Proceedings of the 13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, August 30 - Sept 2, 2016, The Sodoh Higashiyama Kyoto, Japan

Managing Risk by Design

Guy A. Boy

Florida Institute of Technology, Melbourne, FL 32901
USA (Tel: 321-506-5073; e-mail: gboy@fit.edu).

Abstract: Safety is a matter of Technology, Organization and People (the TOP model: Boy, 2013); it is a matter of human-systems integration (HSI). However HSI, especially when systems are life-critical, requires us to make a distinction between Technology-Centred Engineering (TCE) and Human-Centred Design (HCD). TCE leads to situations where human operators need to do most of this integration at operations time. Conversely, HCD is deliberately based on HSI at design time. Of course, the HCD approach does not remove the need for human operators' expertise and experience at performance time, but provides solid foundations for HSI at systems delivery. In this paper, I will use the metaphor of the Orchestra to describe the HCD approach, where human operators (musicians) share the same frame of reference (music theory), have integrated tools and tasks (music instruments and scores) coordinated by HCD specialists (composers), are coordinated at performance time by technology knowledgeable and skilled leaders (conductors), and are knowledgeable about the recipients of their work (the audience). Making and playing a symphony involves many risks that need to be managed. These risks can be managed by design, using the TOP model, where human-centred designers create, test and certify technology, corresponding organization and people's jobs involved. Creativity is not only deliberate; it is often a matter of discovery of emergent properties of what is being created. For this reason, human-in-the-loop simulations and formative evaluations are mandatory. This paper will explain HCD processes that support managing risk by design.

Keywords: Risk, Life-Critical Systems, Human-Centred Design, Technology-Centred Engineering, Human-Systems Integration, Technology, Organizations, People.

1. INTRODUCTION

"Whatever is well conceived is clearly said, and the words to say it flow with ease." This citation of Nicolas Boileau[1] (1636-1711) clearly illustrates what any engineering designer should be able to do. Indeed, a crucial part of design work consists in transforming a purpose into a consistent and meaningful story that will clearly define a new system. First, the problem must be clearly stated. Second, solutions should be elicited from experts. Third, these solutions should be iteratively tested. Fourth, a converging solution should be selected. Fifth, this solution should be delivered. In this paper, I propose a framework that supports a design team to share and build knowledge about experience in possible uses of a life-critical system. In other words, risk is managed at design time through modelling and simulation that enable the involvement of appropriate actors and the generation of both constraints and opportunities. Such an approach of risk management by design requires understanding the shift from Technology-Centred Engineering (TCE) to Human-Centred Design (HCD).

During the 20th century, machines were engineered and manufactured based on tangible hardware. Structure was the basis of design. Drawings were performed on paper in the way architect would do. These drawings required very skilled people trained in descriptive geometry. They led to tracing templates onto metal, wood or any other hardware material. Hardware was finally processed toward manufacturing and production of final products. Computer-Aided Design (CAD) and Computer-Aided Manufacturing (CAM) became the automated version of these processes of design, tracing and manufacturing. We moved from tangible industrial steps to virtual engineering and manufacturing. We built washing machines, cars, airplanes and other things that involved more or less complicated designs. In addition, during the last three decades of the 20th century, electronics and software penetrated mechanical parts and assemblies. This evolution contributed to the implementation of artificial functions into machines. We typically called corresponding processes and additional products "automation." Many jobs based on manual skills became obsolete. New jobs were created based on CAD-CAM for structural design and production, as well as automatic control and computer engineering for functional design and production. More globally, industry was based on linear sequences of processes going from hardware development (the structure account) to software integration (the function account).

