Introduction

The general topic of the workshop on Human Centered Processes (HCP) concerns the problem of understanding and modelling human expertise in industrial settings. The particularity of the HCP approach is focusing on natural cognition and considering as fundamental the contribution of cognitive sciences.

This workshop follows a series of previous Human Centered Processes conferences. The first conference was held in Brest (France, 1999) and was mainly devoted to the application of cognitive approaches in various fields of process control and other complex industrial problems usually managed by Operational Researchers. The second conference, held in Luxembourg (2003), was dedicated to distributed decision and man-machine cooperation. Both these conferences put special emphasis upon cognitive models of decision making. The third conference (Delft, Netherlands, 2008), focused on the human actor and software agent collaboration in safety and time critical systems-of-systems.

The original group has been involved for years in the special working group "Human Centered Processes" belonging to EURO, the Association of European Operational Research Societies. Now the group is reorganizing, and a new emphasis is given to complementing traditional Operational Research practices with cognitive issues, often neglected in industry but relevant, such as knowledge engineering, discovery of rules, updating and maintenance of rules, rule-based systems, models designed to aid complex human decision-making.

This idea poses a challenge over the organization and design of contemporary manifacturing, against the idea of systems totally detached from human cognitive procedures. Theaim is to make possible a joint scientific and industrial research intended for analyzing and modelling advanced manufacturing, information, or actionsystems that are strongly dependent on a balanced integration between human and machine skills: collaborative working, cooperation, user adapted interaction, etc. This original attitude has recently met the development of Cognitive Sciences, with the particular emphasis on a multidisciplinary approach, suggesting the idea of helps realizing Human Centered Designs and Technologies. The complementarity of cognitive and traditional approaches provides a great number of additional dimensions that allow to design and analyze more complete and complex systems.

HCP 2011 is not devoted to a main topic, in order to encouragethe maximum scientific exchange between practitioners in different research topics. These Proceedings collect accepted papers, that have been sorted into different topics or lines of debate fields. The variety of themes does not mean unrelatedness. On the contrary, all of these contributionsare attempts to show how human cognitive processes are essential in designing and supporting complex systems, where the amount of required automatization cannot overcome the underlying cognitive patterns based on human strategies.

A first aspect analyzed in the workshop contributions is the relationship between designer and customer, directly concerning the problem of user support. In particular **M. Norese, E. Liguigli** and **C. Novello** deal with the problem of understanding users' real needs through a direct cooperation between them and designers, and with the creation of cognitive maps to summarize information obtained from different sources. In administration and management new perspectives are opened by using graphs that take into account the role of different contexts in influencing different lines of reasoning: **H.Tahir** and **P. Brézillon** apply this approach in database management, with the goal of developing an intelligent assistant for administrators; **X.Fan, P. Brézillon, R. Zhang** and **L. Li** use a similar approach for a workflow management system applied to the scientists' decision making processes.

Another area of interest relates to improving safety, particularly by understanding the limits of interfaces. **P. Carvalho, J. Gomes** and **M.Borges** examined safety problems arising in nuclear power plants, where most critical processes depend on control room operator's interaction with interfaces. They used cognitive task analysis to examine problems in operator training and propose a new interface design that helps the operator in incident management. **C. Calefato, L. Minin** and **R. Montanari** are interested in the car driving behavior. Their research is aimed at improving the safety systems by constructing a user model that allows to predict the driver's behavior. The aim is to build intelligent vehicles by combining the study of vehicle technology with the performance of drivers. **A. Chialastri**'s contribution describes a new designer perspective in aviation: going beyond the classical "human friendly" design, to create systems better defined as "pilot friendly", that take into account the pilots' experience. In this perspective, a shift is required from "reactive" interventions after accidents to "proactive" actions.

Three papers concern support systems for health and welfare. **S. Apolonio, L. de Deus, M. Borges** and **A. Vivacqua** carry out a study using Google Docs as a collaborative writing aid for hearing impaired people. They found that this tool is effective in helping the interaction with non-impaired people. **M-A. Sujan, S. Pozzi, C. Valbonesi** and **C. Ingram** deal with the general ability of keeping control when adverse events happen. They describe some resilient behaviors in hospitals, using namely personal trading, shared awareness and seeking help. Using qualitative analysis of reports, their study shows that those behaviors are based on personal experience and initiative, often without enough support from the organization. **F. Cavallo, M. Aquilano, G. Anerdi, M.C. Carrozza, P. Dario** and **A. Greco** present the results of a European project aimed at developing a roadmap for intelligent systems providing personal assistance to elder and impaired people. Originally proposed ubiquitous computing techniques ("Ambient Assisted Living") are extended to cognitive aspects.

Along with risk prevention some proposals focus on human error analysis. **E. Schreuder** and **T. Mioch** address the problem of cognitive lockup, a difficulty that arises when a person is faced with a series of disturbing situations. They put this problem in relation to time pressure, and carried out an experiment that shows how an almost finished task may delay the consideration of another more urgent task. **M. Frixione** and **A. Lieto** are concerned with a classical problem, highly problematic in AI and Cognitive Engineering, namely concept representation. They use evidence from experimental psychology to sketch an architecture for a "dual process" concept representation in formal ontologies taking into account the heterogeneity of human cognition.

On the whole, contributions to this workshop show that a human-centered approach is not only useful but necessary in joint academic and industrial research, where analyzing and modelling advanced manufacturing, information, or action systems is strongly dependent on a balanced integration between human and machine skills. In this endeavour, the multidisciplinary nature of cognitive sciences (particularly Psychology, Philosophy, and Computer Science) plays a crucial role in the necessary integration of technical perspectives with ones concerning people.

Elena Carrea, Alberto Greco, Carlo Penco

Making context explicit towards decision support for a flexible scientific workflow system

Xiaoliang Fan (1, 2, xiaoliang.fan@gmail.com), Patrick Brézillon (1, patrick.brezillon@lip6.fr), Ruisheng Zhang (2, zhangrs@lzu.edu.cn) and Lian Li (2, lil@lzu.edu.cn)

(1) LIP6, Box 169, Université Pierre et Marie Curie
4 Place Jussieu, Paris 75005 France

(2) School of Information Science and Engineering, Lanzhou University
222 South Tianshui Road, Lanzhou 730000 P.R.China

Abstract

Scientific workflow (SWF) system is a specific workflow management system applied to science arena. For years, SWF systems are widely applied to many applications, namely in physics, climate modeling, drug discovery process, etc. However, current SWF systems face the challenge to adapt the flexibility and lack of decision support for scientist. We believe the major reason for the failure is due to do not make context explicit. We propose a solution to introduce contextual graphs (CxG) in the four phases of the SWF lifecycle, each of which is expressed in a standard format, including a case study in virtual screening. Contextual graph allows to model scientists' decision making processes as a uniform representation of knowledge, reasoning, and of contexts, so that scientists are closely involved in each phase of SWF lifecycle to maximize the decision support. Finally, we conclude and highlight that using CxG is the key human-centered process for SWF systems.

Introduction

Scientific workflow system liberates the computational scientists from burden of data-centric operations to concentration on their scientific problems (Altintas et al., 2004; Goble et al., 2007). However, it is not yet satisfied, considering that computational science (Roache, 1998) is always reproduced in a flexible and exploratory pattern. Consider virtual screening (Chen & Shoichet, 2009) for example, the choice of one software over others depends much on contextual information that are highly specific of the situation at hand, and where, when, how and by whom the scientific workflow is executed. Thus a strong and sustainable decision support is urged for scientists to transfer hypotheses to discovery.

Workflow flexibility becomes a critical challenge to deal with intermittently available resources, execution failures, and to support human-centric decision-makings. However, identifying how scientists make decisions to address workflow flexibility is a very complicated issue. The ways of scientists make their decision vary from one another: (1) based on their past experience considering successful or failed ones; (2) inherited from the best practices within science communities; (3) from the observed intermediate results; and (4) just follow their own distinguished way. Various approaches (Zhang et al., 2008; Courtney, 2001; Tabak et al., 1985) are proposed to get user involved to describe their decision making processes. Normally in such applications, a decision making (e.g., choose methods, change parameters, re-design the experiment) is measured by a decision node in workflow design accompanying with a numerical value (e.g. IF the variable is greater than 5, THEN execute the activity A, ELSE execute activity B; WAIT for 2 minutes to execute activity C). However, scientific discovery is by nature a knowledge-intensive one (van der Aalst et al., 2005) that scientists' decisions rely not only on data and information available, but also on a learning process in which user's preference, knowledge, and situation are captured to adapt the human-centered processes.

Such challenges mentioned above become an obstacle when scientists are making adaptive decisions to deliver new outcomes with fresh data and its context (Fan et al., 2010). Brézillon and Pomerol (1999) define context as "what constrains the resolution of a problem without explicit intervention in it". We believe that the main reason for this failure is largely due to the lack of context management in an explicit way. In this paper we propose four ways of making context explicit in scientific workflow, by introducing contextual graph to in the four phases of scientific workflow lifecycle. Representing and making "context" explicit in SWF system would provide sustainable decision supports for scientists by formalizing their research, strategies, and customization information, where elements of knowledge, reasoning and contexts are represented in a uniform way.

Hereafter, the paper is organized in the following way. Section 2 introduces the four phases of the scientific workflow lifecycle. Section 3 investigates the possibility of integrating contextual graphs to the four phases of scientific workflow lifecycle through a case study in virtual screening. Section 4 discusses previous works on workflow flexibility in order to point out what is reusable while problems remain to support decision-makings in a flexible scientific workflow system. The general conclusion and future work in Section 5 closes the paper.

Scientific Workflow Lifecycle

Scientific workflow lifecycle is coming from workflow lifecycle (van de Aalst & van Dongen, 2003; Gil et al., 2007; Deelman & Chervenak, 2008). It normally starts from the scientific hypotheses (Beaulah et al., 2008; Tadmor & Tidor, 2005; Claus & Johnson, 2008) to reach a specific experimental goal, which includes four phases (see Figure 1):

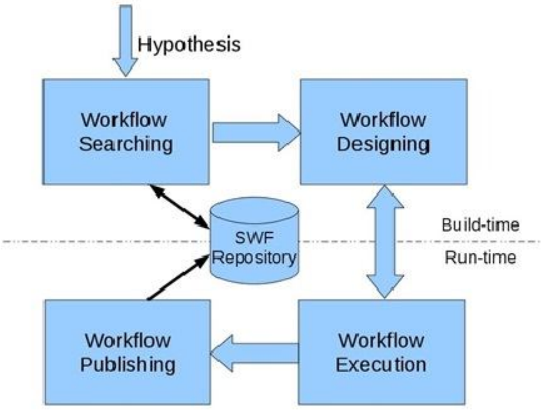


Figure 1: SWF lifecycle

- *Workflow Searching*: before initiating a brand new workflow designing, scientists get used to firstly consult a public SWF repository for searching previously published workflows (Wroe et al., 2007). Once found, it would be easy to reproduce the pre-existing workflow to constitute a new one. Workflow searching results of sharing SWF considered with its context of use. The more shared SWFs are taken place in the SWF repository, the more accurate the searching result would be.
- *Workflow Designing* is then initiated for constructing a workflow model (Ludascher et al., 2009). An abstract workflow model will firstly be designed, in which scientific tasks and their execution orders, as well as data and its dependencies will be described. Secondly, the phase involves the mapping from abstract workflow to concrete/executable workflow where the required resources are selected. By mapping the workflow instance onto the available execution resources, an executable workflow is created for the next phase.
- *Workflow Execution* is the enactment of executable workflow by a *workflow engine* (Deelman & Chervenak, 2008), in which input data is consumed and output data is produced (Tan et al., 2010). Workflow engine follows the order of tasks and their dependencies defined in the workflow model. It is common to re-execute the workflow iteratively, considering the evolutionary changes of workflow model (e.g., in workflow design, adding or skipping tasks, and altering task dependencies) or momentary changes of a running workflow instance (e.g., making local decisions in response to a special situation, alter decision after analysing observed intermediate result, reporting exceptional cases).
- *Workflow Publishing* is a post-execution phase for scientists to interpret workflow results (Tan et al., 2010; Ludascher et al., 2009) and to publish the SWF in its context of use (Wroe et al., 2007; Deelman & Gil, 2006). Depending on the workflow outcomes and analysis results, the original hypotheses or experimental goals may be revised or refined, giving rise to another round of workflow design/execution in an iterative manner. Furthermore, it must then be facilitated to publish the workflow on a repository, so that SWF could be archived for re-use later.

Figure 1 shows the relationship among each phases of scientific workflow lifecycle: hypotheses arrive as keywords to search pre-existing scientific workflow in SWF repository; then scientist begin to design the workflow model and maintain the mapping from an abstract workflow to a concrete one; workflow execution phase enacts the workflow model on available resources according to data and control dependencies; if a change is encountered, there is an iterative process to re-design the workflow model as well as re-execute the workflow instance; if executed successful, scientist will publish the workflow in the SWF repository for the sake of reproduction in the research communities.

Current studies (van de Aalst & van Dongen, 2003; Deelman & Chervenak, 2008) on SWF lifecycle generally result in the weakness to manage the workflow changes and exceptions. We believe that the major failure is due to do not make context explicit in the SWF systems.

Make Context Explicit in SWF Lifecycle

Representing and making context explicit in SWF system is a challenge that could promote a SWF system more flexible and enhance its intelligence to facilitate effective decision-makings. In this section, we discuss managing contexts explicit throughout the four phases of the SWF lifecycle, each of which is described using a standard format including: *motivation*, *realization approach*, *example*, and *discussion*.

The example is represented in the Contextual graphs formalism (Brézillon, 2005) through a case study entitled "Virtual screening research on avian influenza H5N1 virus", which aims to find dozens of drug candidates for H5N1 virus (He et al., 2008), by docking 7.7 million small molecules separately on H5N1 protein (Chen & Shoichet, 2009). Figure 2 shows a docking example, which binds a molecule (ZINC12050767) to a virus protein (H5N1 PA_C Polymerase, known as Bird flu) through the Dock 6.2 software. Virtual screening could be considered as millions of docking procedures on the PA_C protein.

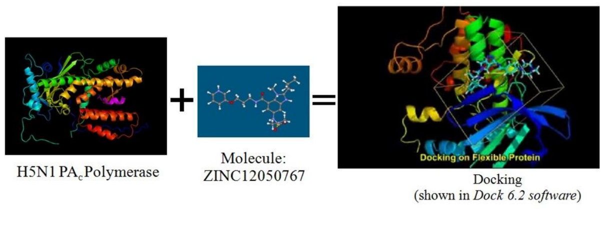


Figure 2: Docking example

The application is not only a time-consuming workflow application in which intensive computing is expected to be performed by docking software, but also a very flexible one that there is no unique solution for each computing because they vary from each other on selecting docking software. For example, scientists should identify the context in which the experiment is organized as a scientific workflow. According to the current focus and context, they link a specific resource (e.g., software, database, and instrument) with the workflow to realize a specific task. The concept of human-centered process is particularly relevant in such domains.

Figure 3 provides the definition of the elements in a contextual graph (actions, contextual elements, sub-graphs, activities and temporal branching). A more complete presentation of this formalism and its implementation can be found in (Brézillon, 2005).

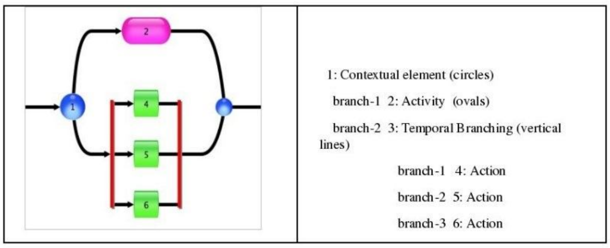


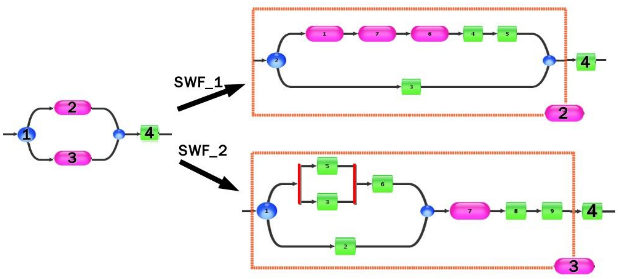
Figure 3: Elements in Contextual graph

Workflow Searching

Motivation: Before the workflow design, context behaves as an interface to determine which SWF should be chosen from a library of SWFs, or a SWF repository. In this case, a scientist plays a role as a context provider to guide the choice of the right SWF model according to current focus and context at hand, so as to largely match what the scientific hypotheses indicate.

Realization approach:

- Scientist firstly searches a SWF from a SWF repository, using keywords which could best describe their hypotheses and are coherent with the context at hand.
- If the pre-existing SWF is exactly what they want, the scientist could skip workflow design phase and just replace with their own parameters for workflow execution directly.
- Otherwise if it is similar to their needs, slight modifications will be carried out shortly in the workflow design.



Context graph: virtual screening on protein PAc
1: Is the protein rigid or flexible?
Rigid 2: Activity: perform first rigid screening
Flexible 3: Activity: perform second flexible screening
4: analyze the result

Figure 4: (Left) Contextual graph of virtual screening on H5N1 protein; (Right) Choosing one SWF from two SWFs (SWF_1 and SWF_2)

Example: In Figure 4 (Left), CE1 is a contextual element (blue circle with number 1). The instantiation of the CE1 (Is the protein rigid or flexible?) leads to the generation of two scientific workflow instances in Figure 4 (Right): one is SWF_1 (i.e. value of CE2= "Rigid"), and the other is SWF_2 (i.e. value of CE2="Flexible"). In the application, if scientists want to do a rigid virtual screening, "rigid" will become a keyword when performing the searching. Thus, SWF_1 will be selected. Similarly, SWF_2 is chosen when searching for a "flexible" screening. As a result, CxGs act as an interface to make decisions to choose SWF from the SWF repository.

Discussion: It is normal to expect nothing from the repository, scientist could move to the next phase to start workflow design from scratch.

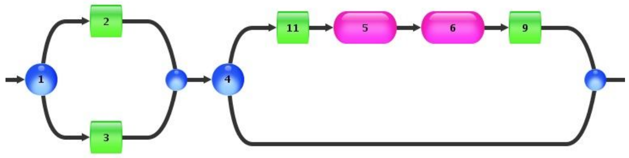
Workflow Designing

Motivation: During workflow design, a certain degree of freedom is given to the user to execute a workflow by offering multiple alternative execution paths. Classical workflow systems reduce the degree of flexibility by offering powerful design constructs (e.g., start, if/else, repeat until, parallel execution, end), in which decision-making is always measured by a decision node accompanying with a numerical value. However, human decision is so complex that a numerical decision is less descriptive than a simple question. As a result, we describe execution paths of workflow in contextual graphs (CxGs) which model contextualized information (CEs) and their dependencies. In a contextual graph, the most appropriate execution path could be selected from those encoded during the execution time to address the context at hand.

Realization approach:

- Firstly, it is necessary to know all the current instances of the CEs at the moment of the application of the workflow. An instantiation is the value that a contextual element can take for a specific instantiation of the focus at hand.
- Then, a group of contextualized information is generalized as a set of CEs.
- CEs are then formalized in a contextual graph by their dependencies. The contextual graph is ready for the workflow execution, when a SWF instance corresponds to a specific execution path under the instantiation of context. In CxG, the execution path is a sequence of actions, connected by the instantiation of the selected contextual elements.

Example: In Figure 5, a scientist designs the workflow of protein preparation as a contextual graph with a set of contextual elements (CE1 and CE4) and their execution dependencies. The possible execution paths are controlled by the value of each contextual element. For example, the instantiation of CE1 (i.e., value of CE1= "Yes") and CE4 (i.e., value of CE4= "Yes") leads to the execution path of "1→2→4→11→5→6→9".



Contextual graph: protein preparation (old)
1: Can you find the protein by yourself?
Yes 2: download it from "Protein Data Bank"
No 3: ask for help until you get the protein
4: Do you need to do "protein preparation"?
Yes 11: enter parameters during "protein preparation"
5: Activity: remove unrelated molecules
6: Activity: add hydrogen and charge
9: store the protein prepared in the database
No

Figure 5: Contextual graph: protein preparation (old)

Discussion: Describing a completely set of all possible execution paths during workflow design might be either undesirable or impossible (Schonenberg et al., 2008). For example, a certain number of possible execution paths are unknown before execution. As a result, late-modelling (Han et al., 1998) could enable to make sub-model dynamically defined during execution.

Workflow Execution

Motivation: Scientists frequently re-execute the scientific workflow by adding or ignoring portions of workflow realized at design time. Context should support the assembling of SWF components, which must be recompiled each time when a new context arrives (i.e., a contextual element takes a new instance). As a result, a new execution path, or even a new contextual graph will be inserted or removed when SWF evolves along with its context.

Realization approach:

- Each time a new instantiation of a CE occurs, the contextual graph is re-executed, and the SWF is recompiled for generating a new SWF instance for execution.
- If the scientist wants to re-design the workflow by adding or ignoring portion of SWF, they first stop the current workflow execution.
- Then, a new group of contextualized information, including the information representing the workflow changes, should be generalized as a new set of contextual elements.
- If a CE with the following activities/actions is added or ignored, a new contextual graph is produced to address the new focus.

Example: Figure 6 is inherited from Figure 5. During the execution phase, the scientist finds something wrong with the intermediate result, because he doesn't take into account whether the protein is flexible or rigid. So he decides to stop the current execution and re-design the experiment. As a result, a new contextual element CE7 (Is it a rigid or flexible screening?) is added. When the value of CE7 is "flexible screening", Activity13 (Activity: optimize the protein) is invoked as a new SWF component. Furthermore, the contextual graph is updated along with the change of CEs, and it is necessary to record such update in a knowledge base for the sake of workflow sharing, which will be discussed in the next section.

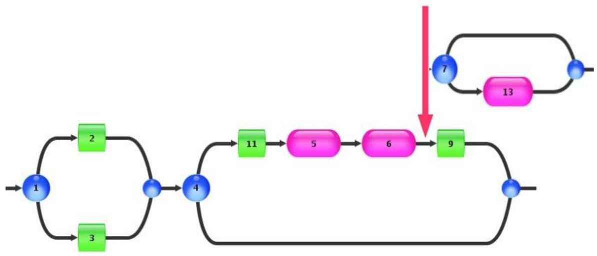


Figure 6: Contextual graph: protein preparation (new)

Discussion: It would be a risk of incoherence between the running workflow instance and results. For example, when you made a decision two minutes ago and the contextual graph chooses an execution path for the workflow. But later, right before the workflow execution, a new context arrives to urge the adaptation of a new contextual graph.

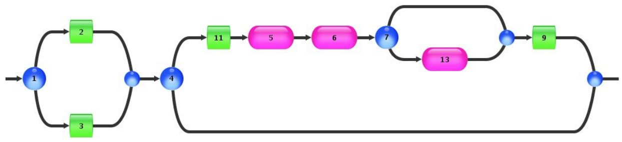
Workflow Publishing

Motivation: If executed successfully, the scientist then try to analyse the results generalized by workflow execution. Type of result analysis includes: 1) evaluate data quality (e.g., does this result make sense?), 2) examine execution traces and data dependencies (e.g., which results were "tainted" by this input dataset?), 3) debug runs (e.g., why did this step fail?), or 4) simply analyse performance (e.g., which steps took the longest time?). After the result analysis process, it is possible to re-design and re-execute the workflow iteratively until the new context is addressed. Incremental knowledge acquisition should be proceeded to make contextual graph growing to be more efficient. Furthermore, one of the motivations what scientists are counting on SWF is the sharing, reproduction, transformation, and evolution of the "old" SWF to be a brand "new" one. It is expected to enable sharing of SWFs according to their contexts of use. In this circumstance, the context defines the status of the knowledge and also maintains the relationship between different kinds of knowledge.

Realization approach:

- A SWF repository is build up to document workflows with their contexts of use.
- When workflow is re-executed, the contextual graph is adapted incrementally to trace the workflow flexibility. Once a new contextual graph is generated, add it as a new scenario to SWF repository.
- Conscientious users might partition the workflow into coherent fragments and publish them.

Example: Once a contextual element is modified, a new CxG is created to address the new focus and its context. Drawn from Figure 6, Figure 7 shows a new contextual graph to be added in a SWF repository for future sharing with other scientists.



Contextual graph: protein preparation (new)

1: Can you find the protein by yourself?
- Yes 2: Download it from "Protein Data Bank"
- No 3: Ask for help until you get the protein

4: Do you need to do "protein preparation"?
- Yes 11: Enter parameters during "protein preparation"
  - 5: Activity: remove unrelated molecules
  - 6: Activity: add hydrogen and charge
  - 7: Is it a rigid or flexible screening?
    - Rigid
    - Flexible 13: Activity: optimize the protein
  - 9: store the protein prepared in the database
- No

Figure 7: Contextual graph: protein preparation (new)

Discussion: Encourage sharing of scientific workflow with its context, would make it as a complementary of paper-based publications. In such a case, scientific workflow would be archived along with paper-based publications. However, the quality of sharing data and workflow becomes a new question.

Summary

Contextual graphs are a formalism of representation allowing the description of decision making in which context influences the line of reasoning (e.g. choice of a method for accomplishing a task). The advantage of contextual graphs relies on that: (i) CxGs provide naturally learning and explanation capabilities in the system; and (ii) CxGs allow a learning process for integrating new situations by assimilation and accommodation. In short, the notion of context is made explicit during the four phases of scientific workflow lifecycle by contextual graphs. Contextual Graphs formalism has been already used in different domains such as medicine, incident management on a subway line, road sign interpretation by a driver, computer security, psychology, cognitive ergonomics, etc.

Related Works

Various approaches, such as BPEL (Zhang et al., 2008), UML (Courtney, 2001), Petri-net (Tabak et al., 1985), are proposed to address the issue of workflow flexibility by getting user involved in representing decision-making. Applications (Yu et al., 2005; Hey et al., 2009) have proven the significance of current systems to handle numerical decision-making as control-flow functions, such as “wait 30 second, and then proceed the next task”, “if the value is greater than 5 then execute the task_A, else execute the task_B”. However, it becomes an obstacle to manage the common but important decisions, such as “are you satisfied with the result?” and “do you need to do the protein preparation again”, which is more comprehensive for scientists.

Context has been considered as a key element to support decision making in human centered processes for a long time (Brézillon, 2003; Brézillon, 2010). To address a coherent formalism of context, Sowa (1984) proposes conceptual graphs with their mechanisms of aggregation and expansion. Then, Sowa (2000) introduces a way to manage the context in conceptual graphs. Brézillon (2005) presents a simpler formalism of Contextual Graphs (CxGs) for representing context. Compared with other approaches, CxGs formalism is good at describing decision making in which context influences the line of reasoning.

In the implementation level, a number of applications exist for preparing formal representation of context. McCarthy (1993) formalizes contexts as formal objects, and the basic relation is ist(c,p). It asserts that the proposition p is true in the context c, where c is meant to capture all that is not explicit in p that is required to make p a meaningful statement representing what it is intended to state. Formulas ist(c,p) are always asserted within a context, i.e., something like ist(c', ist(c,p)): c': ist (c, p). Sharma (1995) gives a list of desirable properties for contexts in a formal language and distinguishes four approaches for formalizing contexts: (1) incrementing arity; (2) variation on implication; (3) modal operator forms; and (4) syntactic treatment. Based on McCarthy's work on context logic, Farquhar et al. (1995) present an approach to integrating disparate heterogeneous information sources.

In Table 1, we compare various approaches to model decision making in workflow, as implementation of “Exclusive Choice workflow pattern” (van de Aalst & Hofstede, 2003).

Table 1: Comparison of various implementations of “Exclusive Choice workflow pattern”

Approach	Decision Element	Decision Value	Decision Type
BPEL (Zhang et al., 2008)	<if>, <pick>	Condition	Numerical value
CxG (Brézillon, 2005)	Contextual Element	Value of CE	Any value
UML (Courtney, 2001)	Decision Node	Condition	Numerical value
Petri-net (Tabak et al., 1985)	Exclusive choice	Arc expression	Numerical value

By comparison, Contextual Graphs plays an equivalent role to other approaches for representing decision making. Furthermore, the advantage of contextual graphs embraces: (1) multiple representations of decision making, not only with a numerical value, but also with any kind of answers to questions to get scientists involved in a local decision-making process; (2) it is directly readable (e.g. generally something as “If the contextual element C has the value V1, thus use method M1, and with the value V2 use method M2”); and (3) it is very easy to have an incremental growth of a contextual graph by addition of contextual elements and branches for representing practices developed by users and not yet known by the system.

Conclusion

The human-centered processes must be considered at a global level to deal with the user, the task at hand, and the context in which the task is accomplished. Take a flexible scientific workflow for example, scientists could not handle the transferring from hypotheses to discovery in the SWF system without taking into account the context.

We propose a solution to introduce contextual graphs in the four phases of SWF lifecycle, each of which is expressed in a standard format, including a concrete example in the area of virtual screening. In our application on virtual screening, we use contextual graphs to model the decision making processes of scientists as a uniform representation of knowledge, reasoning, and contexts. As a result, scientists are closely involved in each phase of SWF lifecycle to maximize the decision support received from the system.

We believe that all of data, information and knowledge should be invoked, assembled, organized, structured and situated according to the given focus, and finally be formulated as the chunk of professional knowledge for scientists to maintain their research sustainability.

The extension of our work includes the development of a prototype interface between scientific workflow system and contextual graphs. Representing and making "context" explicit in SWF system by contextual graph would enhance workflow flexibility by formalizing scientists' research, strategies, and customization information, where elements of knowledge, reasoning and contexts are represented in a uniform way.

Acknowledgments

This work is supported by grants from National Natural Science Foundation of China (90912003, 60773108, 90812001, 61011130212), Centre national de la recherche scientifique (Researcher exchange project with NSFC 2010), and Région Ile-de-France (CP10-201), and by scholarships from China Scholarship Council (2008618047), and Égide (690544G).

References

Altintas, I., et al. (2004), Kepler: an extensible system for design and execution of scientific workflows. *Proceedings of 16th International Conference on Scientific and Statistical Database Management* (pp. 423-424). Danvers, MA: IEEE Computer Society.

Beaulah, S., Correll, M., Munro, R., Sheldon, J. (2008). Addressing informatics challenges in Translational Research with workflow technology. *Drug discovery today*, *13(17-18),* 771-777.

Brézillon P., and Pomerol J.C. (1999). Contextual knowledge sharing and cooperation in intelligent assistant systems, *Le Travail Humain*, 62 (3), 223-246.

Brézillon, P. (2003). Focusing on Context in Human-Centered Computing. *IEEE Intelligent Systems*, *3(18)*, 62-66.

Brézillon, P. (2005). Task-realization models in contextual graphs. In Dey, A.K., Kokinov, B., Leake, D., and Turner, R. (Eds), *Proceedings of Modeling and Using Context: 5th International and Interdisciplinary Conference* (pp. 55-68). Berlin :Springer Verlag.

Brézillon, P. and Pomerol, J.-Ch. (2010). Framing decision making at two levels. In Respicio, A., F. Adam, G. Phillips-Wren, C. Teixeira & J. Telhada (Eds.), *Bridging the Socio-Technica Gap in Decision Support Systems- Challenges for the next Decade*. Amsterdam: IOS Press.

