Annals of Nuclear Energy 52 (2013) 125–136

Contents lists available at SciVerse ScienceDirect

Annals of Nuclear Energy

journal homepage: www.elsevier.com/locate/anucene

Design for safety: A cognitive engineering approach to the control and management of nuclear power plants

Guy A. Boy *, Kara A. Schmitt [1]

Human-Centered Design Institute, Florida Institute of Technology, 150 W. University Blvd., Melbourne, FL 32901, USA

ARTICLE INFO

Article history:
Available online 16 October 2012

Keywords:
Human-centered design
Cognitive engineering
Risk management
Automation

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.

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.

* Corresponding author. Tel.: +1 321 674 7631; fax: +1 321 674 7175.
E-mail addresses: gboy@fit.edu (G.A. Boy), schmittk@fit.edu (K.A. Schmitt).
[1] Tel.: +1 321 626 5323; fax: +1 321 674 7175.

0306-4549/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.anucene.2012.08.027

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 interactions. 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 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

[2] AUTOS means Artifact, User, Task, Organization and Situation.