Even if industrial evolution was continuous and took a few decades, since the beginning of the 21st century, we observe a totally different trend, i.e., going from early integration of functions and structures in virtual environments to the concretization of tangible products. Almost everything designed today is done on computers. We have a tremendous number of software applications that enable us to draw, model and simulate not only virtual structures but also their

[1] Nicolas Boileau (1636-1711) was a French writer and poet who produced this famous citation in *L'Art Poétique*: "Ce qui se conçoit bien s'énonce clairement, les mots pour le dire arrivent aisément."

functions. From the beginning, we have the opportunity to investigate, better understand and design human-systems integration (HSI). In other words, we can assess the articulation of human and technology functions and structures during the early stages of design. Therefore, issues are less "automation" than function allocation and, more recently, tangibility of virtual products.

Life-critical systems, such as aircraft, spacecraft, medical systems and nuclear power plants, have more specific HCD requirements. Systems failures and human errors may lead to serious repercussions. Risk analysis methods were developed for a long time, on both the system side and the human side, but not based on HSI. This paper addresses the difficult issue of "managing risk by design," based on the industrial shift described above.

2. WHAT DO WE MEAN BY RISK?

We always face risk taking, even when we try to manage risk! Why? In any life-critical situation, best behaviour is taking precautions. Best risk takers (i.e., the ones who are still alive!) spend huge amount of time preparing themselves to potential risky situations. They know that routine lead to complacency, and consequently to potential wrong reaction to "unexpected" events. Life-critical systems that lead to risky situations require full involvement and engagement of the various key actors. Risk is taken and managed at various stages of product life cycle, from design to decommissioning. Let's concentrate on risk management at design time.

Risk is typically modelled mathematically as a product of the probability of a risky event to occur by its consequence severity. This definition is very useful and is commonly used in industry. Corresponding formulas are used for all parts of a product individually and incrementally assembled for integrated parts and the overall product itself. However, if this kind of risk definition is usually valid for technology, it does not work when people are involved. Why? This is because people involved in the use of a system induce a tremendous amount of complexity that is difficult and most of the time impossible to model. Of course, when a system is developed, operational procedures are defined. It is expected that users will follow these procedures, but experience shows that people do not always do it. Procedure following discrepancy is not only a matter of discipline (i.e., respect of operational procedures); it is also a matter of context (i.e., procedures are typically context-sensitive and may not be valid in some unexpected contexts).

Risk analysis methods are based on task descriptions. It is then crucial to understand the **distinction between "task" and "activity."** The task is what is prescribed to be done by human operators using systems begin designed. The activity is what human operators effectively do. During the 20th century, it was only possible to carry out risk analyses based on task descriptions. "Surprises" were discovered at operations time (i.e., when real activity was observable.) Today, it is possible to observe activity at design time using modelling and human-in-the-loop simulations (HITLS). We have very advanced models and simulations capabilities that enable to include human operators during design.

Therefore, investigating risk using analytical methods is only one part of the coin: the task side! Risk also needs to be understood by observing human operators' activity. **Analytical methods need to be associated with experience-based methods**. Up to now, experience feedback was only possible when systems were fully developed. We then need to develop methods based on HITLS. Consequently, managing risk by design is a matter of prototyping, formative evaluations and agile approaches. Figure 1 shows several loops going from testing results of design solutions to HITLS, integration into existing environment, prototype and design rationale.

Indeed, it takes time to reach **maturity of technology**, as well as **maturity of practice**. What is the difference between early and current HITLS technology? During the 1990s, we already developed and used cockpit simulators at design time in the commercial aircraft industry. However, it was very difficult and long to modify parts. We did not have the level of software flexibility that we have nowadays. Consequently, we often made decisions based on what we could do at design time. Today, it is possible to modify software in a few hours, and therefore it is possible to generate all loops of Figure 1 effectively and rapidly. In other words, it is possible to carry out agile design processes (i.e., develop and integrate software – and often hardware – in a few weeks, run tests and redesign quickly after.)