Chen, Y., & Shoichet, BK. (2009). Molecular docking and ligand specificity in fragment-based inhibitor discovery, *Nature Chemical Biology, 5 (5)*, 358-364.

Claus, B., and Johnson, S. (2008). Grid computing in large pharmaceutical molecular modeling, *Drug discovery today*, *13(13-14)*, 578-583.

Courtney, J.F. (2001). Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS. *Decision Support Systems, 31(1)*, 17–38.

Deelman, E. and Chervenak, A.L. (2008). Data Management Challenges of Data-Intensive Scientific Workflows. *Proceedings of the 8th IEEE International Symposium on Cluster Computing and the Grid* (pp. 687-692), Lyon: IEEE CS Press.

Deelman, E. and Gil, Y. (2006). Managing large-scale scientific workflows in distributed environments: Experiences and challenges. *Proceedings of the Second IEEE International Conference on e-Science and Grid Computing* (pp. 144), Washington: IEEE Computer Society.

Fan, X. et al. (2010). Context-oriented scientific workflow system and its application in virtual screening. In Respicio, A., Adam, F., Phillips-Wren, G., Teixeira C., and Telhada J. (Eds.). *Bridging the Socio-Technical Gap in Decision Support Systems- Challenges for the next Decade* (pp. 335-345). Amsterdam: IOS Press.

Farquhar, A, Dappert, J, Fikes, R and Pratt, W. (1995). *Integrating information sources using context logic* (Tech. Rep. KSL-95-12). Palo Alto, California: Stanford University, Knowledge Systems Laboratories.

Gil, Y., Deelman, E., Ellisman, M., Fahringer, T., Fox G., Gannon, D., Goble, C., Livny, M., Moreau, L., Myers, J. (2007). Examining the Challenges of Scientific Workflows. *Computer*, *12(40)*, 24-32.

Goble, CA., et al. (2007). myExperiment: social networking for workflow-using e-scientists. *Proceedings of 2nd workshop on Workflows in support of large-scale science* (pp. 1-2). NY: ACM.

Han, Y., Sheth, A., Bussler, C. (1998). A Taxonomy of Adaptive Workflow Management. *In Workshop of the 1998 ACM Conference on Computer Supported Cooperative Work*. Seattle, Washington, USA.

He, X., et al. (2008). Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature*, *454*, 1123–1126.

Ludascher, B., Altintas, I., Bowaers, S., Cummings, J., Critchlow, T., Deelman, E., De, R.D., Freire, J., Goble, C., Jones, M., Klasky, S., McPhillips, T., Podhorszki, N., Silva, C., Taylor, I., and Vouk, M. (2009). Scientific process automation and workflow management. In Shoshani, A., and Rotem D. (Eds), *Scientific Data Management: Challenges, Existing Technology, and Deployment, Computational Science Series*, chapter 13. Chapman & Hall/CRC.

McCarthy, J. (1993). Notes on formalizing context. In Bajcsy, R. (Ed.), *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence* (pp. 555-560). San Mateo, California: Morgan Kaufmann.

Roache, P.J. (1998). *Verification and validation in computational science and engineering*. Albuquerque, NM: Hermosa Publishers.

Hey, T., et al. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, Washington: Microsoft research.

Schonenberg, H., Mans, R., Russell, N., Mulyar, N., van der Aalst, W.M.P. (2008). Process flexibility: A survey of contemporary approaches. In: Dietz J., Albani A., and Barjis J. (Eds), *Advances in enterprise engineering I*. Berlin: Springer-Verlag.

Sharma, N (1995). *On formalising and reasoning with contexts* (Tech. Rep. 352). Brisbane, Queensland: University of Queensland, Department of Computer Science.

Sowa, J.F. (1984). *Conceptual Structures: Information Processing in Mind and Machin*e. Reading, MA: Addison Wesley Publishing.

Sowa, J.F. (2000). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks Cole Publishing.

Tabak, D. and Levis, A.H. (1985). Petri Net Representation of Decision Models. *IEEE Transactions on Systems- Man and Cybernetics*, *15(6)*, 812–818.

Tadmor, B., and Tidor, B. (2005). Interdisciplinary research and education at the biology-engineering-computer science interface: a perspective (reprinted article), *Drug discovery today*, *10(23-24)*, 1706-1712.

Tan, W., Missier, P., Foster, I., Madduri, R., De Roure, D., Goble, C. (2010). A comparison of using Taverna and BPEL in building scientific workflows: the case of caGrid, *Concurrency and Computation: Practice and Experience*, *22(9)*, 1098–1117.

Yu, J. and Buyya, R. (2005). A Taxonomy of Scientific Workflow Systems for Grid Computing. *In special issue on Scientific Workflows, SIGMOD Record*, *34(3)*, 44-49.

van de Aalst, W.M.P., van Dongen, B.F., et al. (2003). Workflow Mining: A Survey of Issues and Approaches. *Data & Knowledge Engineering*, *47*, 237–267.

van de Aalst, W.M.P., Hofstede, A., et al. (2003). Workflow Patterns. *Distributed and Parallel Databases. 14(3)*, 5–51.

van der Aalst, W.M.P., Weske, M., and Grunbauer, D. (2005). Case Handling: A New Paradigm for Business Process Support. *Data and Knowledge Engineering 53(2),* 129–162.

Wroe, C., Goble, C., Goderis, A. et al. (2007). Recycling workflows and services through discovery and reuse. *Concurrency and Computation: Practice and Experience*, *19(2)*, 181-194.

Zhang, H., et al. (2008). Extending BPEL2.0 for Grid-Based Scientific Workflow Systems, *Proceedings of IEEE Asia-Pacific Services Computing Conference* (pp. 757-762), Yilan: IEEE CS press.

Integrated use of linear programming and multiple criteria methods in an engineering design process

Maria Franca Norese (mariafranca.norese@polito.it), Ersilia Liguigli (ersilia.liguigli@polito.it)
Chiara Novello (chiara.novello@polito.it)

Department of Production System and Economics ï Politecnico di Torino
Corso Duca degli Abruzzi, 24 ï 10129 Torino, Italy

Abstract

In an aeronautical sector company, where engineering design process steps and activities are developed, a decision aiding methodology was required to support the processes.
We have proposed the integrated use of linear programming and multiple criteria methods, which can be used to orient the conceptual design of alternative functional and physical solutions and to cope with complex design problems.
Linear programming has been used in relation to some case studies, to generate design alternatives that satisfy the set of the initial requirements, while multiple criteria methods have been proposed to interact transparently with the client, in relation to some aspects that a linear programming model cannot include, to evaluate, compare and select alternatives in order to identify and formalize new expectations that the first solutions have not be able to resolve. The iterative use of the two approaches, in a cyclic procedure of mutual learning, allows the requirements to be defined more clearly and a final satisfying solution to be reached.

Introduction

Engineering design is an iterative decision-making process which is developed to devise a component, product, process or system that meets the customerôs needs (Eggert, 2005). A collaboration with the Marketing and Business Development Department (MBDD), of a company that designs and produces aircraft for civil and military use, has allowed us to understand their main conceptual design activities (Norese et al., 2008a; Norese & Liguigli, 2009) and to propose an integrated use of linear programming and multiple criteria methods in order to aid engineering designers (by means of a logical framework that was used also in Alenia Spazio, 2004 and Norese et al., 2008).

An aircraft is a complex system, but it can also be seen as a component of an even more complex structure, a System of Systems (SoS), in which different systems communicate and work together to achieve specific targets. In an SoS, integration and synergic work may vary from a simple collaboration, in which the single components work alone, to a situation in which the single components are not able to work in an autonomous way, when extrapolated by the SoS.

The MBDD supports product development by managing the relationship with the client in the initial engineering design process phase. The clientôs needs have to be identified, in order to decide whether and how a specific legacy aircraft has to be updated in relation to these needs, or to understand what kind of aircraft has to be designed or (at least partially) re-designed, in order to guarantee its integration in the new SoS that the client perceives as possible or essential for the future.

The client's involvement in the initial phase of the design process is analyzed in the literature in "front end" models of the product development process (see, for instance, Smith & Reinertsen,1992; Reinertsen,1999). Some authors have focused on the concept phase of the process where, through the involvement of the client, it is possible to obtain meaningful improvements (Clark & Fuijmoto, 1991) and to resolve ambiguity and uncertainties in the customerôs requirements that may cause orientation difficulties (Smith & Reinertsen,1998).

In aeronautics, a partial and apparently limited re-design requires years of work (five years on average) and therefore uncertainty concerning the evolving nature of the clientôs requirements is normally present, with an evident impact on the engineering design process. The MBDD asked our research group for suggestions and methods, in order to improve the interaction with the client (who has to understand every step of the process and freely propose his point of view), to reduce time and guarantee quality of the results, which can be solutions and/or a better definition of the needs, objectives, priorities and future scenarios of aircraft use.

We proposed the integrated use of two kinds of Operation Research methods, in relation to some case studies, and the cyclic use of the method application outputs, as new inputs for the other method application (Belton & Stewart, 2002). Linear programming (LP) can be used to analytically define the constraints and aspirations of a client, generate the widest set of design alternatives that satisfy the initial requirements (admissible solutions) and calculate optimal solutions, in relation to specific objectives. Multiple Criteria Decision Aid (MCDA)[1] models can be developed and MCDA methods activated iteratively, in order to transparently interact with the client. Alternative solutions, produced by means of the previous LP application, can be analysed and evaluated, in relation to aspects that an LP model cannot include, such as the perception of a risk (of using a too innovative technology, or to generate new complexity in the future maintenance problems, and so on) in relation to a specific solution. Some clientôs requirements can be identified and formalized when a

[1] More details can be found on the Euro Working Group MCDA website: www.cs.put.poznan.pl/ewgmcda/

solution is not compatible with expectations that were not clear enough before the MCDA analysis and therefore not included in the LP model. At this point, new functional and organization limits may be included in the LP model and the consistency of each solution should also be tested for these new constraints, in the new cycle that it started, and so on until an acceptable solution is reached.

The first section of the paper focuses on the iterative nature of the engineering design process and offers a synthetic overview of the methods, theories and tools that are used by designers.

In the second section, the problem, as perceived by the MBDD, is presented and, in the third section, a set covering model is proposed for the generation of design alternatives.

In the fourth section, some multiple criteria approaches are described, in relation to the evaluation of design alternatives, and the integrated use of two methods is proposed to support communication with the clients, in order to better define their needs and expectations. The possible development of the procedure, in relation to more complex projects and decision contexts, is analysed in the conclusions.

Engineering design process

Several theories and various tools are proposed in engineering design to aid designers in different ways: to understand stakeholdersô needs, improve quality, address variability and uncertainty in the design process or generate alternatives for designers.

The engineering design process, as described by Eggert (2005), is structured in five steps: definition of the problem, gathering pertinent information, generating multiple solutions, analyzing and selecting a solution and, finally, testing and implementing the solution. A procedure of identifying and formally listing the customerôs requirements is usually present in problem definition, in order to define product functions and features. These activities are included in the first step of the described process, but in some cases problem definition is complicated and can be completed only when pertinent information is gathered. And generating and analyzing multiple solutions, with the involvement of the client and some areas of the enterprise, is a way to obtain relevant information on the product design and functional specifications.

Once at least the structural components of the design have been identified, above all with inputs from testing, manufacturing and marketing teams, the design team generates alternative conceptual solutions that are oriented in different ways to achieve predefined goals (i.e. requirements that have to be satisfied).

Considering costs, quality and risk, as the main selection criteria, the most promising alternatives are selected for a further analysis (Dean & Unal, 1992), which enables a complete study to be made of the solutions and elaboration of the final design specifications that best fit the requirements. A prototype is therefore constructed and functional tests are performed to verify and, when necessary, to modify the design.

In the conceptual phase of the design process, it may be necessary to go back to a previous step at any point in the process. The chosen solution may prove to be unworkable for different reasons and may require specification redefinition, new solution generation, the collection of more information or, in the worst situation, redefining the problem. This is a continuous and iterative process.

Several tools are commonly used to aid designers. Methodologies and theories that have been proposed in the literature, usually offer a more analytically rigorous support for engineering designers. *Concurrent engineering* may be the most practical methodology to improve the design process. The approaches that are most frequently suggested to obtain input from stakeholders in the design process are the Pugh Method (Pugh, 1990), Quality Function Deployment (Akao, 1997) and the Analytical Hierarchy Process (Saaty 1980; 1994), which always incorporate subjective judgments. Others are used to generate alternatives for designers, such as TRIZ (Altshuller, 1988) and the C-K Theory (Hatchuel & Weil, 2009).

Problem Statement

The problem definition step, in the engineering design process, is critical when the client has to face an evolving situation and cannot clearly communicate needs that are no well defined. This criticality is often present in aeronautics, where many years are required to create a new aircraft, but also to innovate some elements of a legacy system.

The problem definition step is developed in the MBDD by anticipating some activities that pertain to the successful steps of the process (as described in Eggert, 2005) and using them to acquire essential, but latent or fragmented, knowledge elements.

Clearly understanding the point of view of the client, at a functional level but also in organization terms, is essential to identify and structure the requirements that orient the design. The MBDD arrives at a complete problem definition through a procedure that involves the organization of a client in a comparative analysis of some promising draft solutions. These solutions are elaborated in the MBDD, in relation to general technical requirements, and then the strength and weakness elements of the solutions are discussed with the client or, more precisely, with some organization-client key actors (for example, a pilot or whoever is in charge of maintenance).

Even if the innovation is related to a single aircraft component, the future use of the aircraft in an integrated SoS has to be analyzed. Various types of aircraft, but also satellites and maritime or ground systems can be involved, in order to achieve an assigned target in missions of various kinds (i.e. military, civil or a combination of the two situations). Innovation is often required in order to specifically facilitate coordinated work and communication in the SoS.

The MBDD procedure includes two subsequent stages: in the first one, some ñfunctionally acceptableò solutions are identified or elaborated, in relation to the functionalities that are required. In the second stage, the clientôs attention is focused on these solutions, in order to evaluate the associated costs (which are not only monetary), their economic sustainability and specific benefits and risks, as proposed in (Office of Aerospace Studies, 2002). This analysis orients the elaboration of a better solution for the client, but at the same time defines the overall problem and identifies pertinent information and/or information sources. A representation of how the cyclic procedure develops is proposed in Figure 1,with indications on the main activities that are included.

In the last few years, some clients have required the use of an Operations Research tool, in order to facilitate comparisons of the solutions in a multiple criteria analysis. Having found the tool very interesting, the MBDD asked to our research group for a method to help the generation of ñinterestingò and acceptable solutions, in order to reduce time and guarantee the completeness of the acceptable solution set. We analyzed their use of the tool and the weak and strong points of their applications. We then proposed the integrated use of LP and MCDA models in a procedure that fits the MBDD approach to the problem, but also improves the interaction with the client, who can propose his point of view (in terms of limits of the solutions and opportunities that have to be stressed), in a simple but formal language, and who can almost immediately analyze all the new solutions that are consistent with the new vision.

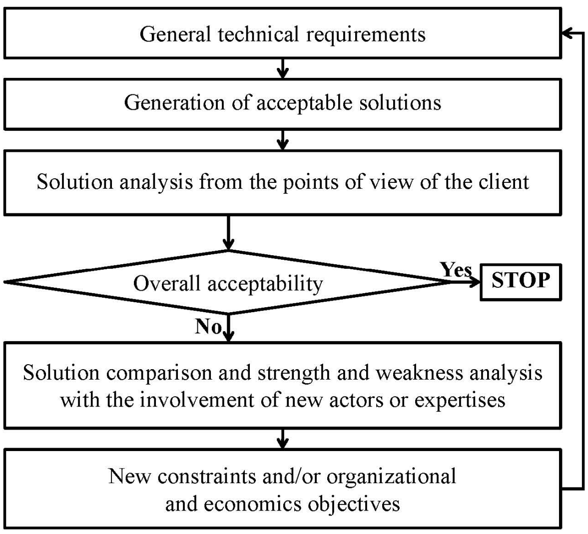


Figure 1: Cyclic procedure

Linear programming application

The request of a client can be very generic and the real needs are not always easy to understand. In order to reduce uncertainty, a request can be expressed in terms of mission types that the new system (or asset) has to face, as a component of an SoS.

From a technical point of view, an asset is a system that guarantees specific functionalities. The assets may be component parts of a single aircraft that have to be integrated to complete a mission, or to be integrated with other assets in other aircraft or in systems that operate on the ground. The assets may also be specific kinds of aircraft (or other resources) that have to be activated together in a specific mission. In all these situations, the integrated assets can be seen as an SoS and both the performances of the assets and the relationships between them allow the missions to be performed.

The purpose of the analysis is to define a mathematical model in which the variables are the different assets that can be activated to accomplish a mission. The functionalities that have to be guaranteed (or guaranteed at a required level), in relation to the ñnatureò of the mission, can become the constraints of the model.

The different objectives, in relation to a specific decisional problem, can be: minimize the costs, maximize the effectiveness, minimize the risks of a mission and so on. A combination of assets that is acceptable because it guarantees the Required Level of Performance (RLP) for each functionality (i.e. for each constraint of the model) becomes an admissible solution, which is called architecture (of the SoS). The optimal solution is an admissible solution that minimizes (or maximizes) the objective. If a single admissible solution does not exist, the need for a technological innovation (i.e. at least a partially new asset) is underlined. A new product, or an improvement in a legacy system, satisfies the clientôs needs if all the missions that the client had proposed to describe his needs can be faced with a minimum cost.

The problem can be represented by a linear programming model, if all the constraints and the objectives are linear functions. If there is only one objective, the oldest and most famous method of Operations Research, the Simplex method (Dantzig, 1963), can be used to obtain the optimal solution. If there is more than one objective, the multi-objective linear programming methods (Ehrgott & Wiecek, 2005) can be used.

At the start of the model setting, the assets that have to be included in the model and a list of functionalities, i.e. the constraints of the model, are defined in relation to the (generic or specific) request of the client and above all using the Universal Join Task List (UJTL) Report[2].

A complete list of about 720 functionalities, in terms of ability to perform a task, is proposed in the UJTL Report, in relation to the strategic, operational and tactical level of mission in a military context. The UJTL was developed for the U.S. Armed Forces, but it has been used by several other countries and international military organizations, such as NATO. The MBDD has structured and adapted the Report to facilitate its use with the clients. The MBDD synthesizes all the coordination, monitoring and controlling functionalities for military missions in the Mission Management macro functionality and Find-Fix-Track is the code that is used to indicate the set of functionalities which, at different levels of detail, allow the area of interest to be patrolled, in order to indentify and trace the target. Using this framework, **xx** main functionalities, that have to be guaranteed in a military mission, are always present as model constraints. When the mission requires a specific and not usual functionality or for non military missions, the UJTL Report is used directly as a check list.

If the adopted objective is to minimize the number of assets that have to be involved in the proposed missions, the mathematical problem can be re-formulated in terms of a *set covering* problem, which consists in finding the

[2] Report available on the www.dtic.mil website

minimum number of service centers (in our model, the assets) so that the request for each service (the guarantee of a required level of a specific functionality) is covered (Tadei & Della Croce, 2001).

In this mathematical model, the performance p_{ij} of the j-asset for the i-functionality is compared with S_i, the RLP that has to be guaranteed for the i-functionality, in order to define the covering matrix $[t_{ij}]$, in which the elements t_{ij} are equal to 1, if p_{ij}Ó S_i, or equal to 0 otherwise.

The *set covering* problem can be formulated in the following way:

Min × x_j $\quad j = 1,$é $, m$
× $t_{ij}x_j$ Ó q_i $\quad i = 1,..., n \qquad x_j=\{0,1\}$

where x_j has a value of 1 when the asset is included in the solution (which in this case, is an SoS architecture), otherwise it is equal to 0.

The value of the redundancy, for each functionality with redundancy (i.e. a critical functionality that requires more than one asset that is able to satisfy this task, in an SoS architecture), is equal to q_i. For the others, q_i is equal to 1.

We used Xpress-MP, version 2007 (Mosel 2.0.0, IVE 1.18.01, Optimizer 18.00.01), produced by DASH Optimization, to treat models with a single mission or multi scheduled missions that are included in the model. For a multi missions model with 18 variables and 210 constraints, the application has provided six admissible and three optimal solutions in 0.15 seconds.

The model structure and the linear programming application to the problem were tested in relation to some previous military cases, where the solutions and their characteristics were well known for the MBDD. We spent a great deal of time defining and modifying the constraints, in order to have a better fit of some specific requirements, but the immediate calculation of the solutions facilitated convergence towards a good model. The same procedure was then applied to a new case, in relation to the surveillance of a critical sea canal. The model development and PL application were accepted by the MBDD as effective steps of a procedure that can support communication with the client.

At this point attention was focused on the tool that should be used to understand why a solution is not adequate enough.

Multiple criteria approaches

The U.S. Air Force Center of Expertise for Analyses of Alternatives (Office of Aerospace Studies, 2002) suggested a multiple criteria approach in which all the aspects that are related to the effectiveness have to be analyzed and then synthesized in an overall judgment, in a transparent way. The different costs (which are not necessary monetary) of each solution have to be identified and synthesized in an overall cost. Every solution can be graphically shown in a two axe diagram (see Figure 2) where, as is natural, the most effective solution is also the most expensive. One or more acceptability thresholds can be introduced to facilitate a decision that is not easy to make.

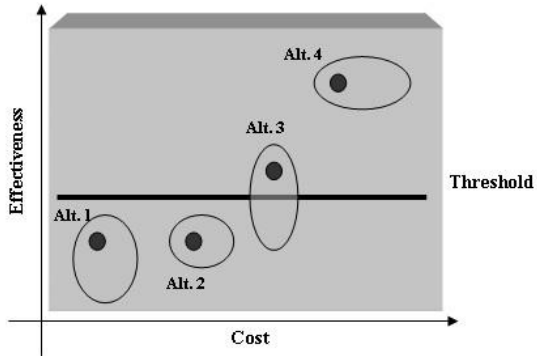


Figure 2: Cost/effectiveness analysis

The MBDD adopted a different approach when a client suggested the use of a multiple criteria method for the comparison of the alternatives. The first application was not totally satisfactory. Some other experiments allowed the MBDD to realize that the correct approach has to involve structuring the evaluation model in macro aspects that can be analysed separately by the organisation actors in charge of each specific aspect. At the same time, the MBDD elaborated a way of translating each personal judgment into an analytical function.

When we analysed the procedure they were using, we noticed that the results were very interesting, in relation to the first aim (improve communication with the client in order to understand his point of view and adequately model his requirements), but very poor as far as the second (transparently arrive at the decision and analytically document the decision motivations) was concerned. In fact, their analytical functions did not result to be consistent with either the original judgements or with the logical structure of the method. At the same time, their need to artificially construct evaluation functions, after the application of the method and in order to explicitly document the process, was analysed together with them, to show them that the wrong method had been adopted.

There are many multiple criteria methods that can be used to aid decision making (see for instance, Belton & Stewart, 2002). It is necessary to choose among the various methods in relation to the specific requests of the decisional problem. In this case, they essentially are: transparency of the process that elaborates a conceptual solution; an objective way of expressing each evaluation and, finally, a treatment of the uncertainty that affects data and judgments.

We proposed the analysis of two methods, Analytic Hierarchy Process (AHP) (Saaty, 1980; 1994) and ELECTRE III (Roy, 1978; 1990), which were tested in relation to the examined case studies. AHP is the method that is already used by the MBDD with its SW tool, Expert Choice, which supports model structuring by means of an easily visualization of the organizational aspects, scenarios, problem dimensions and model criteria. A sensitivity analysis facilitates the identification of model weakness elements and the consequent improvement or re-structuring of the problem and/or the multiple criteria model. An analytical evaluation of the solutions is not required. Comparative judgments are used both to assess the solutions and calculate weights for the compensatory synthesis procedure.

In the ELECTRE III method, unlike AHP, the alternatives have to be evaluated in relation to all the criteria in an explicit and (as much as possible) objective way. Thresholds are introduced when uncertainty is present in some evaluations, to limit the negative effect of the uncertainty on the results. Criteria can have different degrees of importance and, in this case, coefficients of relative importance of the criteria have to be introduced.

ELECTRE III starts by comparing each solution with each of the other solutions. A fuzzy outranking relation, based on the two principles of concordance and discordance, is modelled in phase I of the method through the computation of a concordance index, a discordance index and an outranking degree. The method uses the latter result in the second fuzzy relation exploitation phase, in order to construct two complete pre-orders through a descending and an ascending distillation procedure. Outranking relation modelling offers some interesting advantages, in comparison to other multiple criteria methods: each criterion can use a different ordinal or cardinal scale, since a unique specific scale (such as the cost-benefit analysis monetary scale or the 0-1 utility scale of the multi attribute utility theory) is not necessary and the outranking relation is not compensatory (or partially not compensatory).

A weak point of ELECTRE III is its software package, which does not pay any attention to dialogue with the decision maker, which is essential in model structuring and parameter definition and when the results require a collective analysis. A new product, which is more suitable and includes several multiple criteria methods, is currently being developed in the Decision Deck project[3]. This weak point is related to the original nature of the method which was invented to be used when a problem was well structured, i.e. when:

- a set of solutions is identified, or elaborated, and tested in terms of completeness, admissibility and comparability, and
- a family of evaluation functions (i.e. criteria) which has been created to represent all the different aspects of the problem at hand contains a sufficiently small number of criteria to be a basis for discussion (legibility condition) and to be considered by all the actors as a sound basis for the continuation of the decision aid study; its coherence (exhaustiveness, cohesiveness and redundancy) has to be verified by operational tests (Roy & Bouyssou, 1993; Roy, 1996).

For this reason, ELECTRE III is not normally used until the problem (and/or the model) is structured and only when these conditions are satisfied does it become a powerful method to transparently compare solutions, in relation to all the different criteria, and to rigorously synthesize evaluations that are associated to the consequences of each decision.

Therefore, our proposal was: the AHP would be used in the problem definition step, when pertinent information has to be identified together with the client, by means analysis and selection of conceptual solutions, while the ELECTRE III would be used at the end, in relation to the defined problem, when a decision has to activate the subsequent design process phases.

In the examined cases, the aircraft and the other systems were under production, or at least in the final phases of the production process, and the nature of the missions was clear, since the MBDD knows the decision context very well. Therefore, the principal aspects of the evaluation problem were easily identified (SoS performance in relation to the operational scenarios of the missions, technical effectiveness in relation to the operational management process and life cycle costs) and their disaggregation into organizational and functional-economic components was visualized through the SW Expert Choice and its multilevel decision tree.

In order to support interaction with the client, three models were elaborated during some simulation sessions in the MBDD: an AHP-Expert Choice model that is sufficiently general to be used in different decision situations, with a decision tree that is articulated in five levels and twenty-six elementary components for the comparison of the solutions, and two models for ELECTRE III, with twelve criteria for the first case study and fifteen criteria for the second one. Different decisional scenarios were hypothesized, in order to analyze which impact could have on the result the importance that the criteria assumed.

The AHP-Expert Choice model and the results of some applications were then analyzed to understand their potentiality to facilitate communication between the MBDD and the different clients. The ELECTRE III results were examined in terms of robustness and reliability, and the models in terms of formal validity and consistency with the internal procedures of the company.

Conclusions

A client's involvement in the initial phase of an engineering design process is always important and has to be carefully managed. The temporal horizon to produce an innovation in the aeronautic sector always involves a difficult definition of the clientǿs needs and some risks in translating the needs into formal requirements. The analysis and comparison of some draft solutions is an effective approach to understand the clientǿs point of view and the general structure of his/her preference system. However, this approach requires time to elaborate understandable technical solutions, analyse them with the client and elaborate new solutions for a new collective analysis, in a learning cycle.

Complexity and uncertainty elements can have a negative impact on the problem definition in some decision situations, above all when different, and sometimes conflicting, points of view require the involvement of some specific competences, from the clientǿs organization, as a not easy, but almost obligatory course of action.

A structured procedure can support the acquisition of the different points of view and their translation into mathematical models and then into product requirements, and can prevent, or at least control, ambiguous specifications by an activity that has the aim of verifying the overall consistency of the models.

The opportunity to produce conceptual solutions in a short time (a solution requires only few seconds of

[3] www.decision-deck.org

calculation time), with the guarantee of technical acceptability and specific performance levels in relation to an objective, makes communication possible and effective in the engineering design process.

Mathematical models that use an intelligible language introduce a positive psychological effect, in terms of clear thinking structure and perception of the logical progress. At the same time they facilitate the traceability of the process steps and results.

The integrated use of linear programming and multiple criteria methods can make the active collaboration phase with the client more rigorous (no acceptable solutions are lost and the evaluations can be documented and used consistently) and efficient, because all the structured and partially structured indications can be introduced into the models and transformed, by means of the methods, into information for the decision process.

The MBDD is planning to test the new approach with its clients and our group will be involved in analysing the criticalities and opportunities.

References

Akao, Y. (1997). QFD: Past, present, and future. *Proceedings of the Third Annual International QFD Symposium*. (pp. 19-29). Linköping University .

Alenia Spazio (2004). Multidisciplinary Optimisation (MDO) ï Problem Architecture Definition (*Technical note SDï TNï AIï 085).* Turin: Alenia Spazio.

Altshuller, G. (1988). *Creativity as an Exact Science*. New York: Gordon and Breach.

Belton, V., & Stewart, T.J. (2002). *Multiple criteria decision analysis. An integrated approach.* Dordrecht, Kluwer Academic Publisher.

Clark, K.B., & Fujimoto, T. (1991). *Product development performance: strategy, organization, and management in the world auto industry*. Boston: Harvard Business School Press.

Dantzig, G.B. (1963). *Linear Programming and Extensions*. Princeton: Princeton University Press.

Dean, E.B., & Unal, R.(1992). Elements of designing for cost. *AIAA-92ï 057*. Reston, Virginia: AIAA Press.

Eggert, R.J. (2005). *Engineering Design*. Upper Saddle River, N.J. : Pearson - Prentice Hall.

Ehrgott, M., & Wiecek, M.M. (2005). Multiobjective Programming. In J. Figueira, S. Greco & M. Ehrgott, (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*. London: Springer Verlag

Hatchuel, A., & Weil, B. (2009). C-K design theory: An advanced formulation. *Research in Engineering Design*, 19, 4, 181ï 192.

Norese, M.F., Montagna, F., & Riva, S. (2008). A multicriteria approach to support the design of complex systems, *Foundations of Computing and Decision Sciences*, 33, 1, 53-70.

Norese, M.F, Liguigli E, Montagna F., Protti, M., Prete F., Todino F., & Mascolo D. (2008a). Methodologies of decision support for the design of complex aeronautic systems (in Italian), In G. Felici & A. Sciomachen (Eds.), *Scienza delle decisioni in Italia: applicazioni della ricerca operativa a problemi aziendali* (pp. 223-238). Genova: ECIG.

Norese, M.F, & Liguigli, E. (2009). The conceptual design of a new land monitoring system, *70th meeting of the EURO working group on MCDA*, Moncton, Canada.

Office of Aerospace Studies (2002). *AoA Handbook. A guide for performing and Analysis of Alternatives (AoA)* . www.oas.kirtland.af.mil

Pugh, S. (1990). *Total Design: Integrated Methods for Successful Product Engineering*. Wokingham, England: Addison-Wesley.

Reinertsen, D. G. (1999). Taking the Fuzziness Out of the Fuzzy Front End. *Research Technology Management,* 42. 6, 25-31.

