited to normal, abnormal and emergency, but can also be broken down further. A normal scenario will
not only involve daily routine operations, but also the process of a plant start-up or shut-down. In good design, the system and the user will understand the context of the situation.

4.7. Expecting the unexpected

The Fukushima Daiichi incident has again raised concerns over worldwide nuclear power. Following a 9.0 magnitude earthquake and associated flooding the plant experience partial core meltdown in three of the six units, as well as hydrogen explosions, and fuel pool fires. The sea walls were only built to withstand a 5.7 m tsunami, a far cry from the 14 m tsunami that ensued following the massive earthquake. In addition to the plant being knocked off the grid, the diesel generators were entirely flooded, making the process of removing the decay heat an overwhelming undertaking. Abnormal situations inject immense complexity into accident prevention. Though at the time of print the investigations are still underway, the Fukushima events remind us all of the existence of the black swan (Taleb, 2007). The root and proximate causes, as well as the severity of the outcome, are still to be determined from this event.

TMI, Chernobyl, Fukushima, and the Davis-Besse plants serve as a warning and example to all designers. The industry has a strong safety culture, and safety record. People understand the inherent risks involved in Nuclear power, but to the designers, operators and engineers, these plants were safe. However, they were not designed with the operator in mind. So much so that in the case of TMI, had operators not been present in the control room; the automations themselves would have saved the plant. The operators turned off the automations, because at the time, given their training, this was the correct thing to do.

5. Conclusion and perspectives

Designing for safety is a real issue that deserves extreme attention, reflection and practice. The current answer to protect people from the failure of safety-critical systems is grounded in the development of software-intensive systems. It is an easy way to generate protections but it may yield a vicious circle. The accumulation of software layers increases automation, and thus system complexity and consequently perceived complexity that in turn can generate new types of safety issues. Automation is the idea of removing complexities from the human, and applying them to the machine. By doing this, we add new layers of complexity as users roles change within the system from actor to manager. This is due to the fact that software-intensive system maturity is almost never reached. Software is always evolving and too often used as patches instead of an integrated solution. Safety-critical systems deserve clean and understandable solutions. Why?

We would like to insist on the importance of having “the right” high-level requirements in the early stages of design. This is a very important attribute of maturity. A series of endless modifications are always necessary to update an incorrect design. This is why it is important for formative evaluations to be considered during the entire design and development processes. Summative evaluation is performed at the end of the development of a product. Both types of evaluation may use the same kind of techniques such as interviews, brainstorming, data collection, the AUTOS analysis, and human-in-the-loop experiments. HITLS will also prove useful in understanding the emergent behaviors of the system.

We have seen in this paper that whether in the form of procedures or software automation, is often designed and developed as a response to incidents and accidents. In fact, automation must be an applied as a balanced system between procedures and software where human and machine cognitive functions are properly allocated. This is not possible without appropriate HITLS, formative evaluations and incremental prototyping. We need to make a distinction between the maturity of a product concept and a product that is not complete. Once a mature concept is found, it needs to be developed and finished, i.e., matured in the sense of the CMM. This does not remove the importance of user-centered continuous testing. Sometimes, a concept is not mature because the technology is not ready. Developing a new NPP concept involves a long life-cycle that is important to anticipate. Having the right high-level requirements is crucial. In this case, we are talking about the safety of the nuclear process during energy production through decommissioning.
We have seen that ultra-safe systems work most of the time with a very good passive stability. However, human operators’ vigilance tends to decrease when they do not have much to do. This is why human operators need to be kept in the control loop to maintain reasonable continuous situation awareness.

The design of new NPP will inevitably lead to the design and emergence of new jobs and life profiles. Technology and societies are constantly evolving, putting us in transient states with no hope to reach a steady state as it used to be. Therefore, anywhere passive socio-cognitive stability cannot be reached by design or organizational set-up, safety has to be thought as active stability where people are accountable and in control. Nevertheless, safety needs to be considered and handled by the technology, people and organizations.

The primary points to make are:

1. Nuclear Power is complex, thus the control systems are also complex.
2. Complexity must be understood and handled well in order to design for safety.
3. Procedures are human automation, much as software is machine automation.
4. Automation leads to complexity, thus must be an understood, balanced system between procedures, software and human operators’ appropriate skills and knowledge. By always attempting to identify emerging properties of the human–machine system, function allocation will improve on both human and machine sides, and therefore reduce the problem of procedure accumulation.
5. Complexity can be reduced during design by using the AUTOS pyramid model as an analytical and exploration support.
6. Human-in-the-loop-simulations help to understand emergent behaviors, improving the design of a system.
7. A well designed system is designed through the entire life cycle, from day one to decommissioning with the first phase reaching maturity and a continuous experience feedback regulation process.

6. Acknowledgements

We would like to thank Lynn Weaver, S. Mostafa Ghaasiaan, Jon Nowlin and Melissa Mierzwa who provided astute advice towards improving the quality of this paper.

Dr. Guy Boy is University Professor and Director of the Human-Centered Design Institute at Florida Institute of Technology. He has worked on design for safety for more than 30 years, mostly in the aerospace domain. He is currently the Principal Investigator of a 4-year AREVA research grant on the design of new nuclear power plant control rooms. He was the director of the European Institute of Cognitive Sciences and Engineering (EURISCO) for 16 years, where he organized two summer schools in 1995 and 2001 on human-centered automation and design for safety topic that gathered most recognized contributors such as René Amalberti, Saadi Lahlou, Erik Hollnagel, Edwin Hutchins, Thomas Sheridan, Kim Vicente, Earl Wiener and David Woods. Several references to their work are in this paper. In addition, he gave several tutorials organized by the HUMANIST European network of excellence. These tutorials enabled him to synthesize distributed knowledge and experience on design for safety.

Kara Schmitt is a Ph.D. Candidate with the Human-Centered Design Institute at Florida Institute of Technology. She is currently working on a dissertation focusing on functional allocation within the nuclear power domain. Ms. Schmitt previously worked as the lead structural engineer on Space Shuttle Atlantis where she was awarded NASA's Spaceflight Awareness award, as well as the Quest for Excellence Technical Achievement award for her work on the ARES 1-X Modal Testing. The operational lessons learned from this experience are currently being applied to the research in design.
7. References


PBS (Director). (1999). *Meltdown at Three Mile Island* [Motion Picture].