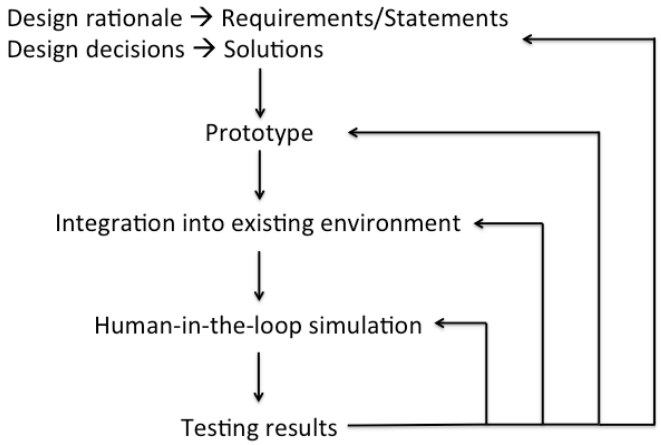

Fig. 1. Agile formative-evaluation-based process.

Risk management then becomes a matter of knowledge and experience. Consequently, such knowledge and experience should be integrated at design time. When designing a new concept, we do not have any experience on its use in the real world, and we need to create appropriate conditions to test it.

3. THE SFAC MODEL

As already introduced, we now have virtual environments based on digital models that can be simulated as pieces of software. For example, an aircraft model can be fully developed as an integrated piece of software related to a physical cockpit, which in turn can be used to develop HITLS (i.e., involving pilots performing scenarios in a real world close to their real world as much as possible).

New systems can then be virtually designed and tested based on integration of their functions and structures, which can be decomposed into sub-functions and sub-structures. Each function and structure can be described in an abstract way and a concrete way. The SFAC model (Structure/Function versus Abstract/Concrete) supports collaborative system design. It is presented in Figure 2. The design team then collaboratively generates four types of things: declarative knowledge (i.e., abstract structures); procedural knowledge (i.e., abstract functions); static objects (i.e., concrete structures); and dynamic processes (i.e., concrete functions).

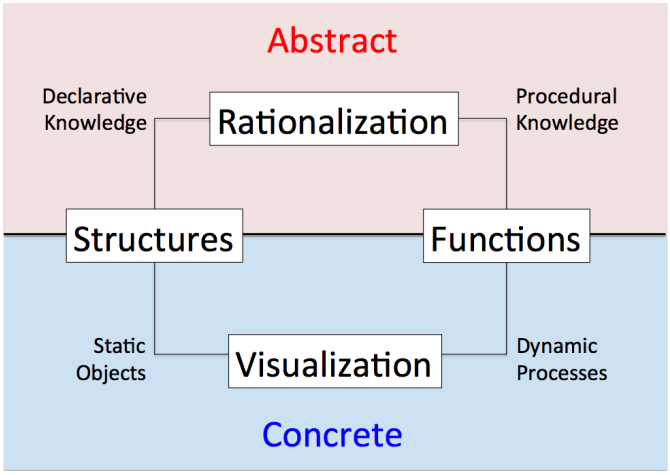

Fig. 2. The SFAC Model.

The abstract part is a knowledge base representing the rationalization of the system being designed. This rationalization can be represented by a set of concepts related among each other by typed relationships. This kind of representation can be called ontology[2], semantic network or concept map. It can take the form of a tree hierarchy in the simplest case, or a complex concept graph in most cases.

At design time, the concrete part is commonly represented using CAD software, which enables the generation of 3D models of the various components of the system being designed. These 3D models include static objects and dynamic processes that enable the visualization of the way the various components being designed work and are integrated together. Later on during the design and development process, these 3D models can be 3D printed and lead to more graspable appreciation of the components being built and their possible integration. Testing occurs at each step of the design process by taking into account concrete parts together with their abstract counterparts (i.e., their rationalization, justifications, as well as various syntactical and semantic relationships that exist among them.

The SFAC model is typically implemented as a mediating space that design team members can share, collaboratively modify and validate. This mediating representation supports incremental tangibility assessment and maturity reaching. SCORE is such as tool (Boy et al., 2016). Such a tool was developed using the latest information technologies (e.g., software for modelling and simulation, computer game generation and rationalization development).