Roy, B. (1978). Electre III: Un algorithme de classements fondé sur une représentation floue en présence de critères multiples. *Cahier du CERO*, 20, 3, 24-27.

Roy, B. (1990). The outranking approach and the foundations of ELECTRE methods. In C.A. Bana e Costa (Ed.), *Readings in Multiple Criteria Decision Aid.* Heidelberg: Springer-Verlag.

Roy, B. (1996). *Multicriteria methodology for Decision Aiding*. Dordrecht: Kluwer.

Roy, B. & Bouyssou, D., (1993) *Aide multicritère à la décision: méthodes et cas.* Paris: Economica, Collection Gestion.

Saaty, T.L.(1980). *The Analytic Hierarchy Process*. New York: McGraw-Hill.

Saaty, T.L. (1994). *Fundamentals of decision making with the Analytic Hierarchy Procesòs*. Pittsburgh: RWS Publications.

Smith, P.G., & Reinertsen, D.G. (1992). Shortening the Product Development Cycle. *Research Technology Management*, 35, 3, 44-49.

Smith, P. G. & Reinertsen, D. G. (1998). *Developing Products in Half the Time: New Rules, New Tools* (2nd Ed.). New York: Wiley & Sons.

Tadei, R., & Della Croce, F. (2001). *Ricerca Operativa e ottimizzazione*. Bologna: Esculapio.

Improvement of database administration by procedure contextualization

H. TAHIR and P. BREZILLON
Laboratory of Computer Science-Paris 6 (LIP6)
University Pierre and Marie Curie (UPMC), Paris, France
{Hassane.Tahir, Patrick.Brezillon}@lip6.fr

Abstract

Database Administrators (DBAs) relieve on a large set of procedures for incident solving in database. However, in the one hand, they have to work under temporal and financial pressures, and, in the other hand, DBAs are continually readjusting these procedures to manage a multitude of specific situations that differ from the generic situation by some few contextual elements. The exceptions are rather the norms. Thus DBAs developed practices that deal with these contextual elements in order to solve the problem at hand.

Capturing and managing practices is far more difficult than procedures. If a procedure-based support system is easy to implement (procedures are well established), a practice-based support system is difficult to design because there are almost as many practices as contextual variants. However, if the context has an infinite dimension, the number of practices is finite. The key elements are the incremental acquisition of knowledge and the learning of new practices, thanks to a software called Contextual Graphs.

The goal of our work is to support DBAs by collecting all the practices developed by the DBAs and proposing them to DBAs to benefit of the experience of the other DBAs and to provide a support system acting as a real context-based intelligent assistant system.

Introduction

The Internet era has enabled organizations to accelerate their changes in order to be ready to the new challenges particularly those related to the management and the development of their information systems. Most of organizations use a database management system (DBMS) to manage their data, and the presence of a database administrator (DBA) today is mandatory. A DBA is the information engineer responsible for ensuring the ongoing operational functionality and efficiency of an organization's databases and the applications that access those databases (Mullins, 2002). This requires him to perform a variety of tasks in different areas including database design, performance monitoring and tuning, database availability, security, backup and recovery, data integrity, release migration. All tasks involve the company's databases.

Database administration is a continuous activity, and DBA spend long days with lots of overtime, especially when there is complex performance problems to resolve. In addition DBAs frequently has to work on weekends and holidays to perform database maintenance and reorganizations during off peak hours. He must be constantly available to analyze and correct failures. Most DBAs carry cell phones and other contact tools so that they can be joined at any time to solve problems encountered to get the applications back up and running, and to avoid database downtime that can completely shut down business processes. Of course, the DBA master all database management fundamentals, but he must be expert in the specific DBMS used in the organization: Oracle, SQL Server, Sybase, DB2, etc… In addition, he needs to acquire and develop exceptional communication skills because he frequently interfaces with many different types of users: technicians, programmers, end users, customers, and managers.

A good DBA participate in all phases of the application development life cycle. For example, in the initiation and requirements gathering phase, he identifies data components of the project to check if the required data already exists elsewhere in the organization or if the data is new. He must also help to determine the final status of the data used by the application which is not longer useful and he is responsible for managing the overall database environment that includes installing the DBMS and setting up the IT infrastructure to allow applications to access databases. Two modes of actions can be distinguished: reactive and proactive. The reactive DBA attempts to resolve problems only after problems occur and he is focused on resolving the main problem confronting him. In the other hand, the proactive DBA implements and develop practices and procedures to avoid problems before they occur.

As far as procedures are concerned, those issued by the database administration manuals cannot all explain because they cannot take into account all possible contexts. Therefore, they are often applied with caution to solve a given problem. In some cases, the DBA prefers to plan his actions in real time (and thus transform the procedure into a practice) to take into account the specific context of the current situation and the problem that he faces.

For example, for a reported anomaly with an error message in log file (or trace file), a procedure may recommend changes in the parameters list, then stop and subsequent restart of the database to make effective the new setting. This recommendation may not always be applied by the DBA because another method may be preferable to apply, method that will not disrupt end users and to affect the data availability during the resolution of the detected failure. Thus, the DBA undertakes a contextualization of the procedure that will become a practice. However, if it is practically automatic and unconscious for DBAs, capturing and managing practices is far more difficult than procedures.

The above motivation shows that an intelligent assistant system for DBAs must, on the one hand, use explicitly contextual information, and, on the other hand, works on an experience base that capitalizes past experience. This supposes that the experience base is developed in a uniform representation of knowledge, reasoning and context. Our study is in the realm of research on context modeling and management particularly that initiated in Brézillon and Pomerol (1999). This paper describes the improvements obtained by contextualizing users' procedures in applications like database administration.

The paper starts by presenting database administration problems followed by a discussion of the context approach and the related work. Then it presents the different kinds of Knowledge in Database Administration. After this, it discusses DBA procedures and practices and how they are modeled using a contextual graph based on example of a simple procedure for applying a database patch. Finally, the paper lists the remaining future work and concludes what has been reached.

Database Administration problems

DBA activities include many tasks. Some of them must be performed on a regular basis, others in response to emergencies or specific user needs. Every day, the DBA follows the main established administration procedures within the company like in the following list:

- Verify that the database is up and accessible to the end users ;
- Verify that backup was successful ;
- Check performance statistics and if there are enough resources for acceptable performance ;
- Check the status of previous night's processes ;
- Look for any new alert log entries to check whether or not new database errors have appeared since the last time.

In addition, he must be constantly available and ready to deal with problems and prevent them before they occur.

One or many technical problems can impact the performance of the entire information system of the company. This includes problems due to the database, the server, the network and/or the application. Many of the database problems fall into the following categories:

1) Unable to connect to the database: Users cannot connect to the database because of locked account or bad configuration file, and sometimes because the database is down. The system administrator and DBA must interact and work together to set shared server and database configuration. The DBA often writes scripts to allow one or more critical databases to restart automatically if for any reason the server reboots. These scripts can include commands to start databases and connection processes (i.e. oracle listener process). If the DBA forgets to include a command to start one connection process, users will not be able to connect to the database even if the database is started.

2) Slow time response and bad performance: Users cannot get the results of their queries in an acceptable time frame.

3) Privileges to access the database: This category of problems can occur in two cases. The first one is when DBA do not grant sufficient rights to the users. The second one is when DBA grants excessive privileges to some users or applications causing serious security vulnerability.

4) Change in database structure: The application can be impacted if any an unanticipated change occurred at the schema level of a database such as accidentally invalidating and dropping objects (table, column, etc …).

5) Bad database deployment: The DBAs may accidentally propagate the changes made to the database in the testing environment to the production system. One of the other reasons of this category of problems is due to the bugs in the DBA's deployment scripts required to move from one environment to another.

6) Database maintenance: Some of problems occur when the required maintenance tasks are not performed by the database administrator. An example of this is when database parameters like memory, size and path of data files need to be modified. Other DBA maintenance problems are due to inattentive space management or erroneous backup scripts resulting in an incorrect data recovery or bad patch updates as it will be discussed later on in this paper.

7) Other problems: This category is about the least frequent problems. Examples of problems include DBMS bugs, data loss and hardware failure, etc ….

Details, descriptions and real examples about database administration problems can be found in (Mullins, 2002), (Oliveira & Nagaraja et al., 2006) and (Wessler, 2002).

Sometimes when of the above problems suddenly appear, companies may lose large amounts of money for each hour of downtime. In such situation, A DBA life can become stressful because of the excessive pressure to solve problems quickly. In this framework, our work will contribute to find ways to minimize those DBA stressful situations. One way to do this is by designing and implementing a context-based intelligent assistant system (CBIAS) to help the DBA by providing him with a comprehensive list of information to resolve critical problems he faced. This system will be developed using a methodology that should take into consideration contextual information about DBA procedures and practices as well as the different situations where he performs his tasks to ensure the availability and the safety of the database.

Context approach

Brézillon and Pomerol (1999) show that context plays an important role in a number of domains since a long time. This is especially true for activities as predicting context changes, explaining unanticipated events and helping to handle them, and helping to focus attention. Context is used to describe knowledge shared on physical, social, historical and other circumstances where actions or events happen. All this knowledge is not part of the actions to execute or the events that occur, but will constrain the execution of an action and event interpretation.

Context has an infinite dimension and it is not clearly defined. To deal with a large number of contextual information, Brézillon and Pomerol (1999) distinguish between three types of context (Figure 1) for a given focus of attention, namely, external knowledge, contextual knowledge, and proceduralized context. The external knowledge is the knowledge that has nothing to do with the current focus. Conversely, the contextual knowledge is the knowledge that is more or less relevant for the current focus of attention. Always at a given focus, part of the contextual knowledge is proceduralized. The proceduralized context is a part of the contextual knowledge, which is invoked, organized, structured and situated according to this focus.

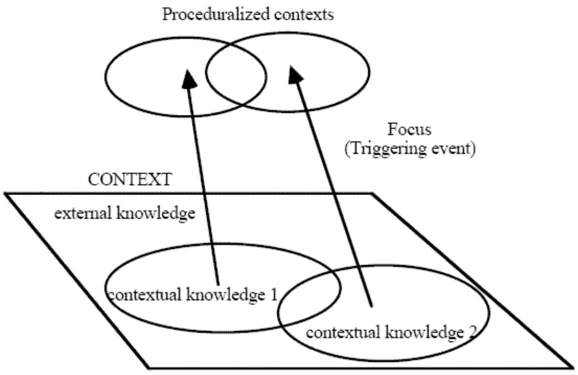


Figure 1: Different types of context.

Context can be modeled using an approach based on contextual graphs (Brézillon, Pasquier and Pomerol, 2002; Brézillon, 2002) where the contextual elements are acquired incrementally when needed. Later on in this paper we show to apply this approach to represent contextual information in database administration procedures.

Related works

Nowadays there are many automated database administration tools that helps the DBA in his decision-making. Traditional expert systems like Oracle Expert offer much functionality such as assisting the DBA by automating routine database maintenance and tuning tasks (ORACLE, 2001). The main principle of these systems is to suggest solutions to correct detected failures based on information already stored in a Knowledge Base. There exists also solutions with a learning mechanism to suggest the most adequate solution to each detected problem like in DBSitter which is based on the combination of two methods: (1) Case Based Reasoning (CBR) to adapt known solutions and enrich the Knowledge Base (2) Mutli-Agent where a set of intelligent agents is used to monitor the Database environment and actuate on it (Carneiro & Passos et al., 2004).

Other recent research works uses Policy-based computing to implement autonomic administration capabilities into a database for enforcing policies to control and decide which changes are allowed and which ones are not. (Qiao & Soetarman et al., 2007) present a framework to define, manage, and enforce policies that are used to isolate a problem into a more specific context, upon which either general or customized solutions are derived. Jabbour and Menasee (2008) introduce the notion of embedding policies into the database itself and enabling these policies to block every attempt to compromise the state of the database, or to alter its configuration in a way that contradicts what has been established and fed into the policy by the system owner. The Common observations concerning all these mentioned works and some similar works not discussed in this paper are the full considerations to technical aspects like using sensors to collect the contextual information about the administration environment like in DBSitter. Most of them seldom consider the contextual information about the preferences of DBA (experience matters, social and cultural, etc. ...) and the dynamics aspects of contexts. Nevertheless, some other research works have performed ethnographic field studies to develop guidelines for tools to better support how administrators actually work as in (Barrett & Kandogan et al., 2004) and (Haber & Bailey, 2007). The present work contributes to contextualize DBA procedures based on user-centered approach in database administration. To improve the usability of these procedures, the contextual graphs can be used. The following section will help understanding the different kinds of knowledge when dealing with database administration procedures.

A Knowledge Categorization in Database Administration

Effective database administration requires mix of three types of knowledge (Wessler, 2002): technical knowledge, business knowledge, and human knowledge as shown in Figure 2. These three main knowledge categories in database administration are themselves divided into sub-categories.

Technical knowledge

Technical knowledge corresponds to the technical skills the DBA needs to possess. It can be broken down into three main categories: database knowledge, application knowledge, and system knowledge.

Database knowledge

Database knowledge is about the RDBMS the DBA is responsible for. He should know everything about the database he monitors on a daily basis and provides all the maintenance for it.

Application Knowledge

Application Knowledge concerns program code affecting the database. The DBA needs to understand what it does and how it connects to the database. He also needs to support the packages and procedures stored inside the database. However tuning the code is a shared responsibility between the DBA and the application developers.

System Knowledge

System Knowledge corresponds to the different aspects of systems administration and networking that affect the database. It is interesting that the DBA should be able to serve as a backup System Administrator if needed.

Business knowledge

Business knowledge corresponds to the business skills the DBA needs to possess. It consists of Organization's Processes and industry trends. So the DBA have to consider and understand the business process as whole to add value to the process and to the organization. On the other hand, he should follow the industry the organization is in and to keep up with the IT industry. This will help him to be in a good position to come up with new ideas.

Human knowledge

Human knowledge corresponds to the people skills the DBA needs to master. It can be broken down into four main categories: Communication, Management, and Problem solving and Education.

Communication

Communication corresponds to the interaction between the DBA and other people, both inside and outside the organization. This will help in avoiding misunderstandings and accidents such as destroying each other's work or imposed system downtime.

Management

Knowledge of the entire system (not just the database) and how it supports the business makes the DBA a very knowledgeable person and often qualifies him to take on project management responsibilities. The people he manages may include not only DBAs but also System Administrators and programmers in support of a project.

Problem Solving

Much of what a DBA does is about gathering information and making judgments to solve both technical and non-technical problems. Often the real source of an error is hidden and will be found only by an experienced DBA who understands well how the database, operating system, and application interact with each other. Some other problems require skills in negotiation and compromise.

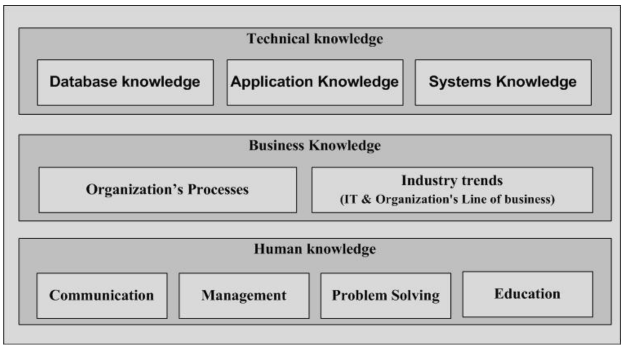


Figure 2: Kinds of Knowledge in Database Administration.

Education

Education concerns the DBA professional certification and training both inside and outside his organization.

DBA procedures and practices

Database administration procedures cannot always be successfully applied by database administrators (DBAs) to perform their tasks and solve the different problems they face. This is why assisted tools based on these procedures cannot always be efficiently used because they cannot take into account all possible contexts. In such situations, DBAs develop their own practices like to correct a failure by considering the context in which that failure occurs.

Degani and Wiener (1997) distinguish procedures, practices and techniques. Procedures are specified beforehand by developers to save time during critical situations. Practices encompass what the users do with procedures. Ideally, procedures and practices should be the same, but the users either conform to procedure or deviate from it, even if the procedure is mandatory. About the finite dimension of practices, the number of practices is of a reasonable size. Later in the paper, we will discuss a software tool called Contextual Graphs which is used to support incremental acquisition of new practices.

Techniques are defined as personal methods for carrying out specific tasks without violating procedural constraints.

The following is an example of a procedure of applying a patch to a database:

1. Shut down the database
2. Apply patch to the Database Home
3. Start the database
4. In case of patch application not succeeded perform steps 5, 6, 7:
5. Shut down the database
6. Roll back the patch
7. Start the database

Of course if the patch is ready, a DBA can perform the above steps during the allowed database change time windows when the users are not always connected to the database. Generally, night is the preferred period of time for generating backups, and also the ideal moment for applying database patch updates. However the DBA may not apply the procedure during the day when the database is running. He should take into account the different contexts related to the given situation. For instance, theses contexts concern the type of database environment (Development, Test, Pre-Production, Production) and whether end users and customers are connected or not. For each context, he may request information or authorization from other colleagues or his manager to perform the required patch actions. Hence, the step 1 to shut down the database cannot be performed if users are connected especially if any user is connected to the database on the production environment. In this case the database administrator may request an authorization from a manager before applying the patch. If the authorization to shut down the database is given, he can start to execute

the actions 1 to 3. So, the DBA has adapted the above procedure to the contextual knowledge he has acquired over time by developing practices to reach the goal fixed by this procedure.
As explained by Brézillon (2003), the construction of the proceduralized context from contextual knowledge is often a process of communication in a community of practice, even if members of this community come from different domains. Figure 3 represents how the different proceduralized contexts built from contextual knowledge during the interactions between the DBA and each category of users.

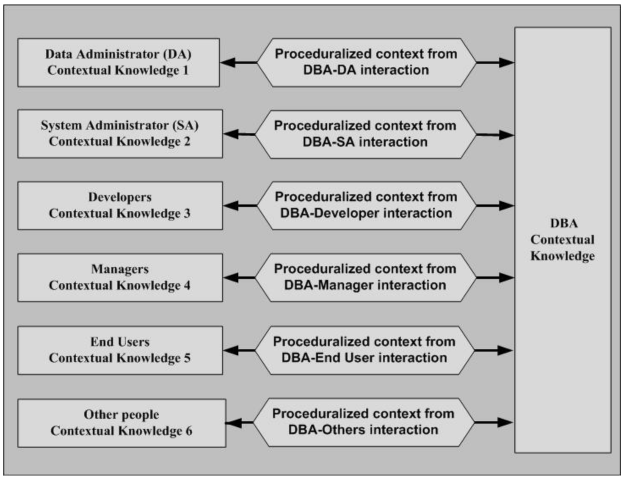


Figure 3: Context in DBA-User's interactions

The different interaction contexts contain elements of the contextual knowledge for the building of the proceduralized context in the focus of attention of the DBA and the corresponding user. These elements of knowledge in the interaction context are extracted from the contextual knowledge of each category of user. Then they are proceduralized by the two persons, and result in a shared chunk of knowledge. The resulting proceduralized context contains all the pieces of knowledge that have been discussed, accepted and assembled by the DBA and each category of user. For example, in the DBA-System Administrator (SA) Interaction, the SA should provide relevant information to the DBA when he requests information about the server such as the disk space and the needed permissions.
As shown in the Figure 3, in most IT organizations, a DBA will interact with five types of users: data administrators, system administrators, developers, managers, customers and end users, other people outside the organization.

DBA-DA Interaction

The role of data administrator (DA) is managing data from a theoretical standpoint to build conceptual data models. The DBA-DA Interaction is about how logical models are actually implemented as tables. The DA functions mainly in larger companies. However, when there is no DA role in an organization (as often in small companies), the DBA must assume the role of Data modeler.

DBA-SA Interaction

The main role of the system administrator (SA) is to keep all the system boxes running efficiently and in a secure manner. If no system administration group exists, the DBA assumes responsibility for DBMS-related system administration and programming. The DBA-SA Interaction is needed where the SA should provide relevant information to the DBA and resources to support the Database especially in the following cases:

- At the moment of the creation of a new database or server ;
- When a DBA requests information about the server's backup strategy, the disk layout, RAID level, machine memory;
- When a DBA needs to know if there are any major non-database applications planned ;
- When DBA needs more disk space ;

The interaction between the SA and the DBA is not always a very close and cordial relationship. There is often contention between the two administrators on some issues such as the moment to reboot the machine, use of system resources (disk space and memory), backup and recovery procedures, and user policy.

DBA-Developer Interaction

The interaction between the DBA and Developers is needed mainly in the following cases:

- When Developers writes and maintain programs, the DBA should implement any required database or environment changes;
- Supporting developers in tuning their database queries used by their programs ;
- Database errors are the result of bad program code, the DBA asks the developers if they have changed anything to help him in troubleshooting.

DBA-Manager Interaction

The DBA-Manager interaction concerns the following:

- The DBA must report to his manager who may be technical or may be from a non-technical background ;
- The DBA explains to the manager the different parts of a system (at a high level) and explaining how they are dependent on each other. He can also explain what it takes to provide high system uptime, both in terms of hardware and in terms of trained IT professionals.
- The DBA may have to provide an explanation to his manager if problems still occur even if they might not even be related to the database.

DBA-End User Interaction

The DBA-End User Interaction concerns the communication between the DBA and a population of users with a different background and needs inside the organization, outside the organization, or even outside the company. Often it is the end user or a customer who is the first to detect a failure in the system, even before the IT staff does. The administrator needs to listen to what is being said and provide support and answers that are understandable.

DBA-Other people Interaction

The DBA-Other people interaction concerns the different communications and contacts between the DBA and RDBMS vendors (Oracle, Microsoft, IBM … etc), and other outside companies. The main objectives of these interactions are the following:
1) To choose products and negotiate support contracts;
2) To ensure that any new purchased software, related to both database and non-database uses, does not adversely affect data protection or availability of the existing databases. It is up to the DBA to do the needed testing to guarantee that no unexpected problems or bugs occur because of a new software addition;
3) To Sign up for the highly recommended training classes, pursuing technical support issues.

The following section presents contextual-graphs and how they can be used in any situation in which database administrators developed practices from the existing procedures adopted by the organization.

Contextual graphs for database administration

Contextual graphs have been initially designed for an application for incident solving on a subway line (Brezillon & Pomerol, 2000). A contextual graph is an acyclic directed graph with a one input, one output, and a serial-parallel organization of nodes connected by oriented arcs. It is an acyclic graph because user's tasks are generally performed in ordered sequences. For example, the activity “Make sure users are disconnected from the database” is always considered at the beginning of an incident solving, never at the end of the incident solving. There are different types of nodes in a contextual graph: actions, contextual and recombination nodes, sub-graphs and parallel grouping. A sub-graph allows the modeling of DBA activities, and thus contextual graphs give a representation of the reasoning directly understandable by database administrators. A path is an ordered sequence of elements (contextual and recombination nodes and actions) of the contextual graph from the input source to the output. Each path represents, by its sequence of actions, a practice.

In a contextual graph (CxG), the dimension of context is finite, limited by the number of practices learnt by the system. The initial structure of a Contextual graph (its skeleton) is defined by the procedure that is established by the DBA company for the problem solving. Different practices can be developped for a given incident depending on the context in which the incident must be solved. CxGs support incremental acquisition of new practices.

Example of a DBA procedure for applying a database patch:

As an example, we use contextual graphs to model the context-based DBA procedure for applying a database patch. The previous procedure can be represented using the contextual graph in Figure 4.

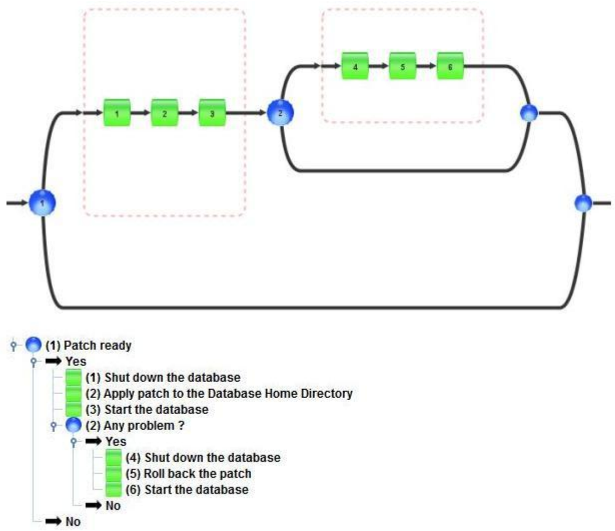


Figure 4: Contextual graph of a database Administration procedure

The graph contains two contextual elements 1 and 2 that can be noted as CE1 and CE2. CE1 is about if the patch is ready or not. In other words CE1can have two values (Yes, No). When the patch is ready (Value (CE1) = Yes), the sequence of actions 1, 2, 3 is executed by the DBA. These actions can be noted A1, A2 and A3. On the other hand CE2 allows checking if any problem occurs after the end of action A3. In the case of problem (or patch fails), the sequence of actions 4, 5 and 6 (noted A4, A5 and A6) is executed. Notice that A1 and A4 are instances of the action “Shut down the database” and that A3 and A6 are instances of the action “Start the database”. The two dotted frames in Figure 5 correspond to the two main DBA activities in the procedure for applying the database patch. Actions A1, A2 and A3 form the “Apply database patch” activity whereas actions A4, A5 and A6 form the “Undo patch” activity. The Figure 5 shows another representation of the DBA procedures using contextual elements and activities.

The procedure represented in Figure 5 can be applied only in the ideal conditions where the non availability of database for a period of time will not affect users and the business of the company which is not always the case. In practice, the DBA adapt administration procedure to the different contexts without losing the main purpose of the procedure. In our example, he should consider contexts such as whether or not users connected to database, and what should be done in the case that the patch is not available. So, in this case for example he should take into account other contexts like if he uses or not web support to download the patch and if the web site is down or not, etc.

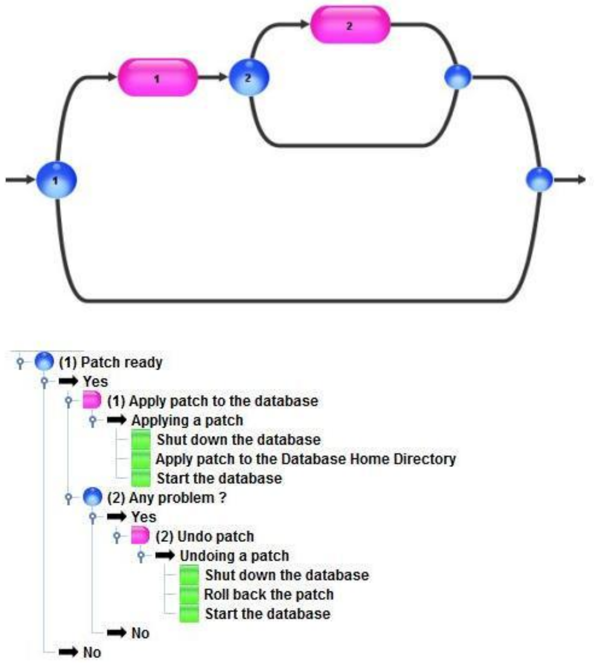


Figure 5: Using DBA activities in contextual graphs

The Figure 6(a) details the different practices developed by the DBA to consider the different contextual elements. The context of the contextual graph presented in Figure 6(a) is given by the elements {CE1, CE2, CE3, CE4, CE5, CE6, CE7, CE8, CE9}. Figure 6(b) and Figure 6(c) explain numbers on Figure 6(a). The contextual graph shows that DBA has developed many practices around the original procedure for applying the database which is not sufficient in the multi-user interactions. We showed that it is important to consider not only the technical contexts affecting the database like the type of environment and the availability of website to download the patch, but also the contexts shared in the different interactions between the DBA and other group of users (Developers, managers, system administrators, etc).

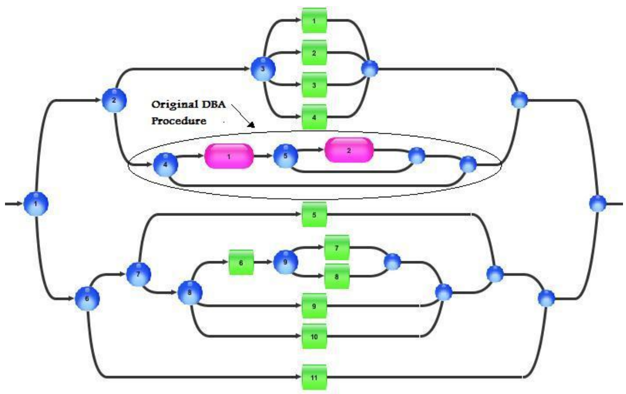


Figure 6(a): Adding DBA practices in contextual graphs

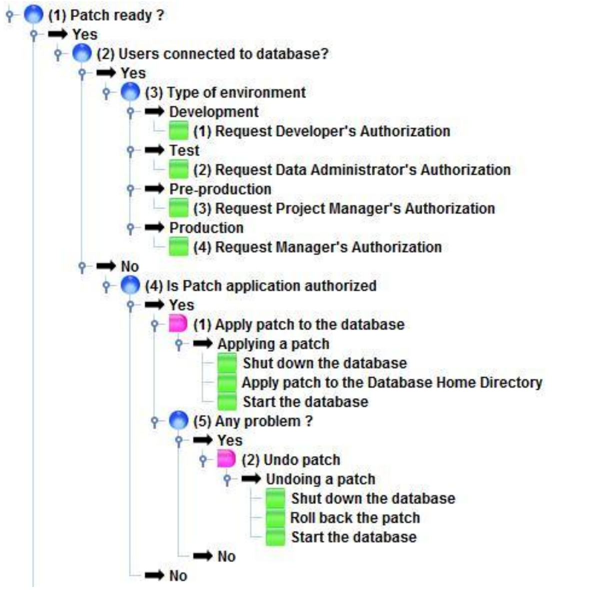


Figure 6(b): Description of practices

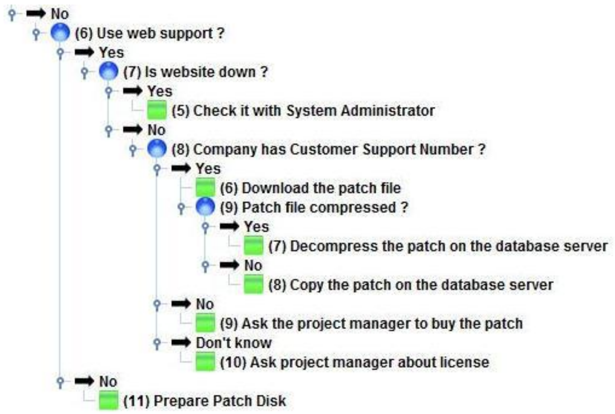


Figure 6(c): Description of practices (continued)

The context of an action (e.g. A7) is composed of two parts: the contextual elements used on the path from the input and the other elements. On the path, some of the contextual elements has a value that intervenes in the practice (CE1, CE6, CE7, CE8 and CE9) and other not (CE2, CE3, CE4 and CE5). The first ones intervene in an ordered way and are called the proceduralized context. The second one is the set of elements called the contextual knowledge.

Thus, the context of the action A6 is defined by:

- Its proceduralized context: {CE1, CE6, CE7, CE8, CE9}, supposing that the actions A5, A6 are realized.
- The contextual knowledge: {CE2, CE3, CE4, CE5}

The main interest of contextual graphs relies on the possibility to introduce easily new practices in contextual graphs. A new practice generally corresponds at a known practice with few changes introduced by contextual nodes. Thus, a contextual graph based system either knows a practice used by a DBA or acquires it when needed.