Tangibility can be defined twofold. Something is tangible when it is **graspable** in the **physical** sense, but also **believable** in the **figurative** sense (e.g., an idea or a concept that cannot be grasped by the mind).

4. DESIGN CARDS

A great way of managing risks at design time is to involve all actors dealing with the system being designed, and making them aware of risks involved. As already explained, the provision of HITLS enables discovering seriousness of risky situations repercussions. This situation awareness issue is better emphasized when people involved can explicitly visualized and document such situations/repercussions patterns. SFAC supports the design team to **document the design process and its solutions**. The concept of active design document (ADD), initially developed for traceability purposes for safety-critical systems, is useful for the rationalization of innovative concepts and incremental formative evaluations (Boy, 2005).

The **design card** (DC) concept supersedes the ADD concept. Like in the ADD, CD includes a rationalization space and an activity space. The activity space provides dynamic visualization capabilities that enable the manipulation of systems being designed and developed. Instead of ADD's task space and the evaluation space, design cards have a structure space and a function space. ADD emphasized HFE issues only; design cards emphasize human-systems integration as a whole concept. This is why structure and function are put to the front. Structure space emphasizes the multi-agent declarative perspective (i.e., the HCD view of modern design), which can be denoted system of systems in systems engineering. Function space emphasizes both physical and cognitive function allocation and support storytelling (i.e., scenario-based design) from the start of the design process.

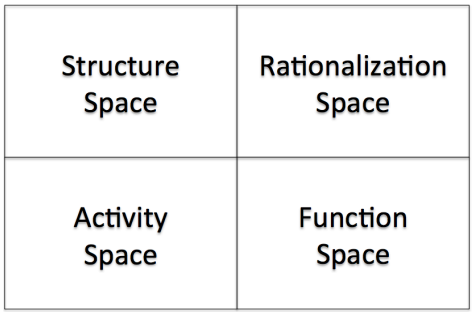

Fig. 3. Design Card (DC).

As ADD, DC supports the design history of the system being designed. Several versions of a DC are incrementally generated and refined. These versions can be traced at any time by anyone in the design team. This feature increases

[2] In philosophy, ontology is the study of what there is, what exists. It is "what the most general features and relations of these things are" (Hofweber, 2011).

inter-subjectivity in the design team (i.e., mutual understanding among the design team members).

A design card (DC) is defined by four entities (Figure 3):

- A rationalization space where the various components of the system being designed (SBD) are described in terms of design rationale, integration and requirements; this space includes declarative and procedural descriptions and statements (e.g., creation date, design rationale, requirements).
- An activity space where the current version of the SBD is displayed; it includes static and dynamic features; this space enables SBD manipulation (e.g., evaluation, scenarios, criteria).
- A structure space where the various components and their inter-relations are formally and declaratively described as systems of systems (e.g., visualization of components).
- A function space where the various functions of the SBD are described in terms of procedural knowledge and dynamic processes involved; this space includes qualitative and quantitative physical and cognitive models (e.g., procedures, checklists and technical explanations).

A given DC presents the state of the design of a TIS at a given time for a given design team member (DTM). It is formally represented by DC (t, DTM_i), where t is time and DTM_i is the design team member i (could be a person or a group of persons).

A DC enables designers to describe the various components of a system and the integrated whole in the rationalization space, display and manipulate them in the activity space, describe and use the navigation and control features in the operational space, and fill in the evaluation space as required after assessment of the system being designed.

Using DCs support solving several problems, such as geographical spread-out of experts of these groups, speed of technology evolution, high personnel turnover, and lack of documentation of the design process. DC generation happens during design. When DCs are documented regularly, they do absorb very little time of the design process. This additional time is compensated by a gain of time due to shared situation awareness of the entire design team. DC quality contributes to the quality of design.

Each DC (t, DTM_i) corresponds to a version of the system being designed and developed. Each time design management has a design review meeting at time t_1 (Figure 4), all DTMs analyse the work done by each DTM and create a synthetic ADD (t_1, DT), where DT is the whole design team. DCs are like scores that musicians use to play a symphony in an orchestra, with the peculiar difference that, unlike scores, DCs are being incrementally defined to get a sound symphony in the end of the design process.

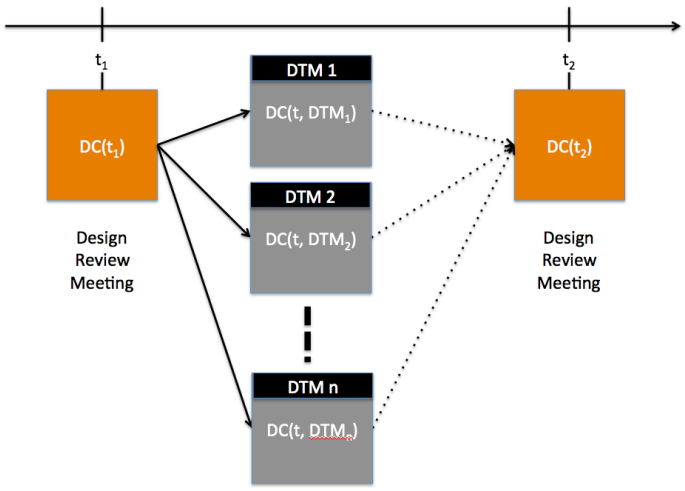

Fig. 4. Design cards generation.

After a design review meeting at time t_1, each DTM_i works on the premises of DC (t_1, DT), and produces their own DC (t, DTM_i) until the next design review meeting is organized at time t_2, where a new DC (t_2, DT) will be produced from the integration of all active design documents created and/or modified by each design team member during the time interval [t_1, t_2].

Each DC is stored into a design database, and can be retrieved at any time by any member of the design team (although some restrictions could be implemented and applied if necessary). Various DC traceability mechanisms can be implemented.

5. COLLABORATIVE WORK

Collaborative work is a crucial activity in a HCD team. Shared situation awareness (SSA) is a key issue in LCS design. SSA has been, and still is, studied at operations time (Stanton et al., 2006), but it requires more attention at design and development times. People may make errors because they are not aware of the current state of the design process. We should provide solutions to answer the following questions: Is DTM_1 aware of current actions and productions of DTM_2 at any time? Is DTM_1 aware of what DTM_3 did at some point in time on the same topic he/she is currently working on or a similar one? How can we create and maintain the best SSA in the design team? As already mentioned above, the SCORE CSCW system was developed to support SSA. A SCORE overview is provided in Figure 5.

SCORE uses components and procedures models. It is implemented using a web based application mechanism, which allows secure and trusted communication via VPN (Virtual Private Network). In addition, effective search mechanisms provide the necessary means to pull appropriate information when needed. It would also be nice to have the appropriate information pushed to the front so potential users are aware of its existence. In both cases, context-sensitive information should be available at any time.