Future work

The paper has shown that it is possible to use contextual graphs to model and represent database administration procedure. This has been illustrated using an example of a patch to be applied by a DBA to update the database in order to resolve incidents encountered by users or detected by the monitoring tools in the log files (or trace). In the case studied, we have pointed not only the technical contexts related to the patch application like the availability of a web site to download the patch but also the contexts about the different interactions between the DBA and other actors. The database patch procedure that has been used is based on a conventional or traditional approach of applying a patch. In our future work, we will continue our research by considering the other existing approaches for that procedure and the following aspects:

1) Complete the contextual graph by considering other contextual elements about:

- DBA-user's interactions;
- Contexts related to the technical details of the database patch (i.e. version number, release number, upgrade number, files copied, and bug) as well as database objects affected by the patch, etc.... Contextual graphs will also be expanded to other administration procedures such as database recovery after a failure causing loss of data. This will help to find out the new practices that will be developed by the DBA;

2) Now by using contextual graphs, we are able to represent a DBA task for applying a patch to the database. This can be extended to represent all the DBA tasks in order to build a real experience base.

3) Explore the possible interactions between Contextual graphs representing different tasks and their consequences.

4) Design and implement a context-based intelligent assistant system (CBIAS) that uses an experience base to help the DBA. The experience base should be developed in a uniform representation of knowledge, reasoning and context.

5) Evaluate the CBIAS by DBA for a set of procedures. The feedback of the database administrator is important for us to evaluate the practical efficiency of the system and to list the set of cases where difficulties may be encountered.

6) To generalize and extend the context-based intelligent assistant system to other domains of applications.

Conclusion

The paper has shown through an example that it is possible to improve the database administration by the contextualization of procedures used by the DBAs. The example concerns a procedure for applying a patch to update a database. It has illustrated that it is so easy to use contextual graphs to model the procedures and practices that have been developed by the DBA to consider the various contexts about situations he faced.

Contextual graphs have been used to represent both the initial administration procedure and the new DBA practices. These practices concern both technical and human contexts, with an emphasis on the various multi-actor interactions (DBA-Developer, DBA-Manager, DBA-End User, etc. ...).

Finally it is important to note that our research work is in the framework of the design of context-based intelligent assistant system not only for database administrators but also to be used in many other areas such as software design support, medicine, energy,…etc.

References

Barrett, R., & Kandogan E., et al. (2004). *Field Studies of Computer System Administrators: Analysis of System Management Tools and Practices.* Proc. CSCW 2004, November 6–10, Chicago, Illinois, USA. ACM

Brézillon, P. (1996). *Context in Human-Machine problem Solving: A Survey*. K. E. Review: 42.

Brézillon, P. (2006). *Some Characteristics of Context.* M. Ali and R. Dapoigny (Eds.): IEA/AIE 2006, LNAI 4031, Springer-Verlag Berlin Heidelberg, pp. 146 – 154.

Brézillon, P., & Pomerol J.-C. (1999). *Contextual knowledge and proceduralized context.* Proceedings of the AAAI-99 Workshop on Modeling Context in AI Applications, Orlando, Florida, USA, July.

Brézillon, P., & Pasquier, L., & Pomerol J.-Ch. (2002). *Reasoning with contextual graphs.* European Journal of Operational Research, 136(2): 290-298.

Brézillon, P. (2003). *Context-based modeling of operators' practices by contextual graphs.* Proceedings of HCP-2003, the 14th Mini-Euro Conference on Human Centered Processes, R. Bisdorf (Ed.), Fonds National de la Recherche, Luxembourg.

Carneiro, A., & Passos R., et al. (2004). *DBSitter: An Intelligent Tool for Database Administration,* Lecture Notes in Computer Science, Volume 3180/2004, 171-180, Berlin-Heidelberg, Springer-Verlag.

Degani, A., & Wiener, E. L. (1997). *Procedures in complex systems: The airline cockpit*. IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans, 27(3), pp. 302-312.

Haber, E. M., & Bailey, J. (2007). *Design Guidelines for System Administration Tools Developed through Ethnographic Field Studies*, Vol6, pp. 388-395, ACM.

Jabbour, G. G., & Menasee, D. A. (2008). *Policy-Based Enforcement of Database Security Configuration through Autonomic Capabilities.* Fourth International Conference on Autonomic and Autonomous Systems ICAS 2008. Gosier ICAS.

Mullins, C. S. (2002). *Database Administration: The Complete Guide to Practices and Procedures*, Addison Wesley.

Oliveira, F., & Nagaraja K., et al. (2006). *Understanding and Validating Database System Administration* http://www.cs.rutgers.edu/~ricardob/papers/usenix06.pdf

ORACLE (2001). *Introduction to Oracle Expert.* http://download.oracle.com/docs/html/A86647_01/intro.htm.

Qiao, L., & Soetarman B., & et al. (2007). *A framework. for enforcing application policies in database systems.* ACM.

Wessler, M. (2002). *Oracle DBA on Unix and Linux*, *for* SAMS.

Human centered design for nuclear power plant control room modernization

Paulo V. R. Carvalho (paulov@ien.gov.br), Jose O. Gomes (joseorlandogomes@terra.com.br), Marcos R. S. Borges (mborges@nce.ufrj.br)
Graduate Program in Informatics
Federal University of Rio de Janeiro
Cidade Universitária, Rio de Janeiro, Brazil

Abstract

The use of nuclear power plants to produce electric energy is a safety-critical process where ultimate operational decisions still relies with the control room operators. Thus it is important to provide the best possible decision support through effective supervisory control interfaces. A human centered design approach, based on cognitive task analysis methods, was used to observe the operators training on the nuclear power plant simulator of the Human System Interface Laboratory (LABIHS). We noted deficiencies in graphic interface design, alarm system and in the integration between the computerized interfaces and the hardcopy (paper) procedures. A new prototype of the interface including graphics, alarms and digital procedures was designed as an alternative to the current hardcopy procedure manuals. The design improves upon the graphical layout of system information and provides better integration of procedures, automation, and alarm systems. The new design was validated by expert opinion and a performance comparison with the existing design.

Introduction

In control theory, systems can be modeled as interrelated components that maintain the system's stability by feedback loops of information and control. The plant's overall performance has to be controlled in order to produce with safety, quality, and low cost. In such an arrangement both controllers (human and automatic) play fundamental roles such as to establish system goals, to know the system status, and its behavior in the near future. This is done through continuous observation/feedback/communication loops where the agents construct their system model of behavior in order to compare with system status, to be able to act on the system to produce the desired outcomes. In this control mode, the human operator has a supervisory role related to the automatic controller. The operator has access to system state information, using the control room indicators, VDUs, strip charts, alarms and the automation controller status, and may have direct ways to manipulate the controlled process, and automatic systems interact with some sections of the plant rapidly and reliably.

However, automatic systems cannot cover the whole operational range of the plant including design basis events. For example, if the configuration of the plant changes for maintenance or accidents, the applicability of the controller might be limited. In that case, humans set up an operational strategy, supervise the automatic systems, and control the plant manually as necessary. Therefore there is a need of a human centered approach in the modernization of current analog interfaces of nuclear power plant control rooms.

The goal of this article is to describe a human centered approach to evaluate and design control room interfaces of safe-critical systems. The research aims the modernization of nuclear power plant control rooms in the design of the graphic interfaces, the layout and informativeness of the alarm system, and the integration of electronic procedures into the control/display environment (Carvalho et al., 2008).

Many nuclear power plants (NPP) around the world are modernizing with new systems and equipment such as upgrading the instrumentation and control (I&C) system from analog to digital technology. Generally, as part of these upgrades, control rooms are being modernized and computer-based interfaces are being introduced, such as software-based process controls, touch-screen interfaces, computerized procedures, and large-screen, overview displays.

This research is connected to the life extension process of a Brazilian nuclear power plant. The plant is a Westinghouse, 600 MWE pressurized water reactor designed in the 60s that suffers a continuous modernization and life extension processes. This overall research aim is to investigate how advanced (digital) interfaces can be used in the modernization of the analog instrumentation and human/system interaction (HSI).

This article is divided as follows. In the next section we review the modernization approach based on control room modernization. The third section is dedicated to methods and materials used in the research of human factors in NPP operations. Section 4 presents the results and a set of recommendations for a new interface aimed to modernization of control rooms. Section 5 presents the evaluation of the new interface design, focusing on human factors, and section 6 presents a discussion and some lessons learned. Finally, Section 7 concludes the paper.

Human centered interface design in NPP control room modernization

The nuclear power plant control room operators observe and manipulate an extremely complex system. The task requires walking along a large control panel, taking readings from gauges and adjusting knobs and levers. Many of today's control rooms have replaced or augmented older, more cumbersome control panels with visual display units (VDUs) with graphic interfaces. VDUs can simplify the human machine interface, but they also introduce new design challenges. Digitalization of previous analog man-machine interfaces imposes new coordination demands on the operational teams (Vicente et al., 1997). These issues lead to new situations of human-human and human–system interaction. In order to run such system effectively, efficiently, and safely, much research has been developed taking into account human performance, new technological possibilities, and types/levels of automation in a system, design of human–machine interfaces etc. (e.g. Sheridan, 2002; Nachreiner et al., 2006).

After the Three Mile Island (TMI) accident in 1979 NPP regulators around the world recommend the use of human centered approach to human systems interface design to ensure that the man-machine interfaces, control room layout, procedures, training and other human related issues meet the task performance requirements, and are designed to be consistent with human cognitive and physiological characteristics (Rouse, 1984). The human aspects related to the control room design such as operating experience review, function analysis and allocation, task and activity analysis, staffing qualifications, training, procedures should be developed, designed, and evaluated on the basis of a systems analysis that uses a "top-down" approach, starting at the "top" of the hierarchy with the plant's high-level mission and goals (O'hara & Brown, 2004).

However, most of the modernization processes has been driven to a large extent by the technology. The modernization of the turbine control in the NPP under study can be viewed an example of technological driven approach. A new computerized turbine control system was purchased to replace the old analog controllers. Although the new system perform its functions better than the old one, it is also true that the installation of computer screen and keyboard along with the analog instruments in the hardwired panel, as shown in figure 1, lead to human-system interaction problems.

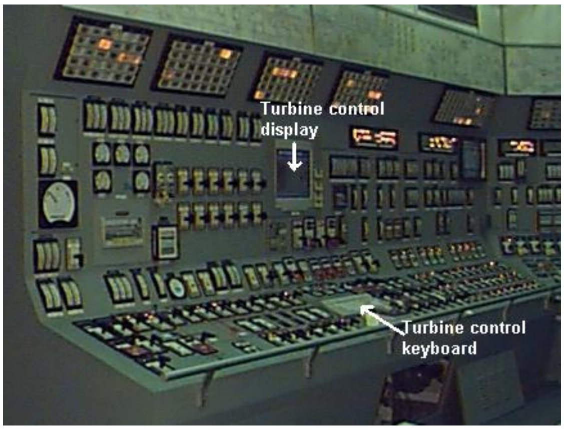


Figure 1: Turbine display and keyboard together with analog instrumentation.

The human-centered approach exploits the technical innovations to achieve an optimum human – artifact interactions, aiming at improving the appropriateness of the technological solutions (Hancock & Chignell, 1995). The human centered approach to the design of human-system interfaces considers the impacts of the introduction of new technology on the humans in the system and on the overall behavior of the system, from the beginning and continuously throughout the design process (Brunélis & Blaye, 2008). The approach requires specific activities that should take place during the system design. These activities are: 1) to understand and specify the context of use; 2) to specify the user and organizational requirements; 3) to produce design solutions and to evaluate designs against requirements. The human-centered design process should start at the earliest stage of the project (e.g. when the initial concept for the product or system is being formulated), and should be repeated iteratively until the system meets the requirements. It is not sufficient to verify the quality with which the design process is carried out (concerned with whether certain design phases were carried and certain documents produced to meet the design requirements). Considering that the in human-centered design approach, technology should be comprehended from the point of view of providing tools for human activity (Flach et al., 1995), it requires a dynamic performance evaluation, to assess the appropriateness of this technology in the aimed use.

Materials and methods

The construction of the NPP under study started in 1972, the first criticality of the nuclear reactor occurred in 1982 and the plant commercial operation started in 1985. Since then, it generated 40 million MWh of electric energy. Into the modernization and life extension plant program an upgrading of I&C and Human System Interface (HSI) systems is planned.

In order to support the application of the human centered design approach in the modernization of the Brazilian NPPs, the Brazilian Nuclear Energy Commission (CNEN), developed an experimental facility for human system interface design and human factors research and development, the Human System Interface Laboratory (LABIHS). LABIHS facility is ready to conduct NPP operators' performance evaluations, and research on human-system interaction in complex domains. The LABIHS consists of an advanced control room, an experimenter's gallery room and other auxiliary rooms. The advanced control room has nuclear reactor simulator software, graphical user interface design software, a hardware/software platform to run and provide the adequate communication between the software, and the operator interface - VDUs and controls needed to operate the simulated process.

To simulate the plant under study, a Westinghouse PWR type digital compact simulator is used. In this simulator, modeling scope and fidelity are equivalent to a full scope simulator, but the full control room is not replicated. An Integrated Hardware/Software Platform runs the simulator program and transfers data throughout the computerized environment. The basic operator workplace is formed by 4 VDUs, each one with mouse and keyboard. An overview display, based on direct beam projector, is also provided in the control room. A graphical user interface design tool (GUI) for HSI design is also available for development and testing of different types of interfaces. The Instructor Station complements LABIHS architecture. The LABIHS control room is shown in figure 2.

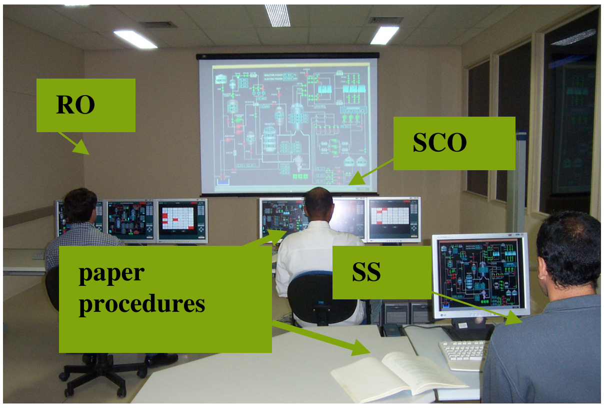


Figure 2: LABIHS control room. RO means Reactor operator, SCO Secondary system operator, SS Shift supervisor

Research method

In this research, we use LABIHS to investigate the nature of operator–system interaction in a digital interface during abnormal events to contribute to operational safety and efficiency through enhanced interface design. We use the interface evaluation procedure proposed by Hollnagel (1985) because it is consistent with most of human-machine interface evaluation requirements in the Human Factor Engineering (HFE) guidelines and programs that are currently used in nuclear industry, such as NUREG -0700 rev1 (O'Hara et al, 1996) and NUREG-0711 (O'Hara et al., 1994). The evaluation procedure has three phases. The first phase is the conceptual evaluation of the interface. It can be carried out by experts using tools like task analysis; operational experience review in similar systems; safety analysis reports; functional specification; drawing showing displays, panels, workstation, graphical interfaces and diagrams showing flows of information. In the second phase an heuristic evaluation is made based on some well known interface evaluation criteria (eg. Nielsen, 1993). The system is represented by samples taken from preliminary performance recordings, using results of runs with the real system or prototype. It is a static simulation. It concentrates on the way in which the information is presented to the operator and involves some form of basic system operator interaction. In the third phase, the entire process is simulated, and the operators' performance is evaluated. In this phase operators have a degree of psychological involvement and we can see how they react to the simulated process in a realistic manner. It requires a simulated work setting, a detailed experimental planning, including training, data acquisition, analysis systems such as computer logs (process state, process events), operator log (human machine interface events, keyboard, mouse) and audio, video recorder (verbal protocols, communication).

A final evaluation occurs during the plant commissioning tests in the plant site. At that moment any changes in control room interfaces will much more difficult and costly than it would be in the early phases (Santos et al., 2005).

Participants

One operator crew participate in this research under different operating conditions: start up, planned shutdown and in postulated accidents. The LABIHS control room operating crew is composed by 3 operators: the Shift Supervisor, Reactor Operator (RO) and the Secondary Circuit Operator (SCO). The Shift Supervisor have a deep background in nuclear engineering, participated in the LABIHS's HSI design, and have a huge experience in the simulator operation. The RO and SCO are instrumentation technicians who have been trained in LABIHS operation for 2 years before this study. They have no previous experience in the reference plant operation.

The operation of nuclear power plants

The operation of a nuclear power plant falls under four basic phases: startup, normal operation, planned shutdown, and emergency operation that begins after reactor automatic shutdown, when incidents/accidents occur. Although important events occur in all modes of operations, we focused the observation on periods of higher activity, such as startup and emergency operation.

Under normal conditions, NPP operations are well coordinated and based on procedural instructions. In this ''nominal'' operating mode, the SS reads the procedural instructions aloud to the RO and SCO who then execute the instructions (Carvalho et al, 2006).

Performance evaluation

During 30 hours of direct observations, we observed how the operators interacted with the simulated PWR in various modes of operation. The LABIHS is equipped with a ceiling-mounted camera which captures the majority of the room, including the two operators' stations and the main presentations screen (fig. 2). We placed a tripod-mounted Mini-DV camcorder to record whichever operator would be likely to have the most active role. A hand-held digital camera was used to film particular details of interest that were not sufficiently captured by the other two cameras.

The research team, with 3 analysts, was divided to pair up with the employees of the simulator. One analyst accompanied the primary operator; the second accompanied the secondary operator; and the third accompanied the simulator supervisor. The operation of a nuclear power plant fall under 4 basic phases: startup, normal operation, shutdown, and incidents/trips (unplanned automatic shutdown)/accidents. Although important events occur in all modes of operations, we focus on periods of higher activity - startup and incident/accident.

During the startup phase observations, we encouraged the operators to verbalize their goals, actions, and concerns to improve our understanding of the technical system. However, during the simulated accidents, we tried not to interfere with the operators so as to elicit true response behavior. During the simulated accidents, the supervisor and two senior LABIHS researchers were also present. This placed noticeably increased pressure on the operators, and also led them to justify their actions verbally after the scenario was completed.

We paid particular attention to the tasks dictated by the procedure manual and to the operators' actual activity. We search for particular deficiencies in the support of operator response to abnormal system states, and then we redesigned the operator interface to improve upon the graphical layout of the information, the navigation across screens, the alarm presentation, acknowledgement and response, and to integrate these with computer-based procedures that dynamically correspond with real-time system information. Comprehensive debriefing interviews with the operators and supervisor and was carried out to validate the conclusions taken.

Results and recommendations for modernization

Graphical interface design

Figure 3a shows a typical control screen of the original interface for one subsystem of the plant, in this case, the Chemical and Volume Control System (CVCS). Multiple objects with bright, contrasting colors compete for the operator's attention on the cluttered screen. In many places in the interface, red is associated with a state of alarm or failure. However, this association is undermined by the red color of some valves, pumps, and indicators which are operating normally (red means valve closed; the same color pattern used in the reference plant). Additionally, the red components are highly salient, even when the components do not require operator's attention. Excessive labels contribute to clutter. For example, the blue RCP Seal information box displays the same variables for each of the three RCP seals, but uses nine labels – one for each variable display field. It increases the visual distance between readouts, making comparisons of the values more difficult. The high salience of the large pump icons detracts from the operator's ability to perceive other elements on the screen. They are not frequently manipulated and they only display two pump states (on and off). The sharp contrast between the white lines representing the pipes which connect system elements and the black background contributes to the clutter of the screen without providing much information. The white-on-black color scheme is also used for pump and valve labels, as well as the system variable values. The similarity in color detracts from the salience of these labels and values. Flow directions of are not clearly indicated. The lack of distinction between pipes with and without flow does not contribute to the principle of pictorial realism, i.e., that a visual representation should accurately symbolize the entity it is intended to represent. To determine the path of coolant, operators must trace the white line pumps through which the line passes to ensure that all are open or on, respectively. While the on/off color distinction is clear, there is no redundant indicator of a valve's state, nor does the interface support the synthesis of individual valve states into an overall depiction of flow; each valve must be independently analyzed, increasing the operator's cognitive load. Label legibility is poor due to all-capital text. This also increases label's space requirement without providing additional information. Also, the shine used to produce the 3D graphical effects for the tanks and reactor core decreases contrast and reduces legibility for the white labels that overlay these graphics. There are many different unit names for the same physical variable (e.g., gallons/minute, liters/second and Kg/second), and are many variables without units. The positions of variable values display and related components (pumps, valves) are not uniform among displays, and the same lack of uniformity and consistence appears on the graphical representation of the plant components. Some plant components are not correctly identified and labels positions and formats are not consistent across displays.

There are also many problems related to navigation among displays. The navigation process using the arrow buttons is not clear, because operators don't know the display they will go on and the History/Previous buttons are not working properly (indicating the previous navigated displays). The interface design does not highlight which elements (e.g., pumps, valves) can be manipulated, which are locked out or which are automatic. The operator may be operating under the assumption that a certain valve can be manipulated, finding out latter, when trying to manipulate the valve, that this it is not possible.

Operators show difficult with the navigation using graphic links. Links between some displays do not represent clearly the process flow. Therefore operators always returned to the Main Menu display, searching for the adequate navigation button, because they prefer the navigation buttons rather than graphic links. This back and forth situation augments navigation time between the displays.

Plant component control (ON, OFF, START, INCREASE, DECREASE, STOP) starts with a mouse click on the equipment icon (valves, pumps). After that, a pop-up window appears on the screen, showing the respective control buttons. Then, the control operation should be carried out by clicking on the respective control button. However, observing the control actions of the operators in valves and pumps, we note that sometimes the pop-up windows appeared on the process viewing area of the displays (not in the control panel area), covering plant variables, and interfering with the readings of displayed information.

The redesigned interface (fig. 3 b) is based on the deficiencies noted in the evaluation. They include improved aesthetics and mock-up designs of new functionality. While we have not coded the components into the simulator software, we do not expect significant compatibility problems. The components consist of borders, text boxes, and colors – all of which are supported by the simulator's graphics builder software. The component functionality is also expected to be compatible, as it largely mimics functions (such as linking, highlighting, and displaying real-time system variable values) observed in the original simulator.

Issues with the legibility of labels were addressed by using mixed-case fonts which use less space and provide redundant coding of written information: the shape of the words provides another cue for recognition, aside from the sequence of the letters. To further aid legibility, the 3D graphical tanks, pressurizer, and reactor core were replaced with simpler, flat representations. This allows for increased legibility of the labels, as well as the inclusion of a graphical indicator for the fluid level in the Volume Control Tank (VCT), Pressurizer (Prz), and Reactor Core. The graphical indicator does not require much visual space on the screen, and provides the operator with redundant information on the fluid level of the component. Understanding the context of a reactor core coolant level of 6.5 meters, for example, is aided by the blue bar showing the level of fluid relative to full (top) and empty (bottom) states. (see fig. 3 b).

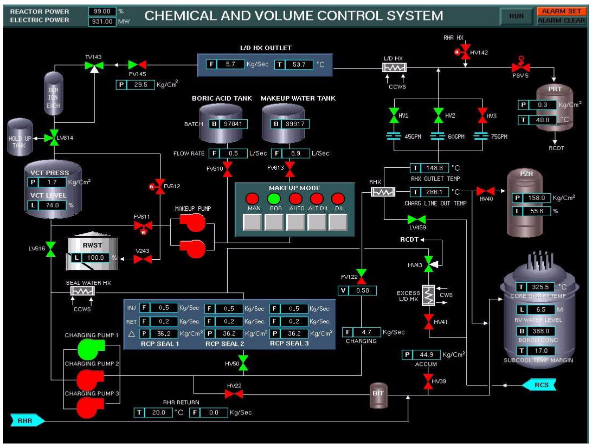


Figure 3a: Original simulator CVCS display.

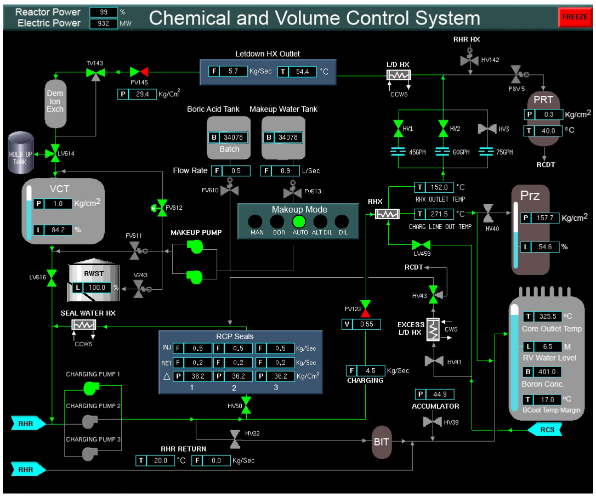


Figure 3b: Graphical improvements on CVCS screen.

The changes aim to improve operator situational awareness, and reduce the likelihood of human error. We remedied the overload of red icons by updating the valve and pump color scheme. Grey is used to reduce salience of closed valves and pumps which are off. Redundant coding is provided by rotating closed valves perpendicular to the pipe, while open valves remain parallel. The size of the pump icons is reduced. While still easy to locate, the off pumps and closed valves do not attract unnecessary attention from a broad overview. The frequently manipulated variable flow valves remain unchanged, providing distinction that helps the operator to quickly locate them. We also simplified the controls for the green “Makeup Mode” control box in the center of the screen. The circular indicators now serve as buttons as well as indicators, obviating the need for the Grey buttons. Also, now only the indicator showing the current mode is lit green. The other indicators which were previously red are toned down to black, so that they do not distract the operator. The RCP seal information box has also been simplified to bring the variable displays into closer visual proximity, and excessive labels have been removed to decrease clutter. The pipes have been re-colored to decrease the salience of pipes which with no coolant flow and to emphasize the pipes with flow. Pipes with coolant flow are bolded and shaded the same color green as the switched-on pumps and open valves. As a result, the emergent feature is a green circuit where there is flow of reactor coolant. The pipes with no flow have been subdued from white to Grey so that they will not interfere with the reading of labels and variables.

Alarm system design

When an abnormal state of a variable occurs, the simulator initiates an audible alarm, as well as a flashing red “Alarm Set” indicator at the top right corner of the screen in use. The alarm indicators are located on two separate specific alarm screens. They are arranged as tiles in a grid where active alarms are indicated by a flashing red tile (fig 4 a). This arrangement reproduces in the simulator the main alarm annunciation tiles used in the reference (analog) interface of the real plant. The existing system does not support quick alarm identification. The text descriptions on the alarms tiles are written in English abbreviations, which may cause delays in the identification for Portuguese speaking operators. The alarm set indicator does not provide any detailed information about the nature of the alarm which is sounding (the same situation that occurs in the actual plant). The operator must always navigate to both alarm screens to determine which alarms were activated. Additionally, the grid arrangement has no apparent organization or order. Related alarms are not grouped on the screen nor are alarms divided logically across the two alarm screens. Finally, all alarms are displayed identically, making it difficult to distinguish between alarms on the basis of severity and importance. All alarms are annunciated by the same sound.

The new prototype interface includes an extensive revision of the original alarm system. The major changes are captured in the revised alarm screen (fig. 4 b). The alarms have been divided into two panels, distinguishing reactor and turbine trip alarms from all others. Within each panel the alarms are organized by the location of their activator in the system. For example, the charging flow indicator is located on the CVCS screen and hence, on the alarm screen, it is under the CVCS column heading. Each alarm tile is a dynamic interface component. This reduces the required number of alarm tiles, allowing all of them to fit on one screen. Instead of a button each for pressurizer pressure high and pressurizer pressure low, the redesign simply uses pressurizer pressure. Depending on the alarm (high or low), the alarm tile displays the appropriate text. Each sounding alarm tile also keeps track of how many seconds since the alarm was set off using a small counter in the upper-left corner of the tile. The trend graphs on the alarm screen saves time and provides better diagnostic information. The acknowledging system has also been improved to allow single-alarm acknowledgement (by clicking on a sounding alarm tile), while retaining the "ACK" button to acknowledge all alarms.

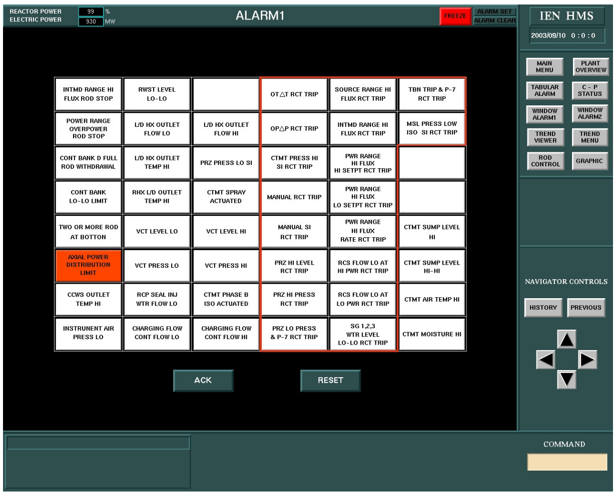


Figure 4a: Original alarm screen.

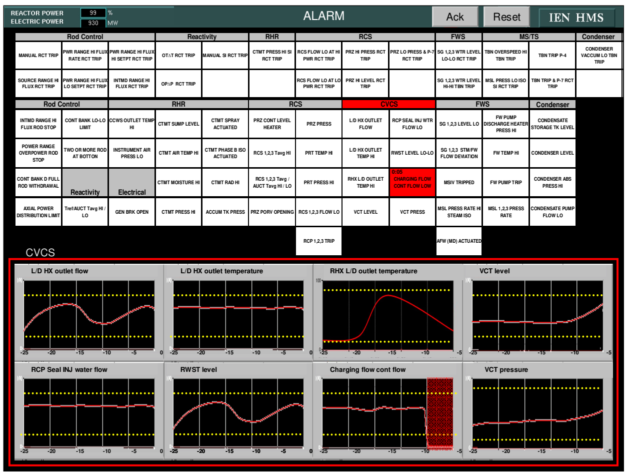


Figure 4b: Redesigned alarm screen.

Each alarm tile acts as a link; clicking the sounding alarm tile navigates to the appropriate screen. On the relevant screen, a red box flashes several times, drawing attention to the area triggering the alarm. Additionally, the alarms relating to the current screen are displayed in chronological order of occurrence as tiles to the right of the schematic diagram. Clicking on these tiles flashes the red box several times box around the area of concern. The navigation buttons have been revised to provide easier access to all the operations screens. While the system is in an alarm state, the related navigation buttons at the bottom of the screen are displayed in red, effectively doubling as an alarm overview. Clicking on the red alarm button navigates to the alarm screen (fig. 4 b).

Digital procedure system design

Procedures guide the operators as they face unfamiliar situations. The simulator uses hardcopy procedure manuals in the form of one-dimensional checklists and step-by-step guides. Non-compliance with procedures was observed frequently. In these situations, operators often improvised around the formal procedures to achieve their system goals, which in some cases can enhance system safety. We observed one operator consistently using a hand-written sheet to aid him through various procedures. The procedures are often constraint-based, requiring the operator to maintain multiple system variables within a specified range. The current interface does not support this task. Instead, it relies on the operator's cognitive ability to monitor system variables and recall acceptable ranges which change frequently during operation. For example, one procedure requires the operator to locate two variables, manually calculate the difference, and judge whether the difference exceeds a safe upper bound which depends on the current mode of operation. Finally, the layout of data in the simulator is inadequate for perceiving and comparing the rate at which a variable of interest is arriving at its limit.