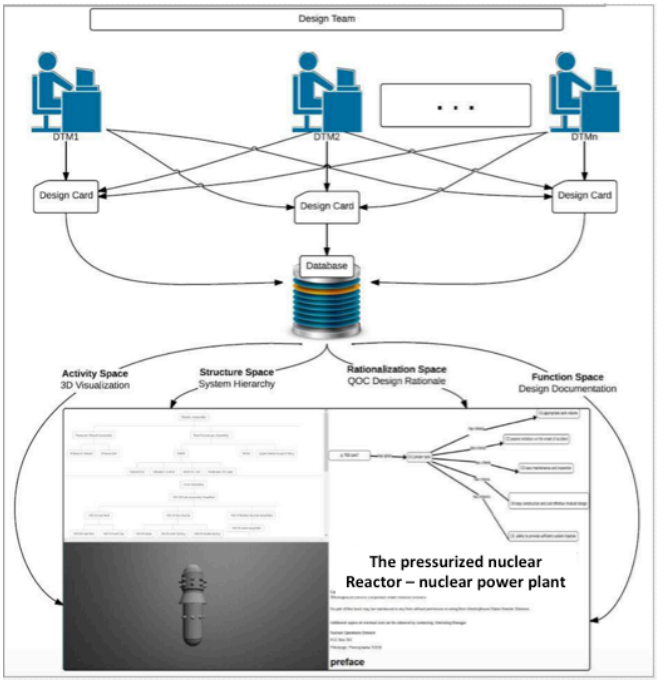

Figure 5. SCORE system overview.

For this reason, at all stages of the design process, any DTM needs to know salient reasons that pushed other people to design systems the way they are. When I carried out an exhaustive study on traceability within a large aircraft manufacturing company (Boy, 2001), I found out that traceability is not only information retrieval, but also deals with awareness that potential knowledge exists somewhere, and finally when this knowledge is found, it must be understood correctly.

6. DISCUSSION

As already said, HCD goes from purpose to means. Therefore, the first thing to do is to get an initial purpose that tells us what we want to do, knowing that it will evolve with the product being incrementally developed and used. Design team members need to work on meaningful things. Meaning is coming from both rationalization of design rationale (setting up tasks) and formative evaluations (figuring out activities). This is not only important for developing a great product, but also create and keep a soul in the project. Indeed, design team members need to clearly perceive and understand what they are doing and why, and they need to project themselves into the future to "see" how the product they are developing can be used.

I^2S-LWR[3] design team included about 30 people coming from different backgrounds (e.g., mechanical, nuclear, electrical and computer engineering). The SFAC model supported the HCD approach. A computer-supported cooperative work (CSCW) system, called SCORE, enabled the rationalization of the various decisions made during the design process and facilitated collaboration among the various DTMs. We also structured the design rationale capture using the QOC[4] formalism (Boy et al., to appear).

Mainstream engineering traditionally focuses on delivering technologically working products. HCD focuses on engaging people and creating value. Designing static objects, such as an architect would do, is the first step to capture user experience (i.e., user's activity). There are many CAD tools that enable designers to produce virtual 3D static objects, such as Dassault Systemes's CATIA, PTC's ProE, Siemens's NX and AutoDesk Inventor. Visualization of concrete objects provides intuitive anticipation of possible activities (i.e., user experience) and impacts on the real world (i.e., emergent properties).

Easy manipulation of 3D static objects provides designers with capability of testing configurations, also called declarative scenarios. They can construct and deconstruct objects, as well as assemble them among each other and disassemble them. For example, a nuclear reactor 3D model, visualized on a screen, can be decomposed into components, such as core assemblies, to geometrical dimensions of rods[5]. This is made possible through the link between visually tangible objects and their formal representations in the corresponding design rationale. Typically, the user typically selects a visually tangible object, gets a description pop-up window that enables selection of sub-components or attributes.

This kind of modelling and simulation capability provides design team members with endless trial-and-error possibilities. In addition, when this graphical capability is connected with a design rationale generator, it tremendously increases the production of meaningful tangible interactive objects. This design rationale tool can be implemented as an annotation mechanism on top of the 3D-static-objects visualization. This kind of feature provides meaningful interactivity with the objects being designed, and enables traceability among the various versions of these objects as well as connectivity among the various components of the systems being designed.

Visually tangible objects enable designers to immediately capture salient features that either confirm design choices or suggest modifications. Design is an iterative process. Visualization suggests confirmations or modifications; design rationale enables rationalization and deeper calculations. Linking visually tangible objects to their abstract descriptions (i.e., design rationale) enables designers to proceed with a

[3] The Integral Inherently Safe Light Water Reactor (I^2S-LWR) nuclear project is funded by the U.S. Department of Energy, in collaboration with several universities and industry partners including Georgia Tech, University of Michigan, Florida Institute of Technology and Westinghouse.