Due to strict procedural adherence requirements, instead of requiring decision support, operators often benefit from tools that reduce errors of omission. The Procedure Guidance Component (PGC) supports operator's process control effectiveness, by converting the procedure manual into an online, navigable guide (Fig. 5). Clicking on any procedure in the left column produces a detailed text description of the procedure. It also reports relevant system statistics and links to useful screens elsewhere in the simulator. This tool adds interactivity to what was previously only a hardcopy procedure manual.

The second component, the Emergency Guidance Component (EGC), is used during emergencies in which the root problem is unknown. The EGC is a reworking of the Strategic Manual Operations flow diagrams provided by LABIHS (for example, see Fig. 6). Clicking on event objects on the left provides response instructions on the right. The operator may scroll up or down through the flow diagram and response instructions using the click and drag technique common to document viewer applications. The continuity provided through the scrolling feature obviates the need for page turning, which takes time and artificially divides what, in reality, is a continuous process. The logic that runs the simulator can be used to support the EGC. Because some decision nodes are based on system variables, the system can often suggest an appropriate decision based on the current system state. The system's suggestion is displayed in a green box to the right of the flow diagram and above the response instructions. It includes the suggested action and the rationale for proposing it. In addition, the operator can trace the decision path because the system fades the paths which have not been taken to a neutral grey, leaving a bold black decision path. Digitizing the emergency procedures enables the implementation of additional support features. The response instructions often involve "if-then" statements. For example, if the pressurizer level reaches 8 meters, then open valves X and Y. Because the simulator knows system variable values, it can guide "if-then" decision-making by placing a red box around “then” actions when the “if” conditional is met.

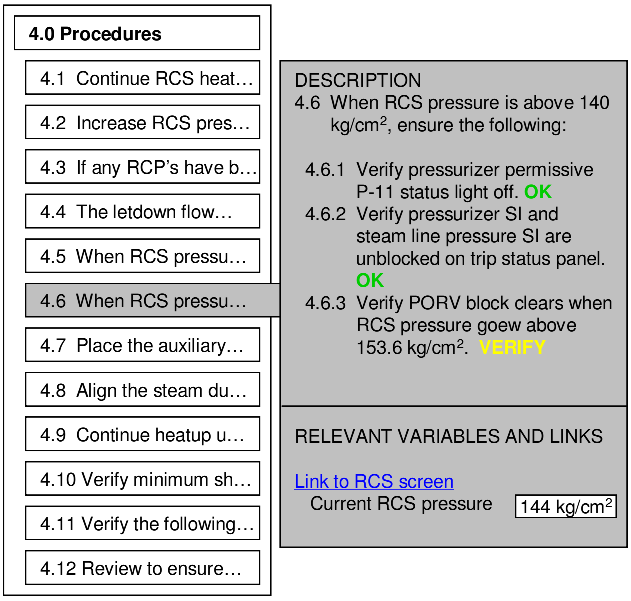


Figure 5: Procedure component guidance.

The Procedure and System Overview (PSO) screen was created to display the PGC and the EGC (Fig 6). The operator may tab between the PGC and the EGC, which reduces short-term memory requirements when compared to hardcopy procedures. On the right side of the PSO, graphical representations of relevant variables are displayed. These are dictated by the current procedure. For example, during a Loss of Coolant Accident (LOCA) the system will keep track of main system pressure, pressurizer pressure, etc. In addition to providing support during emergencies, it aids accident prevention by supporting operator awareness.

In the hardcopy procedures, decision nodes do not have any response instructions because they are implicitly “ifthen” nodes. The digital version shows these “if-then” relationships efficiently by displaying them in the response instructions panel. The response instructions of action nodes include “if-then” relationships as well. Some “if” statements refer to the system state (e.g., if valve X is open) while others ask the operator to wait for a variable to reach a set point before taking action. Unlike the hardcopy version, the new system displays these variables proximally and outlines in red the response instructions when the “if” conditions are met.

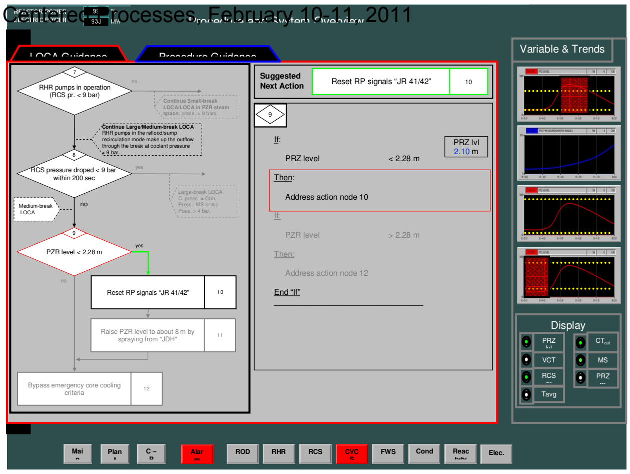


Figure 6: Procedure and System Overview screen displaying the EGC.

Evaluation of the new interfaces

We evaluated operator performance in the new designed interface (figure 6) during accident simulations (Loss of Coolant Accidents – LOCA and Steam Generator Tube Rupture - STGR). A LOCA occurs when there is a pipe rupture in the Reactor Coolant System and the STGR accident occurs when there is a leak in the steam generator tubes. The old LABIHS interface design provided the performance benchmark.

Initially we measure the time that operators need to identify the accident using both interfaces. The time interval between automatic reactor shutdown (reactor trip) and the correct accident identification is very important for a safer operation (Carvalho & Oliveira, 2009). When the reactor is tripped, the operators carry out the standard post trip actions, according to emergency procedures to identify what accident happened, in order to define adequate actions to keep the system under control and minimize the damage that the accident may cause. Using the data obtained from simulator logs it was possible to measure the time interval from the reactor trip until the correct accident identification in both interfaces.

The time spent by the operators to identify the LOCA and SGTR accidents, through the existing interfaces, was 362 seconds and 490 seconds, respectively. The time spent by the operators to identify the LOCA and SGTR accidents, through the new interfaces, was 338 seconds and 428 seconds, respectively. The results show that the time interval from the reactor trip until the identification of the SGTR and LOCA accident decreased when the operators used the new interfaces to identify the accident. The number of screens used during the identification also change. In the existing LABIHS interface the SCO used 13 screens to identify the LOCA and 25 to identify the STGR. In the new interface this numbers fall to 8 and 10, respectively, showing a considerable reduction in navigation actions.

In another experiment, after the LOCA identification, operators are tasked with bringing the system under control by following a LOCA flow diagram procedure. Currently this diagram is available in hardcopy and portable document format. The format requires the operator to shuffle among various pages. The flow diagrams and the response instructions are located on separate pages, either requiring the operator to flip back and forth at least once per node or to take up desk space by laying them side by side. The standard hardcopy procedures are bound, therefore requiring the flip method. Given a medium-break LOCA, to get to step 12 of the diagram requires at least 4 flips between the diagram pages and the response pages and viewing 23 pages (2 diagram pages and 21 response pages). Using the new design, operators can see the flow diagram, the currently selected node's response instructions and the alarm screen together (fig. 7). The redesign requires no page turns, and because it is linked to the alarm system, the operator does not have to search for the appropriate binder or page number to carry out the actions.

Figure 7: Operator working with the redesigned interfaces.

Discussion and lessons learned

We believe that the performance evaluation of the operators` activities in real work is absolutely necessary for human system interface evaluation in nuclear industry.

Activity can defined as the set of behaviors and resources that operators use to accomplish their goals during daily work. Traditional ethnographic methods enable the understanding of activities through observation of communications, gestures and postures. Using ethnographic methods, an observer locates classes of behavior that are recognizable and repeated during work. The methods also allows the observers to identify not only the previous described tasks (prescribed work), but also side activities not formulated in the frame of the task description (Marmaras and Pavard, 1997). The data obtained through direct observation, or with the aid of cameras and audio recorders, is the set of signals picked up by the operators in the information field and how they use these signals to manipulate the control room interfaces. A further analysis of the data set obtained, can show how operators transform the interface information into actions and decisions (Carvalho et al., 2006).

However, the most of methods currently used (including we use in this research) were adequate for describing individual activities developed in a well-defined sequence. However, the work in a NPP control room involves multiple and often conflicting (in goals and time) lines of activities (Carvalho et al, 2006). There are many differences between prescribed and described tasks and real work activities (how the tasks are actually done). Even in a rigid work setting like nuclear power plants, the actual work in control rooms is characterized by adaptations, improvisation and ad hoc procedure modifications (Carvalho et al., 2007; La Porte & Thomas, 1995) because the work demands and resources available rarely correspond to what was anticipated when the task was developed, thereby rendering the task description or operational procedures unworkable (Hollnagel, 2006). Therefore, using the traditional observational methods it is very difficult for the observers to capture the multiple actions pathways of real work activities, describing the many simultaneous tasks and tasks adaptations that people have to do to cope with reality.

Another methodological difficult is the collective/collaborative characteristic of the work done in a NPP control room. The observation procedure normally used is suitable when there is one observer and one worker. However, most work done in control rooms involves multiple operators who use many different cooperative mechanisms (Vidal et al, 2010). Therefore, for an adequate observation of the real work, we need tools to support an observation procedure in which many observers, in collaboration, are able to observe the activities of many subjects (Junior et al, 2010).

Conclusions

The human centered approach in complex industrial system design, evaluation and validation should be applied in the design process in which the system is produced, and in the system itself. In this research we investigate the human system interface of a nuclear power plant simulator to compare design solutions during the early design phase. The methodology used was based on observations of the operators' performance in the LABIHS simulator. Performance evaluations based methods can be used considering the fact that the appropriateness of a given system expresses itself in the quality of the overall performance of the system is assessed. Normally, performance evaluation is something that is carried out towards the end of a given design process. The LABIHS facility aims to conduct the performance evaluation earlier in the design process. A specific goal of LABIHS is to enable the evaluation of system performance as early as possible. Considering that the reference plant human system interface design has not formally started yet, this objective was already achieved with this research. Even considering that is very difficult to say when the performance of a cognitive system is at an acceptable level, our evaluation has shown some improvement possibilities in the existing design. Some of them related to basic human factors design principles such as:

- Displays with information that are difficult to read (inadequate font sizes and formats, color contrast etc.);
- Cluttered or overloaded displays with many numeric information – graphic information would be better;
- Inadequate icons size considering their function;
- Confusing and unstructured presentation of displays with set points and actual parameter values, leaving the task of searching and detecting such deviations to the operator, instead of directly showing deviations of actual values from set points;
- Static information presentation where a presentation of past dynamics (e.g. trends) and future developments of process parameters

(prediction) would be required for an effective task performance;
- Mix of different media to present operational information – digital displays and paper procedures – requiring different cognitive resources to cope with.

As expected the performance evaluation has shown that the design solutions used (alarm systems, procedures, graphic displays) actually have effects on the usage. Therefore we reinforce the claim of the human factors and ergonomics community that the design solutions should be made considering the appropriate use of the system, emphasizing that work practices in real settings. What we really need are systems that support actions of human operators, and their ability to adapt and adjust to novel situations. To do so, systems must be designed considering that the user, and the usage of the system need to be taken account in all the phases of the design process, from the design of process technology to the design of user interfaces, in a user-centered or activity-based design process.

Acknowledgments

The authors gratefully acknowledge the support of the Brazilian Research Council (CNPq) and of Rio de Janeiro Research Support Foundation (FAPERJ). The research was performed at Instrumentation and Human Reliability Division of the Nuclear Engineering Institute, Brazil (DICH / IEN).

References

Brunélis, T., and Blaye, P. (2008). Towards a human centred methodology for dynamic allocation of functions. *Proceedings of the Third International Conference on Human Centered Processes (HCP 2008)* (pp. 243-256).

Carvalho, P., Santos, I., Gomes, J., Borges, M., Guerlain, S. (2008). Human factors approach for evaluation and redesign of human–system interfaces of a nuclear power plant simulator. *Displays 29*, 273-284.

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

Carvalho, P., Vidal, M., Santos, I. (2006). Safety implications of some cultural and cognitive issues in nuclear power plant operation. *Applied Ergonomics 37 (2)*, 211-223.

Carvalho, P., Oliveira, M. (2009). A computerized tool to evaluate the cognitive compatibility of the emergency operational procedures task flow. *Progress in Nuclear Energy, 51,*409-419.

Flach J., Hancock P., Caird J., Vicente K. (Eds.) (1995). *Global perspectives on the ecology of human-machine systems*. Hillsdale, N. J.: Lawrence Erlbaum Associates.

Hancock P. & Chignell M. (1995). On human factors. In: J. Flach, P. Hancock, J. Caird, K.J. Vicente, (Eds.) *Global perspectives on the ecology of human–machine systems.* Hillsdale, N.J.: Lawrence Erlbaum Associates,14-53.

Hollnagel, E. (1985). *A Survey of man-Machine System Evolution Methods.* (HWR 148) Norway: OECD Halden Reactor Project.

Hollnagel, E. (2006). Task Analysis: why, What and How. In: G. Salvendry (Ed.) *Handbook of Human Factors and Ergonomics-3rd ed.* New Jersey: John Wiley & Sons.

Junior L. C., Borges M., Carvalho, P., (2010) A Mobile Computer System to Support Collaborative Ethnography: An Approach to the Elicitation of Knowledge of Work Teams in Complex Environments. *Lecture Notes in Computer Science, Volume 6257/2010, 33-48.*

La Porte, T. & Thomas, C. (1995) Regulatory Compliance and the Ethos of Quality Enhancement: Surprises in Nuclear Power Plant Operations, *Journal of Public Administration Research and Theory*, n.5, pp. 109-137.

Marmaras, N. and Pavard, B. (1997). A Methodological Framework for Development and Evaluation of Systems Supporting Complex Cognitive Tasks. *Journées Européenees des Techniques de l`Informatique, 8, 13-20.*

Nachreiner F., Nickel P., Meyer I. (2006). Human factors in process control systems: The design of human–machine interfaces. *Safety Science,44*, 5-26.

Nielsen, J. (1993) *Usability Engineering*. Boston: Academic Press.

O´Hara J., Higgins J., Stubler W., Goodman C., Eckinrode R., Bongarra J. and Galletti G. (1994). *Human factors engineering review program model (NUREG-0711 rev.1).* Washington.D.C.: US Nuclear Regulatory Commission.

O´Hara J., Brown W., Stubler W., Wachtel J. and Persensky J. (1996). *Human-system interface review guideline (NUREG-0700 rev.1).* Washington.D.C.: US Nuclear Regulatory Commission.

O'hara J. & Brown M. (2004). *Incorporation of human factors Engineering analyses and tools into the design process for digital control Room upgrades* (BNL-72801-2004-CP). New York: Brookhaven National Laboratory.

Santos, I. J. A., Carvalho, P.V., Grecco, C. H., Victor, M. and Mol, A. C. (2005a). A Methodology for Evaluation and Licensing of Nuclear Power Plant Control Rooms. *In Proceedings of the 2005 International Nuclear Atlantic Conference, INAC*, Santos, SP, Brazil.

Sheridan, T. (2002). *Humans and Automation System Design and Research Issue.* Santa Monica: Wiley/HFES.

Vicente, K., Mumaw R., Roth, E. (19997). *Cognitive Functioning of Control Room Operators – Final Phase.* Ottawa: Atomic Energy Canadian Bureau.

Vidal, M.C.R., Carvalho P.V.R., Santos M., Santos, I.J.L. 2009, Collective work and resilience of complex systems. *Journal of Loss Prevention in the Process Industries, 22, 537-548.*

The Design of Preventive Safety Systems: a Cognitive Engineering Problem

Caterina Calefato (calefato@re-lab.it)
University of Turin Department of Computer Science Corso Svizzera 185,
10149, Torino, Italy
Luca Minin (luca.minin@unimore.it), Roberto Montanari (roberto.montanari@unimore.it)
University of Modena e Reggio Emilia, HMI Group DISMI, Via Amendola 2,
42100, Reggio Emilia, Italy

Abstract

In this paper a design framework for preventive safety systems (ADAS) is proposed. The design framework takes into account risk mitigation strategies, advanced driver's model, based on modern approaches and algorithms (machine learning and add-on functionalities), able to capture key aspects of human behavior, such as distraction, and to retain the fundamental characteristics of cognition and decision making.

Introduction

Driver modeling is a scientific area involving several disciplines, such as psychology, physics, computer science, etc. The importance of adding the learning capability to information systems, in order to make them more effective and smarter, is confirmed by the variety of areas in which user's modeling has already been applied: information retrieval, filtering and extraction systems, adaptive user interfaces, educational software, etc.

In relation to the problem formulated above, the aim of this paper is to deeply understand the problem of ADAS (Advanced Driver Assistance System Design) design, the problem of developing an effective driver's and driving model supporting distraction mitigation. Such systems would mitigate the effects of distraction and tolerate the consequences of distraction thanks to a better road and vehicle design (Regan, Lee & Young 2009).

A feasible and promising solution is the use of Add-On functionalities, able to detect driving maneuvers that are indication of distraction, placing them in the framework of a cognitive model of human behavior.

In this paper a design framework for preventive safety systems is proposed, following three main building blocks:

1. **new knowledge about driver behavior**: extensive empirical studies about the sources of accidents and potential counter measures as a basis for the driver model development.
2. **risk mitigation strategies**: implementation of a human error risk based approach;
3. **advanced driver modeling**: development of models for predicting correct and erroneous driver behavior, based on modern approaches and algorithms (machine learning), able to capture key aspects of human behavior, and to retain the fundamental characteristics of cognition and decision making.

The precondition analysis

Accidents occur because multiple factors combine to create the necessary conditions for them. Over the past 30 years, the literature shows a consistent trend in trying to understand accidents in aviation, nuclear power generation, telecommunications, unmanned and manned spaceflight, railroad transport, shipping, healthcare and many other fields (Catino 2002). Regardless of the domain of investigation, there are some crucial questions to which the research is trying to find an answer (Cook & O'Connor 2005):

- *How do accidents happen and what do they mean?*
- *Are accidents foreseeable?*
- *If so, are they preventable?*
- *What role does technology play in accidents?*
- *What role does human performance play?*
- *Are accidents evidence of systemic problems or are they isolated failures?*
- *If accidents are systemic, how can the system be fixed to prevent future accidents?*

What it may be interesting to study is the detection of an accident pre-conditions and which may be the conditions combination (circumstances) that may lead toward an accident. For each combination (that we may label *risk layouts*) a mitigation strategy will be applied in order to avoid a possible accident. "What was lacking was the ability to foresee that circumstances would conspire to create the conditions needed to make these technical features active and lethal" (Cook & O'Connor 2005).

An adaptive system should be able to detect risk layouts and dynamically adapt its behavior in order to avoid accident, but the failure factor in this scheme is *change*. In a complex system formed by a context, a predictable system (including automatic applications) and the human being, the unpredictable factor is the human being behavior and its combination to a certain context.

Although information technology can defend against some types of accidents and failures, the impact of automation on human-machine system performance is a mixture of desirable and undesirable effects (Perry, Wears & Cook 2005).

Systems like ADAS (Advanced Driver Assistance Systems) have great potential for enhancing road safety, but on the other hand the safety benefits of ADAS may be significantly reduced by unexpected behavioral responses to the technologies, e.g. system over-reliance, safety margin compensation and distraction, leading toward an automation failure. The automation failure is a side effect of an effort to produce ‘‘safety’’(Catino 2002).

ADAS: existing applications

In addition to the safety issues associated with the driving task, the proliferation of complex in-vehicle functions itself poses a further challenge for the design of the driver-vehicle interface: one of the current research area in automotive is the development of preventive warning systems, also called ADAS (i.e. Advanced Driver Assistance Systems) adopted with the aim of improving driving safety. These systems are able to detect an incoming dangerous in advance, allowing a time to perform a repairing manoeuvre.

ADAS are aimed at "partly supporting and/or taking over the driver's tasks" (Berghout, Versteegt, van Arem 2003) so to generally provide safer driving conditions. Several functions can be mentioned within ADAS set. In the following, a list of the main relevant ones is reported (Ehmanns & Spannheimer 2004):

- **Lane departure warning**: If certain thresholds (like distance, time to lane crossing) allow a prediction of a lane departure this system warns the driver by means of acoustic, optic or haptic feedback. The detection of the lane markings results from e.g. video image processing. In order to have a robust lane marking detection two needs can be absolved: (i) good visible lane markings have to be provided by the infrastructure and (ii) a robust lane detection sensing system has to be implemented in the vehicles. Both aspects are influencing the complexity of the system on the roadside and the technical level.
- **Near field collision warning**: The near field collision warning includes the detection of especially vehicles in the near field like in the blind spot area. The detection area is very close limited to the vehicle. Suitable sensor systems for the detection of other cars are radar or vision based sensors.
- **Curve & speed limit info**: These systems inform the driver about speed limits and the recommended speed in curves. Possibly the necessary information can be taken from digital maps, image processing communication systems between vehicles and infrastructure.
- **Adaptive Cruise Control (ACC) /Stop & Go:** The ACC and Stop & Go establish a virtual link with the frontal vehicle via a radar-based technology and keep booth vehicle within a safe distance. The main innovation of this systems, that is derived from the well-known cruise-control, is that the distance can be adapted both to the driver's preferences (as in ACC) and to the specific requirements of the urban environment (as in the Stop & Go). In traffic condition as in a queue, the Stop & Go automatically drive the vehicle timely providing vehicles' stops and small movements.
- **Lane Keeping Assistant**: The function of a lane keeping assistant system includes the lane detection and the feedback to the driver if he/she is leaving a defined trajectory within the lane. An active steering wheel can help the driver with a force feedback to keep on this trajectory. The lane is detected by a video image processing system.
- **Local Hazard Warning** : If a hazard occurs far away in front of the vehicle, so that the driver cannot see it, this system will warn him/her. By the means of communication it is possible, to transfer this information over long distances
- **Lane Change Assistant**: Before and during a dangerous lane change process, the lane change assistant will warn the driver. Several stages of such a system are possible from pure warning systems to even haptic feedback at the steering wheel to help the driver following a lane change trajectory.
- **Blind Sport Monitoring:** This function detects if a vehicle is present in the so called "blind spot" area when the vehicle is starting a lane change and/or overtaking maneuvers. A camera is placed into the left rear-mirror and once the incoming vehicle is recognized, a warning is issued to the driver.
- **Obstacle & Collision Warning**: The driver will be warned if a potential collision is detected via radar-based technology (e.g. another car or obstacle). The functional limits of these systems have to be clearly pointed out. The liability problem of these systems grows with the complexity of the detecting scenarios.
- **Obstacle and Collision Avoidance**: This system has an extended functionality compared to the Obstacle and Collision Warning. An autonomous intervention takes over the control of the vehicle in critical situations in order to avoid an accident. Longitudinal and lateral control will be done by the system during the defined time while the dangerous event takes place.
- **Night Vision**: Based on camera techniques like near or far infrared, it allows enhancing the perception of the driver in dark light conditions. The picture of the camera will be shown to the driver by monitors or head up displays.
- **Platooning**: Several cars are connected electronically (e.g. by the means of communication) and follow one after the other in a platoon. An example is the connection of trucks in order to save space, fuel and to increase the traffic flow. As the following vehicles are driven automatically, the system is complex concerning all aspects. The

takeover of the driver at e.g. gateways has to be taken into account as well as the behavior in mixed traffic at driveways.

Designing the trustiness: ADAS research issues

Interaction with these devices is one of the many activities that constitutes driving and so it can represent an additional source of driving-related distraction (Regan, Lee & Young 2009). For example poorly designed collision warning systems may be even more likely to distract drivers; navigation represents a driving-related task with substantial potential to distract (Neale et al 2005), (Dingus et al 1989).

The analysis of ADAS working conditions, architectures and performances leads towards the definition of a proper theoretical framework that is not yet present in current projects.

The reasoning behind is the following: Advanced Driver Assistance Systems, or ADAS, are systems that help the driver in its driving process: they detect a dangerous situation and gives a warning. We can define *analytical* (Andreone et al 2005), the ADAS type that warns the driver suggesting accident-avoiding maneuvers. The ADAS is *behavioral* if it acts in place of the driver, partially taking over a certain driving task (Andreone et al 2005) (Hoch et al 2007)0.

In the case of analytical ADAS we can consider there are two actors playing a role:

- The driver
- The warning system

In the case of behavioral ADAS we can consider there are three actors playing a role:

- The driver
- The *horse* (the artificial system able to drive in place of the driver, see Flemisch et al 2003)
- The warning system

In both cases, analytical and behavioral ADAS, there is a warning systems that detects the dangerous situation and then provides the driver, as safety warning, accident-avoiding maneuvers.

The purpose of these systems is to foresee and detect possible driver's errors and mistakes, due to a misbehavior such as distraction, or resulting from too high workload, missing perception, wrong action/execution or poor operator skills.

The ADAS design is aimed at enhancing the driver's perception of hazards and critical situations (in some cases, by partly automating the driving task as well). Of course the potential of such systems in reducing accidents depends on the effectiveness of their interaction with the driver. For example, in the case of an anti-collision systems it is safety-critical that the collision warning is able to generate the appropriate feedback (e.g. an avoidance maneuver).

Since ADAS can be actually considered recommending systems, the use of an appropriate driver's or/and driving model will improve their effectiveness and consequently, human safety.

Risk mitigation strategies: recommending the accident avoiding actions

ADAS can be considered an application of recommending systems that recommends the driver repairing maneuvers in order to avoid an accident. The most advanced systems are able to directly take part in the driving task, whether the driver doesn't react on time. Also in this case, the systems follows a recommendation formulated by the system itself.

Recommender systems have become a promising research area since the appearance of the first papers on collaborative filtering in the mid 1990s (Hill et al1995), (Resnick et al 1994), (Shardanand & Maes 1995)

Shortly, the recommendation problem is formally represented as a space *S* of possible items that may be very big, ranging in hundreds of thousands or even millions of items in some applications, such as recommending books. An utility function measures the usefulness of each item for a certain user.

In recommender systems, the utility of an item is usually represented by a rating, which indicates how a particular user likes a particular item or how a particular item is appropriate for a certain user, taking care of a set of context conditions. Generally speaking , utility can be an arbitrary function, including a profit function. Depending on the application, the utility can either be specified by the user, as is often done for the user-defined ratings, or is computed by the application, as can be the case for a profit-based utility function (Adomavicius, Tuzhilin 2005).

To each element of the user space *C* can be associated a profile that includes the user characteristics that are relevant for the current application. Similarly each element of the item space S is defined by a set of relevant characteristics.

In recommender systems, utility is typically represented by ratings, therefore, the recommendation engine should be able to estimate (predict) the ratings of the nonrated user/item combinations and issue appropriate recommendations based on these predictions. Extrapolations from known to unknown ratings are usually done by:

- specifying heuristics that define the utility function and empirically validating its performance
- estimating the utility function that optimizes certain performance criterion, such as the mean square error.

Despite of the results nowadays achieved, the existing generation of recommender systems still requires further improvements including better methods for representing user behavior and the information about the items to be recommended, more advanced recommendation modeling methods, incorporation of various contextual information into the recommendation process, utilization of multicriteria ratings, development of less intrusive and more flexible recommendation methods that also rely on the measures that more effectively determine performance of recommender systems (Adomavicius, Tuzhilin 2005).

In the case of services provided on board a car, we can notice that they are rapidly growing. Almost all car manufactures are offering systems that add functionality to route planners, possibly integrated with internet and web access or that support driver in high demanding tasks, in order to increase safety and avoiding accidents. The availability of these add-ons is an interesting opportunity, considering that nowadays the amount of time spent in the car (e.g., for commuting or for work and vacation trips) is very high (Console et al 2003).

If on one hand the driver and the other vehicle occupants can actively use the time spent on the car, on the other hand the use of these services can be distracting and can create serious safety problems (Green 200) (Console et al 2003), contracting societal goals of increasing safety, reducing the number of accidents. As a consequence it is necessary to find a proper compromise between the increasing number and complexity of the services and the need of making the services compatible with the fact the user is driving.

Starting from this consideration, the introduction of personalization and adaptation strategies and techniques should be a feasible solution in the case of services in the car. In fact, by considering the characteristics of the user and the context of interaction, a personalized and adaptive system may tailor the interaction to the way which is most appropriate to avoid distractions, and as a direct consequence, to avoid an accident (Console et al 2003).

In the case of safety-critical systems which should recommend accident avoiding maneuvers, the adaptation of the recommendations to the specific user is crucial, according to the psychophysical parameters that are taken into account (i.e. mental workload, distraction, arousal level, situation awareness). In the case of advanced driving assistance systems one of the most important psychophysical parameter to be taken into account is distraction. The system should be able to assess driver’s distraction in order to estimate accident precondition (risk layout) and recommend driver appropriate actions, or in the case of adaptive automatic systems to perform a proper risk mitigation strategy.

If the recommending engine has not at its disposal a user behavior model, it can formulate recommendation that may lead towards no decisions or wrong decisions.

Whether the system prediction capability is augmented through a user behavior model it is possible to reduce errors and then the risk of accidents. This consideration is of paramount importance in complex safety critical systems as avionic and automotive, that commonly use different kind of recommending services.

The design of cognitive preventive safety systems

Driving is considered as a complex and multitasking cognitive activity that can be summarized by four main sub-processes: perception, analysis, decision and action. To be performed, each phase presumes the achievement of the previous one. That said, it is likely that the demands of one element of driving will interfere with another element.

ADAS new technologies have great potential for enhancing road safety, however, when an ADAS or an In-Vehicle information system (IVIS) is activated and the driver is asked to interact with it, the driver him/herself is distracted from the driving task, that is, his/her attention is moved from the driving task to the secondary task. A relevant part of vehicle crashes are estimated to

The Driver Assistance Systems have to be able to adapt their action to the context and to the driver and vehicle status. Thereby, they need a model of human behaviour that takes into account the model of the system performance and that is able to detect and classify driver’s intention and distraction, in order is essential to facilitate operating mode transition between users and driver assistance systems.

The need of an effective user model is a requirement for any recommending system, as faced and confirmed by the domain literature on user modeling and automatic recommending systems. This requirements is crucial for any recommending system that has to cope with time-criticality, that directly affects safety.

ADAS applications are examples of such systems and they represent a challenging test bed for the implementation and validation of user behavioral modeling systems realized by means of Machine Learning techniques.

Basically, the human behavior is characterized by the interactions between driver-vehicle and driver environment.

The first interaction is related to how the driver interacts with the vehicle and all systems and sub-systems on-board.

The second interaction is related to how drivers perceive and process the data coming out from the surrounding scenario.

Hence, the driver model should be adaptive to different drivers’ style and preferences as well as to the external environment (including learning both from the driving experience and from the surrounding conditions), but overall it should allow to assess and foresee distraction.

Preventing distraction permit to prevent driving errors and accident risk, as a consequence, a risk based design approach (that follows a risk mitigation strategy) is crucial for the design of vehicles and transport systems in order to guarantee safety and efficiency of human mobility.

User modeling (UM) aims at improving system effectiveness and reliability by adapting the behavior of the system to the needs of the individual.

The importance of adding this capability to information systems is proven by the variety of areas in which user modeling has already been applied: information retrieval, filtering and extraction systems, adaptive user interfaces, educational software, safety-critical systems .

Machine learning (ML) techniques have been applied to user modeling problems for acquiring models of individual users interacting with an information system and grouping them into communities or stereotypes with common interests. This functionality is essential in order to have a useful and usable system that can modify its behavior over time and for different users (Langley 1999). As elicited from literature (Tango Botta 2009), (Tango et al 2009) (Tango et al 2010) there is a trend in choosing machine learning techniques in the study of modeling of human behaviors, that is non-deterministic and highly non-linear.

Driver Assistance Systems have to handle crucial aspects like timing and warning, therefore the development of an algorithm for the personalization of such aspects that takes into account for example is needed.

Regarding the drivers' intention prediction, several models have been proposed aiming at reproducing in a virtual environment how the drivers could behave according to specific Driver, Vehicle or Environment conditions , that is DVE model (Tango et al. 2010) . In the domain literature there are different approaches like:

- the IPS (Information Processing System), which has been applied in almost all technological fields to describe human interaction with control systems, at different levels of automation (Neisser 1967)
- the PIPE (Perception, Interpretation, Planning and finally Execution) based on a very simple approach that assumes that behaviour derives from a cyclical sequence of four cognitive functions in brackets. (Cacciabue 1998)

The development of a model of the human machine system is driven by the model of the Driver, which is the most complex element of the system.

Concerning the design of algorithms used to represent the Driver behaviour, previous and ongoing studies propose different approaches based on the real-time monitoring of the drivers' performance (Lolli et al 2009) (e.g. variation of the position on the road, speed, steering wheel movements) or the drivers' physiological status performing primary and secondary tasks (e.g. eye gaze, eye movements, heart frequency variation, galvanic skin response, etc.) (Ji & Yang 2002).0

Basing on these information, these approaches allow to predict specific drivers' profiles (e.g. stressed, aggressive, tired, distracted, high workload etc.) and are developed following machine learning approaches. Thanks to machine learning, information on drivers' profile can be automatically extracted from data, by computational and statistical methods applied to observable information (e.g. drivers' performance data).

On one hand the assessment of the driver's status and consequentially the prediction of his/her next behaviour is easier and most successful using driver's physiological data, as for example the eye gaze and the eye movements measured by means of eye tracker. On the other hand eye-tracking is an intrusive measurement system of distraction, it represents a further equipment and a further cost that stand in the way of a next future mass-marketing. What is really interesting and challenging is to obtain a driver index analysing driving performance data, realising what it may be considered an ADD-On Functionalities.

Understanding driver's maneuvers by the use of Add-On Functionalities

Research in driver comfort and performance improvement understanding driver's maneuvers is very active. Usually, these targets are achieved through the installation of further in-vehicle sensors and devices (Lolli et al. 2009) .

An alternative is the use of the so-called Add-On Functionalities (AOF). They do not require new sensors but only information coming from the on-board network and sub-systems (e.g. the elements of chassis, suspensions, steering angle, etc.). This information is computed in a well tuned algorithm and the results provide some added-value supports to the drivers as AOFs for pre-crash detection. They prepare the vehicle to the impact in critical situation which can not be avoided, for example, by pre-tensioning the seat belt (Lolli et al. 2009). They can be used also to indirectly infer form the driving & driver data crucial information about driver's behavior like distraction, workload and arousal.

Add-On Functionalities can in fact be divided in two main categories:

- **Driving Behaviour**: i.e. Add-On Functionalities related to driving performance. Main objective of these AOF is to estimate driving conditions concerning road, dynamics and current manoeuvre.
- **Driver Behaviour**: these AOF deal with driver current state, mainly intended as mental effort (or workload) related to the driving task.

In order to be implemented, an Add-On function has to satisfy two conditions:

- All Add-On inputs must be available and shareable.
- At least one Add-On output can be received as input by a vehicle device.

An Add-On Function with m inputs and n outputs is defined as:

$$(x_1, x_2, \dots x_n) = F\ (y_1, y_2, \dots y_m)$$

or in a concise form:

$$\bar{x} = F\ (\bar{y})$$

If Y is set from available device output and X is set from available device inputs, the two aforementioned conditions can be expressed as follows:

In order to be implementable, an add-on functionality F, defined as

$$\bar{x} = F\ (\bar{y}),$$

with $\bar{x} = (x_1, x_2, \dots x_n)$ and $\bar{y} = (y_1, y_2, \dots y_m)$ must satisfy the following conditions:

1. $\bar{y} = (y_1, y_2, \dots y_m) \in Y^m$
2. $\exists\, x_i \in \bar{x}, i = 1, \dots, n | x_1 \in X$

where X and Y are respectively set of devices inputs and outputs.

In the study presented in (Lolli et al. 2009) the efforts are focused on the use of AOF in order to collect useful data that will be employed to fill in the driver's profile and status and to log his/her performance. All this information will be used as a trigger for the adaptive automation applied to the in-vehicle information systems (IVIS) or to the Advanced Driver Assistance Systems (ADAS). The development of AOFs has been tested for tuning in a simulated environment using data Matlab/Simulink vehicle model. Identified AOF outputs to improve driver performance and safety are the following:

Driver Stress (DS). From previous studies (Gulian et al. 1989), stress can be related to lateral vehicle control. Then, it could be derived by the steering activity: several indexes for lateral control monitoring were provided, in particular: HFS (High Frequency Steering) (McDonald Hoffman 1980), RR (Reversal Rate) (McLean Hoffman 1975) and SAR (Steering Action Rate) (Verwey 1991).

According to the level of stress it is possible to find strategies to assist the driver in particularly demanding maneuvers by modifying the steering force feedback, braking behavior and inhibition of secondary task to avoid possible safety-risk situations.

Particularly, modifying the steering force feedback or the braking behavior has an interesting impact in the human-machine interaction, because the feedback to the driver is haptic Empirical test confirmed that driving performance significantly improved when the system activated the force feedback models.

These results compared with data arose from other studies in the literature (Steel Gillespie 2001) suggested that, using an intelligent, haptic steering wheel rather than a traditional passive steering wheel, drivers are better able to closely follow a reference path while requiring fewer visual cues (Minin et al. 2009)

For a decade researchers (Bertollini Hogan 1999) have been finding that the presence of haptic feedback on the steering wheel could help drivers to perform a visually-guided task by providing relevant information like vehicle speed and trajectory. Referring to the augmented cognition field, we can assess that when using a haptic assist steering wheel rather than a traditional passive steering wheel, drivers are better able to follow a reference path and at the same time, they required fewer visual cues (Griffiths Gillespie 2004).

Traffic congestion (TC). Through the monitoring of longitudinal vehicle parameters (e.g. brakes and speed behavior) it is possible to state whether drivers are driving in a heavy traffic situation or not. In this case, strategies could be elaborated by optimizing both the engine-fuel management at a low speed and the driver comfort aiming at reduces the level of stress (Lolli et al. 2009).

Road Conditions (RC). The knowledge of road profile characteristics through information coming from specific sensors (i.e. suspensions, Roll Rate Sensor, Pitch Rate Sensor, Sound Sensor Cluster, ESP intervention) allows alerting active suspension system to smooth the impact of obstacles, helping drivers to reduce the effect of this critical situation (Lolli et al. 2009).

The table below (Lolli et al. 2009) reports inputs coming from the vehicle chassis and used to compute AOFs.

Table 1 AOF inputs, parameters and outputs (Lolli et al. 2009).

AOF inputs from chassis	AOF parameters	AOF outputs
Steering Angle	HFS, SAR, RR	DS
Speed	Deceleration Jerks (DJ)	TC
Brake Pressure	Braking Frequency (BF)	TC
Accelerator Displacement	Accelerator Frequency (AF)	TC
Gear number	Gear Index (GI)	TC
Z acceleration	Frontal Obstacle Preview (FR)	RC
Roll rate	Roll Index (RI)	RC
Pitch rate	Pitch Index (PI)	RC
Suspensions Displacement	Frontal Obstacle Preview (FR)	RC

Inputs were selected to compute specific AOF parameters with the aim to describe driver stress, traffic congestion and road conditions. AOF outputs are the result of the balanced sum among parameters; for instance, the Driver Stress (DS) index (2) was developed as follow 0:

$$DS = (RR \times c_{RR}) + (HFS \times c_{HFS}) + (SAR \times c_{SAR}) \quad (2)$$

Where $c_{RR} + c_{HFS} + c_{SAR} = 1$ are the coefficients to be tuned in order to define the final value of the AOF output. Each AOF outputs (TC, DS and RC) and their related computed parameters (see "AOF parameters" in Table 1) were developed in a simulated environment using Matlab/Simulink (www.mathworks.com).

In order to test and tune these parameters, AOF models were interfaced with a Matlab/Simulink simulated vehicle. The whole model (AOF and simulated vehicle) is fed by real driving data coming from a professional driving simulator. Specific tests were carried out, aiming to provide driving situation where each parameter varies significantly; then, their effectiveness was assessed.

According to the result, information monitored by AOF outputs (DS, TC and RC) will be used as a basis for the development of strategies aiming at improve driving performance, safety and comfort.

AOF test and tuning: Driver Stress

In the following, test and tune of Driver Stress (DS) parameters are described. The DS index is the balanced sum of steering angle based parameters, in particular: SAR (Steering Action Rate), HFS (High Frequency Steering), RR (Reversal Rate). A default tuning of coefficients related to these parameters has been applied (C_{RR} = 0,4; C_{HFS} = 0,2; C_{SAR} = 0,4). The effect of the tuning was assessed by comparing the expected stress profile in certain pre-determined conditions (i.e., the points numbered from 1 to 5 in Figure 1) with the drivers' steering activity (Lolli et al. 2009).

Two tests were conducted on a driving simulator where 12 subjects, each of them was asked to drive for 10 minutes. Test environments were characterized by roads with different curve radius, variable visibility (from 100 to 4500 m) reproduced with fog and variable traffic (from 10 to 50 vehicle/km) (Lolli et al. 2009).

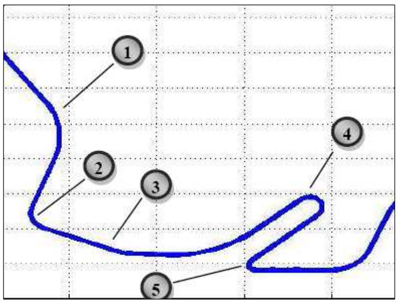


Figure 1 Driving scenario (track) (Lolli et al. 2009).

In order to increase the steering activity, subjects were also expected to complete a secondary visual task, which consisted in pressing the left/right side of a touch-screen on the left side of the vehicle cabin, according to the position of a circle displayed among smaller ones. Data regarding DS parameters were collected and compared with the steering activity in a specific point of the road (the number from 1…5 circled in Figure 2) (Lolli et al. 2009).

Due to the large amount of information, the comparison focused on a sub-sample of 3 subjects; mean values of steering activity and DS parameters are then depicted. An example of steering angle activity is showed in the top side of Figure 2, while Driver Stress index in the bottom. Both are related to scenario coordinates (x-axis). As foreseen, an increased steering activity leads to higher Driving Stress values 0. These peaks are pointed out in particularly critical situations (due to curves, high traffic, low visibility), highlighted in the circled number of the figures.

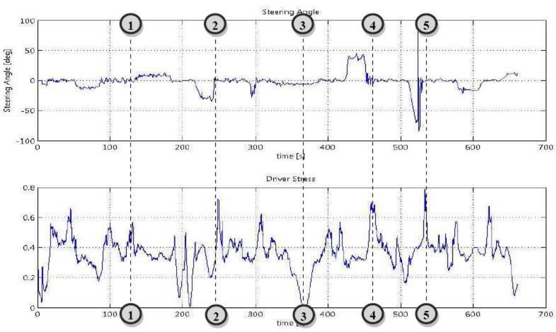


Figure 2 DS compared with steering angle (Lolli et al. 2009)

Results show that the first DS parameters' tuning produced an index able to detect driver stress status. Since the analysis was carried out on a small sample of subjects, in order to increase the significance of the tuning the above deployed comparison will be extended to all subjects (Lolli et al. 2009).

Driving Stress Index appears to be a good starting point for developing a parameter able to detect driver status in real-time even if a deeper test and tuning activity is needed. Furthermore, together with the other AOFs (Traffic Congestion and Road Conditions) can be easily implemented on a vehicle ECU (Electronic Control Unit) or a DSP (Digital Signal Processor) (Lolli et al. 2009).

Conducted simulations are just preliminary tests to find out the most promising indexes. It is very important to improve techniques aimed at monitoring of the status and the performance of both the driver and the vehicle in order to gather all data useful to customize the information provision strategies of the in-vehicle devices. In this way the vehicle may adapt his status, improving comfort or modifying driving performances.

This proactive behaviour paves the way to mechanisms able to infer the driver's distraction and situation awareness, allowing the triggering of adaptive automation strategies. The information provided by AOFs are real-time and allows at dynamically implementing such adaptive strategies and referring them both to the path changes and to the driver's status. The AOFs added-value is the prospect to obtain information in a not intrusive way (Lolli et al. 2009).

The framework for the design of preventive safety systems

The aim of preventive safety system is to support drivers, especially in risky and critical situations, or whenever distraction may occur. The first step for the design of a vehicle able to assess the driver status and intentions is the development a model able to explain and reproduce driver's characteristics. Based on the empirical results presented in the previous chapters, this research work aims at developing a little missing piece of the puzzle of the future intelligent vehicles: namely to identify the main elements for a feasible architecture of a "cognitive driving assistance system" which will substantially advance both integrated safety/assistance systems and the cooperation between human beings and highly automated vehicles.

The feasibility of such an architecture has been investigated analyzing:

- The problem of accident precondition analysis
- ADAS existing applications and research issues
- Risk mitigation strategies for the accident avoiding
- Design issues of cognitive preventive safety systems
- The understanding of driver behavior from driving maneuvers by means of add-on functionalities

Hence the framework to develop an effective model of driver's perception will be include four major functional areas:

1. The **core application**, where motion-planning tools including enhanced personalisation, will be used to explore the maneuver space and to ultimately understand the driver, and if needed, to produce maneuvers compatible to human motion.
2. **Improved sensing of driver input**, where the control input ("input" in control theory sense) produced by the driver, both in longitudinal and lateral directions will be measured. The scope is

to better discriminate between different motion alternatives. The representation of driver input will be given in abstract, vehicle independent way; preferably in terms of the longitudinal and lateral jerk (Nakazawa Ishihara Inooka 2003).

3. A **model for driver perception**, which is an ambitious additional function of the system. The scope of this module is to maintain a representation of the items the driver is aware of, which do not necessarily coincide with the real world. For this module, the feasibility to the expected accuracy is not sure (see risks section), but it does not cost much (here) and if it works it may provide additional very useful information (e.g., understanding that a mistake comes form a missing entity in driver world). The function combines eye gaze observations with information from the perception layer to determine which objects and points the driver watched. The gazed features are introduced into an alternative representation of the world (the driver mental model of the world) and then evolve according to rules that plausibly reproduce the assumptions a driver do about objects he is no longer looking at.
4. An **interaction manager** in the form of a variable plug-in. It will show how a variety of interactions, suited for different types of vehicles and different types of support, can all be built above the same unified situation assessment produced by the core application.

References

Adomavicius G., Tuzhilin A. (2005). *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions*, IEEE Transactions on Knowledge and Data Engineering 17 (6) 2005, 734-749

Andreone, L. Amditis A., Deregibus E., Damiani S., Morreale D., and Bellotti F., (2005). “Beyond Context-Awareness: Driver-Vehicle-Environment Adaptively. From the COMUNICAR Project to the AIDE Concept.” *Proc. IFAC 16th World Congress*, 2005

Berghout, L., Versteegt, E., van Arem, B. (2003). *Advanced driver assistance systems; Results of the State of the Art of ADASE-II*, available on www.adase2.net (consulted on 2sd September 2010).

Bertollini GP, Hogan RM (1999). *Applying Driving Simulation to Quantify Steering Effort Preference as a Function of Vehicle Speed*. Society of Automobile Engineers, 1999-01-0394

Cacciabue P.C. (1998). *Modelling and Simulation of Human Behaviour in System Control*; Springer-Verlag, London, UK.

Catino M. (2002) *Da Chernobyl a Linate. Incidenti tecnologici o errori organizzativi?* Carocci editore, Roma

Console L., Torre I., Lombardi I., Gioria S. and Surano V. (2003). *Personalized and adaptive services on board a car – an application for tourist information*, Journal of Intelligent Information Systems 21(2) (2003), 249–284.

Cook R.I., O'Connor M.F. (2005). “Thinking about accidents and systems”. In: Manasse H.R., Thompson K.K., eds. *Medication safety: A guide for healthcare facilities*. Bethesda, MD: American Society of Health Systems Pharmacists; 2005: p. 73-88.

Dingus, T.A., Antin, J.F., Hulse, M.C., and Wierwille, W.W. (1989). *Attentional Demand Requirements of an Automobile Moving-Map Navigation System*, Transportation Research , 23A (4), 301-315.

Ehmanns, D., Spannheimer, H. (2004). *Roadmap*, available on www.adase2. Net

European Communities (2009). *EU energy and transport in figures, Statistical Pocketbook*, Belgium It can be accessed through the Europa server (http://europa.eu).

Flemisch F., Adams C., Conway S., Goodrich K., Palmer M., and Schutte P., (2003). *The HMetaphor as Guideline for Vehicle Automation and Interaction*, (NASA/TM—2003-212672). Hampton, VA: NASA Langley Research Center.

Green, P. (2000). “Crashes Induced by Driver Information Systems and What Can be Done to Reduce Them” (SAE Paper 2000-01-C008). In *Convergence 2000* (pp. 26–36). SAE Publication

Griffiths, P., Gillespie, R. (2004). Shared Control between Human and Machine: Haptic Display of Automation during Manual Control of Vehicle Heading, *12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (pp. 358-366). Chicago, IL, USA 2004.

Gulian, E., Matthews, G., Glendon, A. I., Davies, D. R., & Debney, L. M. (1989a). *Dimensions of driver stress*, Ergonomics, 32, 585-602.

Hill W., Stead L., Rosenstein M., and Furnas G. (1995). “Recommending and Evaluating Choices in a Virtual Community of Use,” *Proc. Conf. Human Factors in Computing Systems*, 1995.

Hoch S., Schweigert M., Althoff F., Rigoll G. (2007). “The BMW SURF Project: A Contribution to the Research on Cognitive Vehicles,” *in Proc. 2007 IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, June 3-15, 2007.

Ji, Q., Yang, X. (2002). *Real-time eye, gaze, and face pose tracking for monitoring driver vigilance*. Real-Time Imaging (8), 1077–2014

Langley, P. (1999). "User Modeling in Adaptive Interfaces". In *Proceedings of the Workshop on Machine Learning in User Modeling, Advanced Course on Artificial Intelligence* (ACAI ’99), Chania, Greece, 1999 (http://www.iit.demokritos.gr/skel/eetn/acai99/Workshops.htm).

Lolli F., Marzani S., Minin L., Montanari R., Calefato C. (2009). “Not-invasive add-on functionalities for the realtime assessment of driver-vehicle interaction”, in *Proceedings of VI Conference of the Italian Chapter of AIS*, Achieving Fusion in the Interconnected World: Exploring the connection between organizations and technology, Costa Smeralda (Italy), October 2-3, 2009

Mattes, S. (2003). The lane change task as a tool for driver distraction evaluation. In *H.* Strasser, H. Rausch & H. Bubb (Eds.), *Quality of work and products in* *enterprises of the future* (pp. 57-60). Stuttgart: Ergonomia Verlag.

McDonald, W. A., & Hoffman, E. R. (1980). *Review of relationships between steering wheel reversal rate and driving task demand*. Human Factors 22, 733-739.

McLean, J.R. & Hoffman, E.R. (1975). *Steering wheel reversals as a measure of driver performance and steering task difficulty*. Human Factors 17,248-256.

Minin L., Marzani S., Tesauri F., Montanari R., Calefato C. (2009). "Context-dependent force-feedback steering wheel to enhance drivers' on-road performances", in *Proceedings of HCI International 2009*, 9-24 July 09, San Diego, CA, USA

Nakazawa S., Ishihara T., Inooka H. (2003). *Real-time algorithms for estimating jerk signals from noisy acceleration data*, Int.l J. of Applied Electromagnetics and Mechanics, 18, 1-3, pp. 149-163, 2003.

Neale V. L., Dingus T. A., Klauer S. G., Suweeks J., Goodman M. (2005). *An overview of the 100-car naturalistic study and findings*, Paper Number 05-0400, National Highway Traffic Safety Administrationin,Washington D.C.

Neisser U. (1967). *Cognitive Psychology*. Appleton-Century-Crofts, New York

Perry S.J., Wears R.L., Cook R.I. (2005). *The role of automation in complex system failures*. J Patient Safety 2005; 1: 56-61.

Regan M. A, Lee J. D., Young K. L. (2009). *Driver Distraction. Theory, Effects, and Mitigation*, CRC Press, Taylor & Francis Group, New York.

Resnick P., Iakovou N., Sushak M., Bergstrom P., and Riedl J. (1994). "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *Proc. 1994 Computer Supported Cooperative Work Conf.*, 1994.

Shardanand U. and Maes P. (1995). "Social Information Filtering: Algorithms for Automating 'Word of Mouth'," *Proc. Conf. Human Factors in Computing Systems*, 1995.

Tan, P.-N. (2005). *Introduction to Data Mining*. Pearson Addison Wesley, Boston.

Tango F., Botta M. (2009). *Evaluation of Distraction in a Driver-Vehicle-Environment Framework: An Application of Different Data-Mining Techniques*, Lecture Notes in Computer Science, 2009, Volume 5633, Advances in Data Mining. Applications and Theoretical Aspects, Pages 176-190

Tango F., Calefato C., Minin L., Canovi L. (2009). "Moving attention from the road: a new methodology for the driver distraction evaluation using machine learning approaches", *Proceedings of the 2nd conference on Human System Interactions*, May 21-23, 2009, Catania, Italy

Steele, M., & Gillespie, B. (2001) "Shared control between human and machine: Using a haptic steering wheel to aid in land vehicle guidance", *Proceedings of the Human Factors and Ergonomics Society* 45th annual meeting (pp. 1671– 1675). Minneapolis, Minnesota, USA 2001.

Tango F., Minin L., Tesauri F., Montanari R (2010). *Field tests and machine learning approaches for refining algorithms and, correlations of driver's model parameters*, Applied Ergonomics Volume 41, Issue 2, March 2010, Pages 211-224

Verwey, W.B. (1991). *Towards Guidelines for in-car Information Management: Driver Workload in Specific Situations*. Technical Report IZF 1991 C-13, Soesterberg, The Netherlands: TNO Institute of Perception.

Young, K. & Regan, M. (2007). "Driver distraction: A review of the literature". In: Faulks I.J., Regan M., Stevenson M., Brown J., Porter A. & Irwin J.D. (Eds.). *Distracted driving*. Sydney, NSW: Australasian College of Road Safety.

Human-centred design in aviation

Antonio Chialastri (anto.chialastri@tiscali.it)
Aviationlab, Rome

Abstract

This paper focuses on the next challenges that, in the near future, ergonomics has to cope with in the aviation domain. After a short excursus, showing the accidents dynamics along the years and pointing out the relative causes, the paper illustrates the difference between two different conception of automation: a generic human (user) friendly versus a specific pilot-friendly concept. This is useful to evaluate the impact on operational life of the introduction of new technologies onboard in the next generation of airplanes. Some case-studies are shown to give an example of the hidden threats, invisible at the design stage, disseminated through the entire innovation process.

Introduction

Since the beginning of flight, Human Factor specialists have striven to improve the environment in which pilots work. Initially, the upgrading of this environment was made following the accidents' investigation. Air safety was then conceived in a reactive mode; ameliorations and improvements were implemented in the entire system only after a severe mishap and were aimed at avoiding similar accidents.

Safety is conceived today in another way, called "proactive approach". This approach aims at avoiding future accidents, preventing mishap with timely interventions on the areas where possible threats lie, even if no accident occurs. The detection of weak signals helps to understand the menaces' nature, to conceive a set of countermeasures in order to achieve a safer system.

Preliminarily, it is essential to point out which is the safety paradigm that includes our point of view. In fact, during the last seventy years, the safety paradigm changed several times and also the actions taken to achieve risk-free systems, even if a zero accident system has never been experienced. Some conceptions will be briefly discussed as the linear conception, the systemic one and the complex ones (normal accident theory, HRO, resilience engineering).

After having set the frame to our discussion, it will be then described the accidents' dynamics in the aviation domain, to show how the accidents' causes shifted along the years and eventually we describe the macro-area which, according to this paper, represents the next challenge for air safety: ergonomics.

Some case studies will be shown to describe accidents really happened, in order to demonstrate the connection between theory and practice in aviation.

Safety paradigms

"If you have a hammer in your hand, every problem will look like a nail". This is assumed to be a Japanese say and it fits well to describe the situation faced by the investigators: in fact, the spectacles that the investigators don, when they analyze an accident, let them see some items, identified as causes, while neglecting others. During the '30's, according to a "way of thinking" influenced by the Neo-positivistic approach orbiting around the "Wien circle", several disciplines adopted a similar approach to investigate their domain. To synthesize the basic assumptions of that period, every theory should ground its thesis on empirical observation, on measurements, using a language that aims to be universal. During the same period, the industrial domain adopted the scientific management, fostered by Frederick Taylor, based on measurement and optimization of the workers' performance. Psychology, as well, saw the dominance of behaviorism, in which the psyche's inner dynamics (called the black box) were disregarded to focus on observable e measurable acts displayed by the behavior. Safety discipline, too, was influenced and the main tool to explain an accident was the "error's chain", developed by Heinrich, to explain how a single event, originated far away, propagates to affect every other system's component as in a "domino effect".

This metaphor hold on until it was replaced by more functional theory, based on different paradigms. In fact, from the '60's on, the linear explanation was subject to harsh criticisms. In philosophy of science, philosopher as Hans Kuhn proposed a different way to explain the scientific revolution as a paradigm shift, based on collective enterprise either in proposing or in accepting new theories. Moreover, the studies of Von Bertalannfy gave a new impulse on the systemic approach that influenced a lot of disciplines, especially in the biological domains. The stress on the collective thinking fostered a series of new approaches, spanning from industrial domain where a new way of management (team work, total quality) emerged. Even the safety science evolved, shifting from an attitude where the single operator bore the blame for the accident (usually the front end operator, the nearer to the final event leading to the mishap), to a more general approach looking at the different stages of the organizations where hidden traps lie, waiting for a trigger to produce the conditions leading to an accident. This is the theory fostered by James Reason, the Swiss Cheese Model, where safety is seen as a result of different stages acting serially to assure freedom from risks. Every organizational level is seen a barrier fit to intercept any dynamics potentially hazardous for the entire system. Since every barrier has a human component inside, it is prone to errors. This structural condition represents a hole (or set of holes) in the barrier, as in a Swiss cheese. From the initial development of the accident dynamics, the error path passes through all these barriers, eventually causing the accident. This is a more general approach, compared to the preceding one ("name and blame approach", focused on people to charge them legally and morally) attributing liabilities at a much higher level, from the political level, to regulators, to the top management, to middle management and then front end operators.

Nevertheless, this paradigm is still systemic but not yet complex.

Complexity is a new paradigm, emerged from late '80's on, following a bare necessity felt by biological sciences (genetics, biology, medicine) where a reductionist approach was insufficient to put under scrutiny thoroughly the domain. One of the main philosophers that convincingly has proposed a new approach based on complexity is Edgar Morin. On his conception, complexity is difficult also to define, but, as a general way of thinking, it has some common characteristics. It refuses the reductionist and engineering approaches, based on an over-simplification of the reality. The level of observation at which we decide to stay, influences our point of view and determines also our tools to investigate the reality and has its own laws, not necessarily applicable at different levels.

Some scientific disciplines are almost "forced" to adopt such an approach, as genetics, but also in the field of management new theories are emerging to improve performances and comprehension of the organizations.

The safety science followed with different theories in competition to explain the dynamics in complex organizations. To comply with the paper' length requested we cite just three approaches: the normal accident theory (proposed by Charles Perrow), the High Reliable Organizations (studied mainly by James Woods) and the Resilience Engineering approach (Erik Hollnagel is one of the most appreciated authors in this field).

Perrow holds that "zero accident" is not achievable, because of the inner nature of complex system. Too many elements in interaction, give way to unpredictable (and sometimes, unmanageable) situations. Since some domains are not completely under control, such as nuclear plants, they should be closed because the damage arising from an accident is by many times higher than benefits we could gain from their use.

Conversely to what is thought to be a pessimistic approach (or just realistic?) the High Reliable Organizations are some empirical examples of how the man made organizations could be substantially risk-free. They are based on professionalism, on a continuous feedback from the operational levels that is capitalize from top and middle managements. Experience is highly considered as the communication between peers to exchange points of view and to share knowledge. Awareness of an accident is so high that everyone is sincerely committed to safety. Woods studied some organizations revealing that the "safe mentality" is pivotal in assuring a low (if none) rate of accidents. On the contrary to the common say: "No new is good news" these organizations rely on the assumption that "No news is bad news" and when no weak signals of pathogen elements present in the whole system are detected, the management strive to (and push the operational levels) to scrutinize in a deeper way.

Last but not least, we mention the resilience engineering approach. It conceives a safe system as the one who can cope with unexpected events. It has to adapt itself in a flexible and still robust way to respond reliably to the challenge given by a complex system.

Man, in this conception, is not the flaw in the system, but is the main resource to assure flexibility, acting as an intelligent part of the system.

The safety conception assumed in this paper is grounded on the resilience engineering point of view.

In fact, aviation is a complex system in which men, equipments and environment interact. Every of these element is complex in itself.

How should we approach the safety system in aviation, then?

A brief history of accidents

(Graphic's explanation: decades on the x-axis, accidents per million take-offs on the y-axis. Source: Flight Safety Foundation)

Most of the corrections to existing systems or procedures, in aviation, were introduced following severe mishaps. So the path of the entire industry has been a kind of "trial and response" dynamics: innovation, mishaps, correction. According to the statistics, the human error has played a pivotal role in the accidents, with a higher rate, compared with other factors as environment (meteorological conditions, Air traffic that induces mid-air collision, and so on), mechanics (i.e.: structural limit exceeded, poor cockpit design) security (high-jacking, bomb onboard, etc.).

Starting from the '40's, investigators wondered why airplanes crash. Taken for granted that the pilots were the fallible factor in the entire system, someone started to analyze "why" pilots did so many errors. At the beginning, till the mid '50's, the main cause of accident was identified as "Loss of control". This category includes situations in which pilots lost the airplane control such as: reaching (and exceeding) the structural limits, conditions in which the airplane stalls, overbanks or experiences an unusual attitude that put in jeopardize the flight progress. The root cause of lost of control spanned from fatigue, to distraction, to excessive workload, to sleepiness, and so on. Briefly, the problem was identified in the main area of "human performances and limitations".

The solution thought to fix this kind of problems was the engineering, to provide more systems, more aids and more technology.

The technological approach focused on two sides: innovation of ground-based aids and implementation of new instruments onboard.