[4] QOC is a method that enables design rationale visualization of various design questions, possible options (or solutions) and criteria used for the choice of the best options (MacLean et al., 1991).

[5] This example illustrates a human-centred design student project carried out at the Human-Centred Design Institute of Florida Institute of Technology in the Fall 2014, involving the following graduate students, Saad Almesalm, Gopal Jani, Nicholas Kasdaglis, Joan Savage, Neha Suri, Golnoosh Torkashvand, Ruthvik Adloori, and Joseph Torkaman.

convergent process when the design team is proactive, competent and collaborative.

6. CONCLUSIONS

The holistic approach to HCD should be better-denoted cognitive-function-based design, or activity-centred design (Norman, 2005). The cognitive function paradigm provides an explicit representation to what is commonly implicitly done in design. It enables us to rationalize both deliberative (i.e., a priori defined) and emergent (i.e., discover when the socio-technical system is put at work) cognitive functions. The concept of cognitive function encapsulates the concept of agent. People can be called human or natural agents, which use natural and artificial cognitive functions. TISs are artificial agents, which also use natural and artificial cognitive functions.

The claim that **people adapt to technology** is not enough to eliminate **tests** that lead to technology improvement, and more generally incremental design of natural and artificial cognitive functions. This orchestration of incrementally generated cognitive function networks is the key for harmonious human-systems integration. Orchestrating cognitive and physical functions is crucial in HCD, where a concept map of natural and artificial physical and cognitive phenomena needs to be developed; and cooperation/coordination rules needs to be discovered and effectively used. It is like an orchestra where the musicians would have scores that would evolve with time to fix the harmony of the symphony being developed (Boy, 2013). As an example, the increasing density of air traffic strongly suggests that aircraft cannot be controlled the way they are now, i.e., centralized control. Moving toward **decentralized management** of aircraft in dense air traffic can be compared to a flock of birds, and therefore new types of cognitive functions would need to be discovered and implemented. Such implementation can lead to specific human learning/practice and/or development of tangible interactive systems on-board aircraft that automatically assist pilots in separation assurance and collision avoidance. Therefore risk management by design not enables solving existing safety-critical problems, but also contributes to suggesting emergent innovative solutions.

REFERENCES

Boy, G.A. (2016). *Tangible Interactive Systems*. Springer, U.K.

Boy, G.A., Jani, G., Manera, A., Memmott, M., Petrovic, B., Rayad, Y., Stephane, A.L. & Suri, N. (2016). Improving collaborative work and project management in a nuclear power plant design team: A Human-Centered Design approach. *Annals of Nuclear Energy*. Elsevier, ANE4864.

Boy, G.A. (2013). *Orchestrating Human-Centered Design*. Springer, U.K.

Boy, G.A. (2005). Knowledge management for product maturity. Proceedings of the International Conference on Knowledge Capture (K-Cap'05). Banff, Canada. October. ACM Digital Library, New York, USA.

Boy, G.A. (2001). Organizational Memory Systems. Plenary Paper, *Proceedings of IFAC-HMS 2001*, Kassel, Germany.

Hofweber, T. (2011). Logic and Ontology. *Stanford Encyclopedia of Philosophy*. (retrieved on September 13, 2015) http://plato.stanford.edu/entries/logic-ontology/#DifConOnt.

MacLean, A., Young, R.M., Bellotti, V.M.E. & Moran, T.P. (1991). Questions, Options, and Criteria: Elements of Design Space Analysis. *Human-Computer Interaction*, Lawrence Erlbaum Associates, Inc., Volume 6, pp. 201–250.

Norman, D.A. (2005). Human-Centered Design Considered Harmful. *Interactions*, 12. *4, (July + August), pp. 14-19. Also available at* http://jnd.org/dn.mss/human-centered_design_considered_harmful.html.

Stanton N.A. et al. (2006). Distributed situational awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics* 49, 1288–1311.