On the first side, two main innovations were provided:

- the air traffic controllers were equipped with radars to monitor the airplanes approaching the airports and;
- the installation of ground based equipments such as ILS (Instrumental Landing System) gave a strong help to pilots in order to land as precisely as possible.

On the other side, namely the introduction of new technologies onboard of the airplanes, the introduction of auto-pilot, auto-throttle, flight director, helped to:

- lower the workload, when too much attention was needed to carry on the task, or;
- relief the pilot from monitor boring activities, reducing duties related to monotonous operations.

The effects of these innovations were successful, since the rate of accident sharply dropped. Nevertheless, during the '70's the accident rate started to rise again, but with a different dynamics. In fact, the main cause of accident shifted from "Loss of Control" to CFIT (Controlled Flight Into Terrain). In this kind of dynamics, a perfectly efficient airplane hit an obstacle in the nearby of the airport when full in control of the crew. Furthermore, we have to consider that most of the accidents happen during the approach phase. The investigations revealed that a poor decision making, a loss in the situational awareness, a conflict (open or concealed) was in progress between the pilots. In short, there was a problem in the human interaction onboard.

This time the solution didn't pass through technology, but applying a new approach, based on psychological assumptions on what is thought to be a good team work. We should mention that, on that period, other new technologies were introduced in the aviation system, but it is generally assumed that the psychological approach was pivotal in improving the system's safety. Courses of CRM (Crew Resource Management) were implemented in most of the main airlines to enhance the interaction between the pilots (and, later, also between the entire crew, cabin attendants included).

The accident curve dropped again, but during the '90's it raised again, even if in a smaller magnitude compared with the past decades. The problem is that the overall dimension of the air transport, nowadays, has inflated in the last decades and even a small amount of accidents (lower than in any other transportation domain such as roads, railway, sea, etc.) could be unbearable for some reasons. Firstly, the human, legal and economic cost of an accident is huge and could destroy an airline's stability, leading it out of the economic contest. Secondly, an air accident has a worldwide resonance and could distort the real perception of air safety in the public opinion.

Whatever the consequences of air mishaps, it is essential to understand why they keep on happening. During the '90's, the main cause of accident shifted once again, as a pendulum, swinging back to "Loss of control", but in a different shape, compared to the one experienced during the '50's. In fact, today the pilots have so many technological aids that is hard to conceive how they can lose the control of the airplane. Actually, the implementation of so many systems is the consequence of the engineering approach to safety in which the pilots are seen as the weak ring in the industrial chain. So, automatisms are intended to substitute many functions played usually by pilots.

There is a widespread opinion among authors studying human factor in aviation that in this case we may talk about "over-redundancy": too many instruments induce a low workload that could provoke complacency, inadequate training make the pilots unable to override the automatisms in case of their failure or misbehavior.

Case studies

Here are briefly presented two case studies illustrating the relationship between pilots and technology: one related to the misuse of instruments by pilots induced by a poor designed system and the unpredictability of a system behavior when in the real operational context.

The first case involved an Airbus A-321 operated by Air Inter who crashed in Strasbourg after the captain misunderstood the descent profile usability because of the similarity between the flight path angle function and the vertical speed function. In fact, both were displayed via a two digits figure in the same feed back window. For instance, 3.3 could represent either a vertical speed of 3300 feet per minute or 3.3 degrees of vertical path. The captain selected 3.3 being sure to descent with a vertical path selected, while he was descending with 3300 feet per minute, a much steeper path than the desired one. The approach was conducted among high terrain around the airport and such an error gave the crew no way out to recover timely. After that disaster, the display onboard was changed and now there is no way to misunderstand similar functions during the approach phase. Furthermore, after the accident the French authority requested, as mandatory, the installation of the GPWS (Ground Proximity Warning System), which warns the crew in case of excessive approach rate to the ground. It is designed to avoid unintentional collision with obstacle, when not in landing configuration. Today this apparatus has been improved, becoming EGPWS, which is linked to the satellite indication. This allows the system to realize if the low altitude is consistent with the airport location and with obstacles scattered in its vicinity. All the relevant information are displayed to the pilots, who immediately could be aware of the presence of mountainous terrain close to the aircraft position.

The second case involved an A-300 approaching Miami. Due to bad weather around the airport the crew expected to enter an area of turbulence. The crew was instructed to hold over a radio-facility. During the descent, with engine at idle thrust, the auto-throttle (managing the engine thrust, via an automatic movement of the throttle governing the necessary thrust) disengaged with no evident signal displayed to the crew. In the proximity of the holding pattern, the airplane leveled-off, reducing its speed well below the minimum required to sustain the flight. During the initial turn in the holding pattern, the airplane stalled, down-spiraling and losing about three thousand feet. This is a very serious condition for a wide-body aircraft. While spiraling downward, the crew lost all the attitude indications for few seconds, that looked (according to the captain, interviewed after the incident) an eternity. In fact, the only useful instruments in such a situation are the attitude and the speed indicator. The attitude indicator was, by design, conceived to go blank in case of oscillations exceeding some amplitude and frequency. This assumption, made at the design phase, comes from the idea that such oscillations are very unlikely in the airline flight. Reality, alas, is much more unpredictable than the engineer's fantasy.

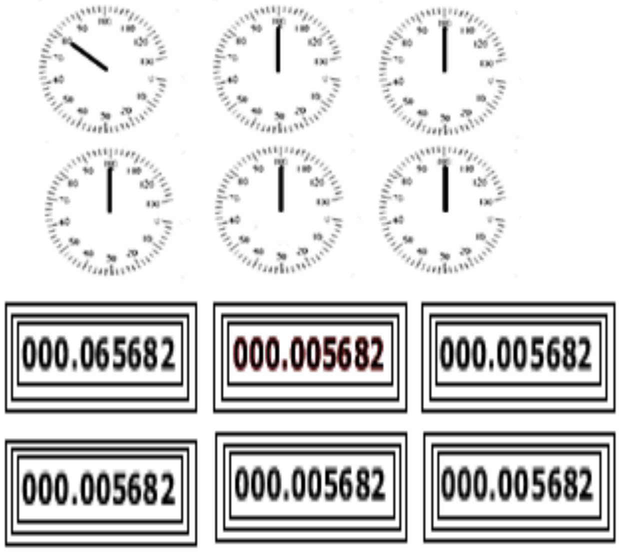
Human factor and technology

There are different conceptions of Ergonomics, as emerges from the evolution of the discipline along the years. Initially, ergonomics was conceived as corrective ergonomics: expert tried to understand how to make system better, after the misuse of something badly designed.

Here it is an example: the design of an airplane with variable wings. In the engineer's mind, it was quite simple to conceive an airplane with variable wings, setting them from straight wings to swept wings. Actually, the straight wings are used at low speed, whilst the swept wings are useful at high speed. To one person observing an airplane is intuitive to understand how to imagine the command lever to change the wings configuration: putting the lever forward, you get straight wings, if you put the lever backward, you get swept wings. It looked quite simple, but some accidents happened cause by pilots' misuse of the command lever. In fact, for a pilot's point of view, every action linked to the idea of speed leads him to move forward: increasing the thrust? Throttle forward. Increasing the speed in case of sudden loss? Pitch down, putting the yoke forward. So when the new system was implemented, a lot of pilots misused it, following their mental pattern related to the speed.

Nowadays, human factor experts are involved at early stage in the design process, to keep the system user friendly. Actually, what is required is the expertise of someone who can translate an engineering necessity in an operational suitable system. Let's think about the number display onboard.

According to the Gestalt principles, human mind is more concerned about general configuration rather than in analytical vision. This is more than true inside a cockpit, because the number of the displays, the short time available to detect every single variation, the process of interpretation of multiple data. In a pilot's mind, symmetry is more important than a precise indication. Here it is an example:



Given the same figures, it is obviously easier to spot a difference on the left side display, called "field vision", versus the "analytical vision" on the right side.

The same applies to the speed indicator, such a speed tape, set on the left of the modern attitude indicator (PFD: Primary Flight Indicator). They have the great advantage, compared to the older version (analogue indicator) of speed indicator: it can represents also the speed related to the entire operational envelope, such as flaps and slats operating limitations, over speed, approach to stall warning et cetera. The problem, as a philosophy of flight is that things appear to go better when the workload is low (inducing perhaps complacency) while they go worst when there is a main failure. In fact, all those useful indications are removed from the speed tape, leaving the pilot to strive with a higher mental workload.

Conclusion

In this short introduction to the problems arising from the implementation of new technology in a modern cockpit, this paper tried to point out the difference between the user friendly concept, as imagined by the airplane designer, and the pilot friendly concept, that follows a mental pattern given by experience and knowledge of the sharp end operators. To obtain a higher level of safety, everyone should strive to make it resilient. The history of airplanes' accidents shows quite clearly that new solutions bring new problems. In this phase we may say that an excessive use of technology could make the entire system less resilient. In fact, the pilots are used to have knowledge of the airplane they fly, based on a kind of "over-learning". This ample knowledge gives the pilot some flexibility, allowing the user to utilize the machine in a non standard way, whenever necessary. At the time in which new generation of airplanes (Fly-by-wire, dark panel, Flight Management System) were conceived, the pilot has been set at the edge of the innovation process. That induced some kind of accidents due to poor interaction and basically to a misunderstanding of the system inner logic.

Paradoxically, to many instruments, thought to be a substitute for humans, could bring two main problem, from a pilot's point of view. Firstly, they induce a low workload when things are running normally and this low workload could induce complacency on the system's reliability. Over-reliance is at the core of some accidents, when pilots could not regain the full control of the aircraft after the automatisms failed.

On the other side, when pilots are in emergency they need more help. Conversely, much of the aids normally available to pilots are removed during an emergency situation. We may, in short, say that the paradox of automation onboard could be said as: "When good, better; when bad, worse".

In my experience, I see that to enhance safety via an engineering approach, it is necessary to take into consideration the pilot's point of view, to implement new systems at the same time useful and usable. But, before introducing new technologies, we should first set the frame to make clear which is our safety paradigm and which is the intended outcome.

The expertise given by the final user is, in this context, highly valuable, since it represents the necessary connection between aims and tools.

References

Cooper, G.E., White, M.D., & Lauber, J.K. (Eds.) (1980) "Resource management on the flightdeck," Proceedings of a NASA/Industry Workshop (NASA CP-2120)

Dekker, S., Johan Rignér, (2000) "Sharing the Burden of Flight Deck Automation Training", The International Journal Of Aviation Psychology, 10(4), 317–326 Lawrence Erlbaum Associates, Inc.

Dekker, S. (2001) "Reconstructing human contributions to accidents" Technical Report - 01, Lund University School of Aviation

Dismukes, Berman, Loukopoulos (2008), The limits of expertise, Ashgate, Aldershot, Hampshire

Goeters k.M. (2004), Aviation Psychology: Practice and Research, Ashgate Publishing Ltd. Aldeshot Hampshire.

Hollnagel E., Woods D., Leveson N. (2006) (a cura di), Resilience Engineering – Concepts and Precepts, Ashgate, Aldershot Hampshire

Hollnagel Erik, (2008) "Critical Information Infrastructures : should models represent structures or functions ?", in Computer Safety, Reliability and Security, Springer, Heidelberg

Hollnagel Erik, (2009), The ETTO Principle – Efficiency-Thoroughness Trade-Off, Ashgate, Surrey, England

ICAO, (1998), Human Factors Training Manual, Doc 9683-AN/950.

IATA (1994), Aircraft Automation Report, Safety Advisory Sub-Committee and Maintenance Advisory Sub-committee

Norman Donald (2005), La caffettiera del masochista, Giunti Editore, Firenze

Ralli Marcello (1993), Fattore umano e operazioni di volo, Libreria dell'Orologio, Roma

Reason James, (1990) Human error, Cambridge University Press, Cambridge

Tichauer E. R., (1978), The Biomechanical Basis of Ergonomics: Anatomy Applied to the Design of Work Stations, New York: John Wiley & Sons,

Resilience as Individual Adaptation: Preliminary Analysis of a Hospital Dispensary

Mark-Alexander Sujan (m-a.sujan@warwick.ac.uk)
Warwick Medical School
University of Warwick, Coventry CV4 7AL, UK

Simone Pozzi (simone.pozzi@dblue.it), Carlo Valbonesi (carlo.valbonesi@dblue.it)
Deep Blue Human Factors Research & Consulting
Rome, Italy

Catherine Ingram (Catherine.Ingram@hhtr.nhs.uk)
Hereford Hospitals NHS Trust
Hereford, UK

Abstract

This study aimed to identify and to describe common forms of resilient behavior in a hospital dispensary. 16 narratives submitted by pharmacy staff were analyzed qualitatively. Common forms of resilient behavior identified include: personal negotiation, creating shared awareness, thinking ahead, seeking help, prioritizing activities. Most of these forms of behavior rely on personal initiative and experience and there is little organizational awareness and support.

Introduction

Hollnagel (Hollnagel, 2006a) argues that "to understand how failure sometimes happens, one must first understand how success is obtained". Such a statement well represents the line of thought known as *Resilience Engineering*, advocated by Hollnagel himself, Nancy Levenson, David Woods and Sidney Dekker among others (Hollnagel, Woods, & Leveson, 2006). These authors maintain that safety is better managed by also focusing on what the system does well, rather than simply concentrating on the failures. Even though resilience has been defined in many different ways, depending on the main emphasis the authors want to convey, a commonly agreed definition may read as follows: "the intrinsic ability of an organisation (system) to maintain or regain a dynamically stable state, which allows it to continue operations after a major mishap and/or in the presence of a continuous stress" (Hollnagel, 2006b, p. 16).

Resilience engineering thus emphasises the ability of a system to maintain control even when faced with (major) disruptive events. It also specifies how such an ability should be able to cope with both internal and external events, namely with its internal variability (e.g. technical failures, human action, etc.) and stress engendered by external variability (e.g. weather conditions, problems of nearby systems, etc.). Some authors (Woods, 2006) also highlight how a resilient system should to able to adapt not only to known disturbances, but also to problems outside of the "design envelope", that is to problems that were not anticipated by system designers and happen in a totally unexpected manner (or timing).

Compared to more traditional approaches to safety, resilience engineering stresses the need to analyse in an integrated way what makes a system work as well as what causes it to fail. Success and failure stem from the same processes, or system characteristics, hence they should be understood as generated by the same system properties.

Hollnagel particularly criticises all those approaches that try to curb or constrain human variability as a source of disturbances. According to Hollnagel, people and organisations always need to adjust flexibly to the operating conditions, in order to make optimal use of finite resources and time. Human variability is the core driver of such a flexible adjustment, so it should not be eliminated, but rather seen as extremely useful. It may also engender failure, but most of the times it ensures that the system adjusts successfully to internal or external disturbances and keeps operating at a satisfying level of performance.

In healthcare, where Reason's (Reason, 1997) model of organizational accidents has been highly influential in shaping many patient safety initiatives, it is rather uncommon to look at how people successfully cope with disturbances and disruptions. A literature search in *Quality & Safety in Health Care*, one of the leading journals dealing with patient safety, produced few hits on the topic of resilience. There is, therefore, a need for empirical studies that explore and describe resilience in healthcare settings.

Cook et al. (Cook, Render, & Woods, 2000) introduce the useful notion of *gaps* or discontinuities in care. Due to the structural characteristics of healthcare and the intrinsic complexity, a major activity of healthcare workers is to cope with the resulting gaps and discontinuities in care. In other words, normal successful everyday performance is not the result of prudent system and safety barrier design only, but rather of the technical work of people within the system who anticipate, detect and bridge the various gaps they encounter. It is important to bear in mind the emphasis here on everyday performance, rather than on failure. These activities intended to deal with complexity and bridging the gaps are so tightly interwoven with other technical work that often they cannot be distinguished from it (Cook et al., 2000).

In a previous paper (Pasquini, Pozzi, Save, & Sujan, 2010), we elaborated a model where risk factors were pushing the system out of control, counter-acted by resilient behaviours (or other resilient features). The model was based on the authors' experience of how a real safety critical system works and achieves its functions, but lacked a detailed description of the resilient characteristics.

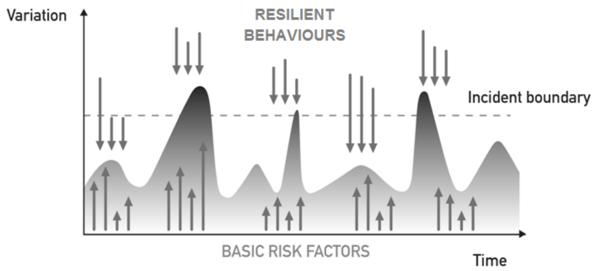


Figure 1. The variability of a system needs to be maintained under control, by counterbalancing disturbance factors (upward arrows) with resilient behaviours (downward arrows).

In this paper we present preliminary results of a qualitative analysis of resilient forms of behavior in a hospital dispensary. The next section describes the setting, as well as the data collection and analysis methods used. Then the results of the analysis are presented with empirical examples. The concluding section outlines further work.

Methods

Setting

The hospital is a main provider of acute services for the West of England and parts of Wales and has a capacity of 259 inpatient beds. The pharmacy department employs 50 staff, the majority of which work in the dispensary on a rotational basis, and there are 8 staff who are based permanently in the dispensary.

Data collection

Data was collected as part of the Health Foundation Safer Clinical Systems Program (SCS). The program was commissioned in 2008 and involves 4 NHS organizations with the aim of developing systems approaches to delivering more reliable and safer care. The data used in this paper was collected by inviting staff in the dispensary in one hospital to submit narratives about something that caused them hassle during the previous week. Staff were encouraged to use their own language and style in order to promote the idea of the narratives being a kind of reflective "safety diary". No further guidance or restrictions were provided and the submitted narratives varied in length from one paragraph to 5 pages. In the first instance, 16 narratives were submitted (by 13 out of 34 members of staff that had been approached).

Analysis

The submitted narratives were analyzed qualitatively using the Nvivo software package. The software tool facilitates deep analysis of non-numerical or unstructured data, such as narratives and interviews. It supports a range of qualitative research methods including grounded theory, the approach taken in this project. The preliminary analysis and coding were done collaboratively by a domain expert (CI) and a safety expert (MAS). The codes were generated from the data. Emerging themes were discussed in a review meeting with human factors experts (SP, CV).

The last part of the analysis was conducted separately by two analysts (MAS, CV) on the basis of an emerging research hypothesis. The two analysts later compared the results of their work with a third HF expert (SP), (i) to agree on the coding of specific events and (ii) to establish a shared set of codes (which will be used to continue the analysis in following research studies).

The purpose of the above process was to prevent idiosyncratic interpretations by a single analyst by involving three partially separate strands of work. To counterbalance potential divergence between the three analysts, the initial research hypothesis served as a guide to orient the separate strands in a common direction. A third requirement was not to spoil the richness of field data by imposing an overly strict *a priori* interpretation – for instance, ideas deriving from previous researches or from the theoretical framework.

The theoretical resilience engineering framework was brought into play only as a common theoretical understanding of the data, but did not orient the identification of the codes themselves. As stated above, the codes were generated by the data, with an empirical bottom-up approach. A detailed comparison of the results with "standard" resilience engineering frameworks will be performed in follow-up studies.

Results

The preliminary analysis identified three main themes, under which most of the codes could be clustered: disturbances, feelings of frustration, coping strategies. These are explained below.

Disturbances

Situations that were described as disturbed or challenging in the narratives were coded as *disturbance*. This node arose after merging initially used nodes such as *concurrent activities*, *excessive demand* and *absences* into a single higher-level node. A disturbance can range from mild or frequently recurring disturbances to a crisis situation. Disturbances can be caused by internal (absence due to annual leave, people engaged in multiple activities) or external (absence due to sickness, external demand) factors. The immediate consequence of disturbances is a rise in demand and queues building up. Disturbances require adaptation and coping strategies. Depending on the success of these coping strategies the consequences of disturbances may be negligible (successful adaptation), or cause delays and lead to frustration. An example is provided below:

"*We were short staffed due to sickness and annual leave, the phones never stopped [...] a technician rang from the ward he wanted a couple of green profiles podding down to [their ward] - a 2 second job, well so I thought. When I got there the draw was bare, someone had obviously used the last one without photocopying anymore. So, I had to leave an already short staffed dispensary to go to the copier to copy some green profiles which takes a little longer because you have to swap the paper in the machine etc.*" (Example 1: disturbance)

In this narrative, the subject points at internal disturbances (e.g. annual leaves, a phone call, the draw being empty) and to external ones (e.g. sick leaves), showing their impact on an already stressed system "*we were short staffed […] I had to leave an already short staffed dispensary*".

Feelings of frustration

Feelings of frustration express the personal emotional reaction to working situations and fall under the larger category of consequences of disturbances. However, in order to emphasise the emotive nature of frustration, it was decided to keep it as a separate category.

Frustration can vary in its intensity ranging from a feeling that one isn't getting anywhere to being very annoyed and upset with oneself or colleagues. Frustration expresses (is caused by) a dissatisfaction with the performance or behaviour of others, of the organisation or with oneself. The consequences of frustration are not clear from the narratives, but may have a negative impact on the coping strategies outlined below, including unwillingness to communicate with others (ask for help, provide help) and not sharing information. An example from the narratives:

"*Phoned [Location A] with a query about a prescription, which I had to explain the full story to 3 nurses, only to find by the third nurse that the patient had already gone home and the prescription was no longer required – very frustrating!*". (Example 2: feeling frustrated)

In this example, the individual reports the feeling of frustration following an episode of time consuming communication with different individuals at another location that turned out to be needless since the information was no longer required.

Coping strategies

Strategies to cope with disturbances often involve personal negotiation and sharing of information about the current situation in order to create a shared awareness:

"*Lead technician made me aware that the CT scanner had been down and there were 37 patients waiting for an appointment, if the scanner was fixed later today, we may see an impact. This would increase the workload on an already busy day. I told lead technician that I'd chase this up with [the Clinical Director] to find out if there was anything we could do to prepare for this.*" (Example 3: creating shared awareness)

In this case the disturbance is an internal one (e.g. a technical failure like the CT scanner being down) and the subject clearly anticipates a potential problem for the system (e.g. workload is going to increase). S/he then warns the Clinical Director, in order to ensure that required actions are taken before the workload increase actually happens.

"*About 11am [the Senior Dispensary Assistant] came to let me know that the pharmacist had a lot of work in his tray and the dispensing [Dispensary Assistants] were looking for work to do. I noticed that one of the ward-based technicians was dispensing and so asked [the Senior Dispensary Assistant] to see if she could check any patient's own drugs to help the pharmacist.*" (Example 4: personal negotiation)

In this case, people report to fellow colleagues that the workload is unevenly balanced between two roles (i.e. the pharmacist and the dispensing Dispensary Assistants), making sure that a shared awareness exists of the current system status. As a consequence to that, people flexibly re-adjust their roles, by shifting one technician to a support role for the pharmacist.

Other strategies to deal with disturbances include prioritization of activities and seeking help from others / offering help. Coping strategies can be proactive (e.g. freeing up resources in case they will be needed later) or reactive (prioritizing activities). Prioritization can be done based on pre-defined work flows or ad-hoc. Seeking help can differ in terms of the type of help that is required, e.g. extension of one's own capabilities, a different set of skills.

"*In the afternoon a nurse came to the hatch and said that there was water on the floor in the waiting area, I was really busy catching up with my databasing and filing but as there were quite a number of people waiting and in view of patient safety I left my post to go and mop it up. It looked like a cup of water from the water machine had been knocked over.*" (Example 5: ad-hoc prioritisation)

"*People off work sick and then there were pre-arranged meetings to go to. [...] We managed to keep up with our work flow of fast track items and out patients (this hadn't given the clinical check pharmacist a chance to do standard track).*" (Example 6: prioritisation based on pre-defined workflows)

"*About 1.20pm ward based technician came to ask if I'd like some help ACTing as there wasn't much work in the dispensing tray - I said yes please!!*" (Example 7: offering help)

The three examples show three different ways of prioritising work to optimise time and resources. In the first case, optimisation takes the form of interrupting the current activity to carry out a more urgent one (no harm can arise from temporarily interrupting the database work, while someone can slip on the water). In the second case, a staff shortage (caused by sick leaves) is managed by relying on two pre-defined workflows (fast track items and standard track items). The third case is a case of help self-offer, based on the recognition that workload is currently low in the ward, but it may be instead high in the dispensary.

Conclusion

The hospital dispensary is an environment that faces challenges both due to internal as well as external factors. Internal factors include issues such as staff on annual leave, staff being unavailable due their being out on wards or being engaged in multiple activities. External factors relate predominantly to the nature of the work (incoming prescriptions) that is dependent on work flows in other parts of the hospital. The narratives describe an environment that is frequently very busy and stretched, but that is at the same time able to adjust and to adapt to these challenges.

The preliminary qualitative analysis of the narratives identified a number of coping strategies that enable the dispensary to deal with the challenges in a resilient way: personal negotiation, creating a shared awareness, prioritization of activities, offering and seeking help from others. As maintained by Hollnagel, human variability is a key feature to adjust and adapt to current demands, as there are cases of role swapping, dynamic prioritization based on local demands and relative urgency, active monitoring of workload uneven distribution, or anticipation of likely problematic future demands.

It is interesting to note that only prioritization of activities is supported at an organizational level through pre-defined urgent and standard workflows. The vast majority of resilient forms of behavior exhibited by the dispensary are the result of personal initiative, negotiation and experience. No training is provided for such skills and there are no mechanisms in place to capture and to document valuable experiences.

The analysis is a first step towards a more comprehensive, empirically constructed framework of resilience in healthcare environments. Such a framework should allow healthcare organizations to identify training opportunities in non-technical skills as well as to institutionalize resilience.

Acknowledgments

This work has been supported in part by the Health Foundation (Registered Charity Number: 286967) as part of the Safer Clinical Systems program. We would like to thank all the pharmacy staff who contributed narratives.

References

Cook, R.I., Render, M., & Woods, D.D. (2000). Gaps in the continuity of care and progress on patient safety. *British Medical Journal*(320):791-794.

Hollnagel, E. (2006a). *Resilience-The challenge of the unstable*. Aldershot, UK: Ashgate Publishing Limited.

Hollnagel, E. (2006b). Resilience – the Challenge of the Unstable. In E. Hollnagel, D.D. Woods & N. Leveson (Eds.), *Resilience engineering : concepts and precepts* (pp. xii, 397 p.). Aldershot, UK; Burlington, VT: Ashgate Publishing Limited.

Hollnagel, E., Woods, D.D., & Leveson, N. (2006). *Resilience engineering : concepts and precepts*. Aldershot, UK; Burlington, VT: Ashgate Publishing Limited.

Pasquini, A., Pozzi, S., Save, L., & Sujan, M.-A. (2010). Requisites for Successful Incident Reporting in Resilient Organisations. In E. Hollnagel, D. Woods & J. Wreathall (Eds.), *Resilience Engineering in Practice: A Guidebook*. Aldershot, UK: Ashgate.

Reason, J. (1997). *Managing the risks of organizational accidents*. Hampshire, UK: Ashgate Publishing Limited.

Woods, D.D. (2006). Essential Characteristics of Resilience. In E. Hollnagel, D.D. Woods & N. Leveson (Eds.), *Resilience engineering : concepts and precepts* (pp. xii, 397 p.). Aldershot, UK; Burlington, VT: Ashgate Publishing Limited.

A Collaborative Approach for Supporting Interaction among Hearing Impaired and Listeners

Samuel O. Apolonio (samuel.apolonio@gmail.com), Luciane C. J. de Deus (lujasmin@gmail.com), Marcos R. S. Borges (mborges@nce.ufrj.br), Adriana Vivacqua (avivacqua@gmail.com)
Graduate Program in Informatics
Federal University of Rio de Janeiro
Cidade Universitária, Rio de Janeiro, Brazil

Abstract

Hearing impairment was indicated as the second largest proven deficiency in world population. People with such disabilities face many obstacles in their communication and interaction with listeners, even with equal level of cognitive development. Collaborative tools offer the possibility to improve the interaction between listeners and hearing impaired subjects. This study presents an approach to support this type of interaction aimed to augment the participation of hearing impaired in collective activities. The approach has been materialized in a collaborative writing tool on which participants from both categories work together to produce a common document. The approach was evaluated by means of an experiment conducted with listeners and hearing impaired students in a learning environment. Analysis of the results obtained in the experiment show that the collaborative writing enhanced the interaction and contributes to a more harmonious and real coexistence among listeners and hearing impaired.

Introduction

The WHO (World Health Organization) estimates that 10% of the population of developed countries is composed of people with a disability. In developing countries this rate is estimated between 12 and 15%. According to Smith (2003) among the world's population under the age of 15, approximately 62 million have permanent and hearing loss. Olusanya (2005) indicated that two-thirds (41 million people), live in developing countries. The incidence of hearing loss in newborns, according to White (1993) is 1.5 to 5.95 per 1000 births. In the literature, we also found that *presbycusis* - hearing loss due to age - is the main cause of hearing impairment in the elderly, with an incidence of about 30% of the population over 65 years of age. The noise, especially in the workplace, is appointed as the second leading cause of sensorineural hearing loss among adults. Hearing impairment affects five percent of adult population in U.S., according to NCHS - National Center for Health Statistics. In Brazil Hearing impairment stands out as the second largest proven deficiency in Brazilian population.

Hearing impairment is characterized by total or partial loss of ability to hear. It manifests as mild to moderate hearing loss and severe or profound deafness. It is considered one of the main disorders that can interfere with language development and speech. According to the American Speech-Language-Hearing Association it represents 60% of communication disorders. Thus, this difficulty interferes directly in the interaction with hearing individuals.

People with hearing impairment may have affected their learning and their integral development, since language plays an essential role in perceptual organization and the receipt and structuring of information, learning and social interactions of human beings. For Gatto (2007) the hearing is a pre-requisite for development and acquisition of language. Hearing and language functions are interrelated and interdependent.

Despite these data, the hearing impaired subjects are not easily recognized in society as people who have distinct needs. This recognition occurs when they need to communicate, because of the difficulty they present using oral colloquial language. This difficulty directly interferes in the interaction with listeners.

According to Capovilla (1998), technology should aim to solve human problems and the solution should not discriminate any type of person. It should look universally the situations faced by humans, assuring them full participation in the environment they live. Thus, we should also use technology to help those who have disabilities in having a life as normal as possible.

New technologies offer good alternatives to facilitate interaction among individuals communication. In this scenario, there is a powerful mechanism of interaction among individuals: the collaborative writing. This mechanism allows the diversity of knowledge and skills and helps individuals who have difficulty to formalize certain knowledge or build a solution alone. Thus, problems can be better solved by a group of individuals working collaboratively, than by a single individual (Howard, 2000).

The aim of this study is to analyze the contribution that collaborative writing can offer to improve the interaction among hearing impaired and listeners. To support this claim an experiment is proposed by mixing in each group listeners and hearing impaired subjects enrolled in a common task: the production of a document in a collaborative way.

This paper is divided as follows: Section 2 describes the communication aspects of hearing impaired persons. Section 3 presents the collaborative writing process and its features that help collaboration among individuals. Section 4 describes an experiment conducted to evaluate the benefits of technology supported collaboration in overcoming individual deficiencies. Section 5 analyzes the results of this experiment and Section 6 presents the conclusions.

Hearing Impaired Communication

Social networking has had a clear impact in the world, connecting people to find and create new friends, share ideas and organize events. Although social interactions are activities that have always happened naturally, on the Internet this process becomes more open and fluid affecting the way people interact. These processes find their "natural" environment on the Internet, where social collaboration and dissemination of information is facilitated by recent innovations.

According to Adams (2008) “The driving social behaviour now though is collaboration. The networks are established and have become part of the scenery. As people become confident social agents in online networks they begin to act, to organise, to create.”

Collaboration in social cooperation between individuals may have different objectives, both to make life more pleasant and to supply their deficiencies. The computer is seen as pro-cognitive activity of structuring knowledge representations and also on emotional development (Oliveira, 1996). It is a resource for children with learning difficulties despite its shortcomings and limitations develop their cognitive capabilities and possibilities of their own.

Hearing impairment is a kind of sensorial restriction, whose main symptom is an atypical response to the sound stimulus, being classified accordingly to the degree of hearing loss (Marchesi, 1996). However, Ciccone (1996) states that the hearing impaired is an individual with potential normal cognitive, although the hearing loss implies, often, in serious obstacles to its interaction with listeners.

According to Couto-Lenzi (1997), in interaction with hearing impaired, listeners should consider that among the hearing impaired, some are born with hearing loss and others lose their hearing after birth, during the pre-linguistic stage or after learning mother tongue, resulting then, in different prognosis in the learning process and development of writing.

This distinction becomes relevant when interacting with deaf people since it implies a greater or lesser degree of difficulty in the use of colloquial language oralized, practiced by individuals listeners.

In accordance to Northern and Downs (1999), any symbols that emerge in society are conditioned to a language of listeners. In the hearing impaired these symbols lose their meaning and sign language will always be the most prevalent.

This difficulty of interaction may also be reflected in collaborative writing, since the hearing perception is correlated with the acquisition of written language, due to the relationship of sounds to the graphic symbols that characterize the natural language. Thus, the formal writing produced by deaf people is based on a different way of thinking and basically sign languages.

As Gotti (1991) affirms, the phrase structure of the hearing impaired is disjointed, without connecting elements, often without verbs, due the deficiency in logic. This is one of the major challenges of the interaction between deaf and hearing people through writing.

There are studies that link the technology-mediated communication as a facilitator of integration. As reported in the work of Santarosa (2003) with deaf and blind people, that integration is provided communication interfaces to facilitate and support the interaction among these subjects. , with evidence from a blind person that corresponded with a colleague, using technological resources, not knowing that this was deaf. In another study Santarosa (2002) stresses that the use of electronic means, primarily e-mail, allows the advance in development of the deaf, with a view to written communication and social interaction.

Collaborative Writing

The human is in continuous biological development, influenced and influencing the social environment where he lives and exercises their interactions. He has natural biological altruism as an individual and need to be part of human groups and to operate by consensus with them. This need acts as a motivating factor in interactions with other individuals and thus the hearing impaired seeks to overcome the difficulties of interaction. For this, the use of new technologies that facilitate this process is seen as a good alternative.

Many efforts have been made to develop solutions to facilitate this process. Among them, we highlight the development of multimodal interfaces, the use of tools that make the association between text and video and software that seek to provide collaborative practices, the groupware, more precisely the collaborative editors, object of this study.

The basic development concept in multimodal interfaces for the disabled is the idea of modality replacement, which is the use of information originating from various modalities to compensate for the missing input modality of the users (Moustakas, 2006).

The tools that make the association between text and videos can be used to associate texts in the mother tongue videos with the same text into sign language, accessible to the hearing impaired.

Among the groupware, we highlight the cooperative editors that allow the creation of texts in cooperation with two or more users.

Editing documents collaboratively or jointly with others is a common task. Often, the documents we produce are reviewed by someone or receive some kind of contribution (Tammaro and Mosier, 1997). The collaborative writing of documents allows participants to interact during the construction of texts, generating new ideas and modifying them still of development work (Howard, 2000). Thus, participants always have the possibility to suffer interference in textual exposition of his ideas, creating a new text composed by several participants. This context of interaction among participants in a cooperative process of editing promotes acceptance of differences that exist between individuals and could be further explored when we think of interaction between listeners and deaf people.

Besides promoting greater interaction among co-authors, collaborative editing environments can promote critical thinking, helping people to learn from each other and strengthen the social relationships of those who write together (Mailhiot, 1968). Such benefits could also be targeted to facilitate interaction between deaf and hearing. Thereafter, the editing process could become a cooperative way for the development of a sense of cooperation and acceptance of hearing impaired by listeners.

However, it is necessary not only thinking on the benefits or problems that this approach represents, but also about the changes in the perception of reality and the changes that occur even in the way of writing. If writing together with individuals who use the same grammatical structure of language can become a complex process, with individuals who use different grammatical structures the level of complexity may increase further.

In contrast, once aroused the senses to the value of cooperation or interaction and understanding its validity, it will never be forgotten by an individual. Thus, the benefits of this interaction could overcome the difficulties presented.

These characteristics can become great allies in this process of interaction through collaborative writing. According to Ellis, Gibbs and Rein (1991), until they have established interpersonal relationships based on acceptance, interdependence and complementarity, the groups are not ready to develop a cooperative work.

The Experiment

Aiming to examine the contribution that collaborative writing can offer to improve the interaction between deaf and hearing people, was conducted an experiment that involved the production of a text in a cooperative manner between these two types of co-authors.

The experiment was conducted in the computer lab of the School Professor Olga Teixeira de Oliveira, pole of inclusion of students with hearing impairment located in Duque de Caxias in Rio de Janeiro.

The moderator of this experiment is a school teacher, who initially explained to the participants the process of collaborative writing. On this part of the experiment, the moderator writes: "They had the opportunity to better understand the reality of two worlds (hearing world and deaf world). Students listeners were impressed with the writing of deaf students, then had to explain about this condition of the deaf colleagues. In short, I explained that Libras (Brazilian Sign Language) is the first language of the deaf person and written Portuguese is a second language, and how they think in Libras have difficulty in organizing writing with the grammatical structure of Portuguese. But that does not make them worse than the listeners, but different." LM.

The Tool Used in the Experiment

According to Tammaro et al. (1997), a cooperative editor should provide mechanisms to assist in the interaction that occurs when people are working collaboratively on editing documents. Thus, their characteristics should include:

- Be flexible to meet the usability needs of users;
- Allow editing of synchronously and asynchronously;
- Maintaining the integrity of the document being edited by different users and the possibility of merging to combine the contributions;
- Have an access policy that defines roles and their permissions and restrictions on the handling of documents;
- Multiple versions of a document should be maintained so that someone can come back with a version where appropriate. It is also important that the record which was the collaboration of each co-author;
- It is important to offer options for comment where the user can make their comments to a document or specific parts of it;
- It is also important to allow communication between the co-authors, so that they can discuss and exchange ideas;
- Allow the creation of a workflow for the construction of the document and publication of the same co-authors;
- When a document is created and shared, users who have access to it should be notified;
- Similarly, when a document is changed, all users that share should be alerted; and
- Provide monitoring a document allowing, for example, who is now visiting, who never accessed etc.

With the aim of using a tool that would meet the largest possible number of the above characteristics, was chosen editor GoogleDocs® to be used in the experiment. It needs only a registration email and is a system for easy access.

Also, lets you invite others to join the work, being a tool of collective works, which also offers basic editing tasks with a desktop simple and easy to understand, also allows these operations can be performed in conjunction with other participants in real time, and, provides a synchronous or asynchronous editing.

According to Machado (2009), this tool fosters interaction, exchange of ideas and collective production of texts. The exchanges can be established positively enabling creativity, critical thinking, responsibility and collaboration.

Google Docs provides ease of use, storage and online editing of files, access via the browser on multiple platforms, gratuity, requires no software installation and simple interface and be accessible over the web (Machado, 2009).

Development

For the development of the experimental parameters were considered, as described below:

- The chosen theme for the activity;
- The age group of invited participants: 13-18 years old;
- The theme of the text to be produced: "A whole country in the World Cup." The choice of this subject consider the proximity of the 2010 World Cup and also the age group of participants. The theme is easy to

understand and refers to a subject widely reported and discussed in the media, serving as a motivating factor for the interaction of participants;

- The number of participants: ten co-authors, five and five deaf listeners;
- All deaf participants are congenital deafness, and they were born with hearing loss were illiterate and without the reference of phrasal constructions and sounds of their own natural language;
- The participants had a period of seven days to build the text in a collaborative way.

Following the construction of the text in GoogleDocs®, we observed that the interaction occurred both synchronously, whereas some young people interact within their school environment in the same place and at the same time, as asynchronous distributed, because other interactions were carried out at different times and places.

Regarding to the contributions of the co-authors, was noted that participants complemented and not changed the text of the other participants, i.e., took care not to modify the text of another. On that point, the moderator reports: "Initially I was a little worried about the way that listener students understand the writing of deaf colleagues. But, after the explanation, I noted the positive reaction of the group. Demonstrated a willingness to help and decided not to correct spelling and concordance of texts developed by deaf students. And on the other hand, the deaf were more attentive to the listener colleagues' writing, and sometimes tried to correct himself." LM.

The participants answered a questionnaire consisting of eight questions, which together with the historical contribution of the participants contributed to the tool used to analyze the following measures:

- The level of knowledge of participants on the theme chosen for the preparation of the text;
- The degree of difficulty in using the tool;
- The level of interaction between participants in the collaborative writing process;
- The level of interaction between listeners and hearing impaired.

Analysis of Results

Data collected through the questionnaire, as well as the history contributions of participants in the cooperative editor used, were analyzed in a qualitative way. Thus, it was possible to consider whether the collaborative writing process developed by the group contributed to improve the interaction among hearing impaired and Listeners who participated in the experiment.

As shown in Figure 1, the majority of participants, both listeners as hearing impaired, has at least some knowledge on the topic "A whole country in the World Cup", chosen for the construction of the text. This indicates that the level of knowledge of participants on the topic facilitated the group interaction in this experiment.

Figure 1 also shows that most listeners had no difficulty in using the tool and the majority of hearing impaired had little or no difficulty. It was possible to note that facility presented in using the chosen tool for this experiment contributed to the group interaction during the editing cooperative process.

It was also noted that the difficulty that hearing impaired present to organize the writing, using the Portuguese grammatical structure, can be unknown even by listeners of the same social group, as is noted in the following testimony of a co- authors: *"I didn't know that deaf people write so differently."* MVS (Listener). In this aspect, the experiment allowed the participants co-authors could learn about the different ways to use their mother tongue.

Participants of the experiment lso report that has never participated in a collaborative writing process. However, all claim that the process of writing together facilitated the construction of the text. Regardless of the group is formed by people with different phrasal constructions, they note that the collective knowledge that allowed everyone to have gain knowledge on the topic: *"They spent more info for those who had little knowledge of the subject."* JHM (Listener). *"Knowing the subject matter and passing colleagues."* TF (Deaf). These reports demonstrate that the differences in the construction of phrases and expressions did not affect the group interaction.

Most deaf people said that interaction with the listeners was good and all the listeners said that interaction with the hearing impaired was excellent, as shown in Figure 01. This indicates a great level of acceptance and understanding of the listeners with respect to the difficulties in writing of the hearing impaired, as reported by the participants: *"To know a little deaf."* LR (Listener). *"Listeners and deaf good relationship."* SGO (Deaf).

It was observed that the collaborative writing process not only facilitated the group interaction in this activity, but also promoted greater acceptance of hearing impaired and contributed to the growth of the social bond, as reported by the participants: *"Learn to write better listener colleague."* TF (Deaf). *"Because I met so my deaf friends."* SK (Listener). *"As one completes the other."* DAO (Deaf).

Conclusions

The experiment has achieved the goal of simulating a real situation of interaction between listeners and the hearing impaired. According to the moderator, *"this experiment has enabled work on the theme World Cup, writing, creativity, respect and appreciation of their own productions and those of their colleagues."* This report confirms that it is very important that the effects of oral language on cognition are not overvalued by listeners about the performance of the hearing impaired, which would complicate its inclusion and real chances of a productive interaction. However, the experiment proved that through writing cooperative deaf participants were able to interact with listeners and together build a textual product, despite their differences in the use of the grammatical structure of Portuguese Language.

Regarding the phrasal construction, was observed in the experiment that participants did not change the sentences presented differently. This confirms that although there is significant difference in the level of management of instrumental language, the process of communication between participants was not harmed. This validates the collaborative writing as a mechanism of interaction between deaf and hearing, and shows that this mechanism favored a greater acceptance of differences among individuals and helped to improve the cognitive development of participants. On that point, the moderator noted: "The result achieved was excellent, and I know that through this tool GoogleDocs® students can further develop their written productions. I consider this tool as another way in the pursuit of cognitive development of my students, and enjoyed both working with her, I'm thinking of developing a job with a new group, involving all students." LM. This report shows that through this experiment it was possible to extend the possibilities of interaction between deaf and hearing through collaborative writing.

The reports presented in this paper guides analyze of results and confirm the acceptance of the collaborative work proposed by the experiment. They clearly show that the experiment helped to improve the interaction between deaf and listener not only in the construction of the text, but in building a more harmonious and real coexistence.

Future Work

This work does not exhaust the subject and, therefore, points out some future work that may be performed in order to contribute more to the interaction between hearing and deaf people using collaborative editing:

• Conduct the experiment with a control group, allowing comparison of results obtained by the groups, enriching the analysis;

• Measure the level of contribution of the participants in relation to the interaction of the texts submitted by others, checking with the resources offered by the tool for collaborative editing, the ability to follow through the recording of interactions made, which can later be measured and analyzed;

• Conduct the experiment with a larger group of participants, allowing also perform a quantitative analysis of the subjects addressed in this work.

Acknowledgments

This work was partially supported by CNPq (Brazilian Research Council and by FAPERJ, the Rio de Janeiro Research Council.

References

Adams, K., Adey, M. (2008). *Colaboração Social: Unindo Forças na Fronteira Digital*. Available at: http://www.scribd.com/doc/8997505/Whitepaper-Social-Collaboration. Accessed: 28/08/2010.

ASHA:American Speech Language Hearing Association (1989). Committee on Infant Hearing: guidelines for audiologic screening of newborn infants who are at risk for hearing impairment.

Gotti, M. O. (1991) Portuguese for the hearing impaired, Editora UNB, Brasilia (In Portuguese).

Marchesi, A. (1996) Comunicação, linguagem e pensamento, In C. Call, J. Palácios & A. Marchesi, "Desenvolvimento psicológico e educação" (pp.200-216), 1996.

Capovilla, F. C., Gonçalves, M.J. and Macedo, E. C. (1998). *Tecnologia em (Re)Habilitação Cognitiva: uma perspectiva multidisciplinar*. São Paulo, Brasil.

Ciccone, M. (1996), "Total Communication" 2nd ed., Rio de Janeiro, 1996 (In Portuguese).

Couto-Lenzi, A. (1997) "A integração das pessoas surdas." Informativo técnico-científico do INES, Rio de Janeiro, v.7, n.7, p.22-25, jun.1997.

Ellis, C.A., Gibbs, S. J., Rein, G. (1991) Groupware: some issues and experiences. In: *Communications of the ACM*, 34, 39–58.

Gatto, C. I., Tochetto, T. M. (2007). *Infantile hearing loss: implications and solutions.* Rev CEFAC, São Paulo – Brasil.

Howard, R.M. (2000). "Collaborative Pedagogy", In: Collaborative Pedagogies: A Bibliographic Guide, Ed. Tate, G., Ruppier, A. and Schick, K., Oxford University Press.

Machado A. C. T. (2009). The Tool Google Docs: Construction of Knowledge Through Interaction and Collaboration. Scientific Journal of Distance Education.

Mailhiot, G.B. (1968) Dynamique et genèse des groupes, actualité des découvertes de Kurt Lewin, Paris, France, Éditions de l'Épi.

Moustakas K., Nikolakis G., Tzovaras D., Deville B., Marras J., & Pavlek J. (2006). Multimodal tools and interfaces for the intercommunication between visually impaired and "deaf and mute" people. eNTERFACE'06, Dubrovnik, Croatia

NCHS - National Center for Health Statistics. Available at: http://www.cdc.gov/nchs/fastats/disable.htm Accessed: 28/08/2010.

Northern, L. J., Downs, P. M. (1989) "Audição em crianças." 13.ed. Trad. de Maria Lúcia Maciel França Madeira et al. São Paulo, 1989.

Olusanya BO, Luxon LM, Wirz SL. (2005) Detection of permanent childhood hearing loss in a developing country. Avaiable at: http://www.saferhealthcare.org.uk/ihi. Accessed: 28/08/2010.

Smith A. (2003). Preventing deafness: an achievable challenge. The WHO perspective. International Congress Series.

Santarosa, Lucila Maria Costi, Souza, Andréa Poletto, Loureiro, Cristiane de Barros Castilho. (2003). *Surdos e Cegos: Comunicação Mediada pela Tecnologia*. IV Congresso Iberoamericano – CIIEE.

Santarosa, Lucila Maria Costi, Lara, Alvira Themis S. (2002). Telemática: Um Novo Canal de Comunicação para Deficientes Auditivos. Faculdade de Educação – UFRGS, Porto Alegre – Brasil.

Tammaro, S. G., Mosier, J. N. (1997) Collaborative Writing Is Hard to Support: A Field Study of Collaborative Writing, The Journal of Collaborative Computing 6: 19–51.

White, R. K, Vohr B.R, Behrens T.R. (1993) Universal newborn hearing screening using transient evoked otoacoustic emissions: results of the Rhode Island hearing assessment project. Semin Hear.

Technological and cognitive roadmaps for Ambient Assisted Living

Filippo Cavallo (f.cavallo@sssup.it), Michela Aquilano (m.aquilano@sssup.it), Giuseppe Anerdi (g.anerdi@sssup.it), Maria Chiara Carrozza (carrozza@sssup.it), Paolo Dario (paolo.dario@sssup.it)
BioRobotics Institute
Scuola Superiore Sant'Anna, Pisa, Italy

Alberto Greco (greco@unige.it)
Lab. of Psychology and Cognitive Sciences
University of Genoa, Italy

Abstract

As a result of the continuous increase of ageing population, a growing number of frail and impaired people (both in physical and cognitive functions) need assistance and pose dramatic challenges for the health and welfare systems. The projects in the Ambient Assisted Living (AAL) field aim at developing intelligent systems able to provide personal assistance to elder and impaired people, allowing them to continue living in their familiar environment and preserving as much as possible their independence. AALIANCE is one European FP-7 project developed in the AAL context established for creating a common vision of AAL and developing a roadmap of future steps and projects on the way to AAL systems. In this paper the AALIANCE roadmap document is presented and the main trends towards AAL are addressed, both from technological and human-centered (psychological, cognitive) points of view.

The AALIANCE Project and the AAL Roadmaps

The industrialised world is experimenting a spectacular increase of ageing population, mainly consequent to the baby boomers generation crossing the threshold of retirement. Such a change in the demographic structure will affect all aspects of the life, including a larger number of frail and impaired people (both in physical and cognitive functions). In particular the risk associated with cognitive impairments is raising with the age and goes with elevated social and economic burdens relevant to its chronic nature and to the progressive loss of autonomy.

At the same time, while representing dramatic challenges for the health and welfare systems it will offer innovation and business opportunities for technology providers, fostering, among others, the development of ICT-enabled assisted living paradigm or "Ambient Assisted Living" (AAL). AAL refers to intelligent systems of assistance and represents a paradigm shift – in society as well as technology – that will walk hand in hand with "human centered computing", where the emphasis is on user friendliness, situation awareness, distributed service support for human interaction.

The AAL system through adaptive and distributed user-system interfaces, body and environment sensor network and AI subsystems, is designed to be able to infer about the activities of daily life (ADLs) the user is carrying out and about the context in which such activities are taking place.

In societal terms, AAL is focused to enable the containment of the overall cost of assistance, offering truly acceptable and usable solutions to the growing need for personal assistance, mainly related to the steadily growing cohort of elderly people, while increasing their independence and ultimately the quality of their lives.

In this context AALIANCE ("The European Ambient Assisted Living Innovation Alliance"), a European FP7 project ended in 2010, had the goal to create a common vision of AAL that provides and defines the necessary future R&D steps and projects on the way to AAL. It aimed at developing such a roadmap and strategic guidance for short-, mid- and long-term R&D approaches in the context of AAL. In this paper the AALIANCE Ambient Assisted Living Roadmap document (Van Den Broek et al., 2010) is presented, addressing the main trends towards AAL, both from technological and human-centered (psychological, cognitive) points of view.

The Roadmap starts with the analysis of the main trends towards AAL, analyzed from a demographic, economic and technological point of view, and the barriers for their deployment, identified for each stakeholder of AAL, i.e. users and caregivers (primary stakeholders), organizations offering services (secondary stakeholders), organizations supplying goods and services (tertiary stakeholders) and organizations analyzing the economical and legal context of AAL (quaternary stakeholders). The needs of elderly and disabled people were studied and identified considering the two fundamental aspect/expectations of the human, that are the *autonomy* and the *independence* in different contexts of daily life.

Starting from these reflections, three main application domains of AAL technologies were identified: AAL for Persons, AAL in the Community, and AAL at Work. These areas reflect the three fundamental contexts of human daily life, that are the personal and health care considered from the physical and cognitive aspects; the participation to the Community activities; the work considered not only as means of economical subsistence but also as instrument for the expression of the own personality, attitudes and self-fulfillment.

Unfortunately, physiological deficits and pathologies invalidate the autonomy and the independence of elderly and impaired persons in these domains, causing problems for both the single subjects and their families and also for the Community that should satisfy an increasing request of services and support. Advanced technologies (ICT,

Table 1: Application Domains defined in AALIANCE AAL Roadmap.

AAL APPLICATION DOMAINS		
4 Persons (@home; @mobile)	@ Community	@Work
- Health, rehabilitation, care - Safety and security - Activity management/cognitive support - Autonomy / physical support - Person-centric services	- Mobility - Social Inclusion -Entertainment and Leisure	- Collaboration, cooperation

Ambient Intelligence, Ubiquitous and Service Robotics, etc.) can be useful tools to help directly elderly people in their ADLs and to support caregivers and sociomedical workers in their jobs.

The AAL Roadmap describes some realistic scenarios related to the single application domain in which ICT and technologies can contribute for helping and guaranteeing people independence (see Table 1). Then technologies and innovations that should be implemented to accomplish these scenarios have been identified describing current technologies and the technological objectives foreseen for the short-, mid- and long-term and how academic and industrial S/T researches should evolve to obtain them. This analysis was carried out looking in particular to all main functionalities that compose the AAL integrated systems: sensing, reasoning, acting, interaction, communication, power supply (Table 2).

Furthermore, the document delineates possible directives for the integration of technological support and contribute in the service settings supplied by public and private service providers (novel service models) and also economical and legislative actions that the society should perform for introducing the AAL technologies in the community guaranteeing and preserving at the same time the safety and the welfare of all citizens.

From AAL Roadmap to the Scuola Superiore Sant'Anna research trends

The AALIANCE AAL Roadmap has been considered by the European Commission as good guidelines for the development of novel European S/T researches so that the Work Programme 2011 Cooperation, Theme 3, says that the next Coordination frameworks developed in the context of the Challenge 5, Objective ICT-2011.5.4 "ICT for Ageing and Well-being", Target C, should "...take into account work already started under the AALIANCE innovation platform (ref. http://www.aaliance.eu)" (European Commission, 2010).

Several research groups are using the AALIANCE AAL Roadmap to address their researches in the AAL fields and service robotics in order to design and develop technological solutions useful and exploitable in society to support elderly and disabled people in their ADLs. In particular the Scuola Superiore Sant'Anna (Pisa, Italy) is carrying out several researches related to AAL and service robotics making reference to AAL Roadmap.

In the context of the calls for experiments organized by the ECHORD Project funded by FP7, the Scuola Superiore Sant'Anna is involved in ASTROMOBILE project (Cavallo et al., 2010; Nepa et al., 2010). This project aims at designing an assistant robot, starting from the robotic platform SCITOS G5 (MetraLabs GmbH), able to provide some services to elderly people (such as monitoring the presence at home and the health status, connecting the user with outside people, reminding him/her about drugs to be taken and events, etc.), to cooperate with users in executing ADLs at home, to interact with them adopting strategies similar as much as possible to natural human ones, and that can be considered usable and acceptable by this particular target of users. The ASTROMOBILE system foresees the integration of the mobile robot with a sensor network and an ambient intelligence infrastructure. During this project an extensive experimentation in realistic indoor environment, that is DomoCasa living lab, and with real elderly people will be carried out.

The project follows a User Centered Design methodology and so it will involve the older persons, through meeting, interviews and other surveys, since the

Table 2: Summary of the main functions and enabling technologies described in AALIANCE AAL Roadmap.

MAIN FUNCTIONS AND ENABLING TECHNOLOGIES	
Sensing	New sensors (health parameters, environment, localization, etc.) Sensor networks
Reasoning	Ontologies Event stream processing Probabilistic reasoning Event prediction
Acting	New actuators Home robots and mechatronic devices
Interacting	Multimodal, natural, persuasive interfaces
Communicating	New protocols and standards for communication network

beginning of the project, in order to address the design of the ASTROMOBILE system (aesthetics and appearance of the robot, way to carry out the domestic services, human-robot interfaces, etc.) for being really suited for elderly people and satisfying the users expectations.

RITA is a different project, developed in the context of Tuscany region. It has been conceived by researchers of Scuola Superiore Sant'Anna coming from different research fields, robotics and law, that are designing novel ICT technological solutions to support and monitor elderly people living alone, and are investigating how these technologies can be exploited and inserted in the context of socio-medical services supplied from public regional providers, and how to deal with both legal and socio-ethical aspects.

For this project a sensor network and ambient intelligence infrastructures will be developed to monitor health status of old subjects, to localize them both indoors and outdoors, to monitor the domestic environments, to recognize potentially dangerous events and to activate alert actions for providing promptly help to the elderly user. All these technological tools will be designed according to indications provided from real elderly people, their caregivers and socio-medical workers, involved through the socio-medical service provider of Pisa's areas, that subsequently will also test the systems.

Moreover, the Scuola Superiore Sant'Anna is carrying out other researches oriented to the robotics and ICT technologies for "Ageing Well" because strongly believes that the technology can help people to live better and is so advanced to be ready to be implemented in real life because acceptable and safe.

Towards a cognitive roadmap

In AAL, products and services, provided by a heterogeneous set of disciplines, are based on selected standards that allow the interoperability of applications and designed in a user-centered way. The main objectives that should direct such a research effort are: from a personal view, to allow impaired people to continue living in their familiar environment; from a social point of view, to reduce assistance costs.

In the first phase technological aspects have been mainly addressed. There are, however, a number of issues to be raised from a human-centered perspective: psychological, cognitive, social, etc. In particular, in the present section we would like to emphasize some psychological and cognitive points that should be integrated with the technological guidelines in the next developments. Such approach belongs to the cognitive science domain, because the contribution of different disciplines is needed.

Acceptance and attitudes

The first implication is about the *psychological acceptance* of an artificial companion. In a previous work (Greco, Anerdi, Rodriguez, 2009) we conducted a preliminary empirical study in order to assess the acceptance of an animaloid robot during a simple interaction session with cognitively impaired elders. Starting from this study we claimed that such acceptance should be considered as a multifaceted attitude, where affective, cognitive, and conative aspects have equal importance. In our opinion, the next steps in the cognitive slope of a roadmap firstly include psychological surveys planning, concerning the acceptability, and possible implementation approaches,

(i) of a pervasive monitoring of daily activities,
(ii) of the presence of an AAL system,
(iii) of interaction with such a system.

In fact, the presence of a continuous monitoring system may give some individuals a suffocation feeling; the possible negative effects of other components of an AAL system must also be carefully scrutinized. And even if the system is accepted, misunderstandings or deceptions can occur about its intended purpose or use. Willingness to interact also should not be taken for granted.

The methodological tools for these surveys include standard interviews, questionnaires, (directed also to relatives), systematic observation sessions, related assessment tools, and also new pilot experimentation on the field.

If this first line of inquiry is just aimed at collecting information, a second line to be developed concerns tools aimed at attitude formation and change. A human-centered AAL system must be part of a *positive attitude system*. This can be developed by enhancing in potential users:

(i) a correct *knowledge* about the system, not only from the technical point of view but also concerning its general philosophy and purposes;

(ii) a positive *feeling*, obtained from a well-done affective computing, able to correctly recognizing user emotions, and to behave in an emotionally sensible way;

(iii) a motivated *intention* to use the system.

We claim that this attitude can only be developed through *interaction*, where the system and the user progressively adapt themselves and each other in order to achieve the final state, where the user has built an increasingly positive attitude, as explained above, and the system has built an increasingly refined user model.

User model

The *user model* is a key concept in this context. It can be developed as a set of general frames, available at the start as a standard toolkit, to be tailored and customized on the way to the user needs, through *interactive learning*. Each frame should include general expectations concerning a different area, belonging to the different enabling technologies for AAL. In order to establish these expectations, an assumption of normality could be the right starting point. In other words, the model could start with the default assumption that a user has a particular skill unless the contrary is proved, possibly in test situations.

We shall consider now some technological aspects specified by the Roadmap (see Table 2). The next mentioned functional components should neither be conceived as standalone mechanisms carrying out computations for their own use, nor just as replacement or improvement devices for faulty intellective processes and resources. Each of them, instead, must be designed as a true subsystem devised to interact with the user in order to build a personalized model of this particular person.

Sensing

Sensors are planned to be placed everywhere, in-body or on-body, in-appliance or on-appliance, or in the environment (see Figure 1). The user also, on his part, has a sensing system, and both systems should interact. In the technical Roadmap much attention is correctly posed to the requirement that sensors be seamless, integrated, less disturbing as possible. It is also important to realize, however, that the huge potential quantity of information conveyed by a complex network of sensory systems is not seamless from the cognitive point of view, and that more and more complex links with high-level processes are needed, so that truly relevant features in context are selected. The fundamental problem here is how to shape a potentially very rich model to be “aware” of the personal limitations of a specific user.

Reasoning

It concerns aggregating, processing and analyzing data, transforming it into knowledge within different and often cross-connected spaces (body, home, vehicle, public spaces). These processes, in fact, may assume a function that can integrate some well known deficits in verbal and visuospatial cognitive capabilities: attending relevant task aspects and inhibiting irrelevant ones, dealing with cognitive workload in complex environments (Newell et al., 2008).

Higher levels of reasoning, as illustrated in Figure 1, concern categorization of user activities and situation recognition. To this purpose, a cognitive task analysis may be useful to help identify critical steps in action performance. Here, again, the aim is not to reach a representation of the ideal performance, but exactly recognizing where the user deficits lie. And this should be done dynamically, because such deficits are not fixed once for all, but can change with time and contexts.

Acting

It is based on sensor-motors (e.g., cognitive robots) and actuators that, based on their perception of the environment, on behalf of their users and on information gathered by some other device and/or background knowledge, could perform concrete actions. These agents should be able to learn the personal habits of the user, that make him/her different from other users, and also individual styles, that often make the same person to perform the same thing differently. It should be able also to detect which actions are performed as automatic or controlled processes.

Interacting and Communicating

An intelligent interaction with systems and services is devised, in order to cope with the abilities of users. Sensors and actuators are connected to one or more reasoning systems that in turn might be connected (even dynamically, e. g. a person moving from home to a vehicle and then to some public space) to other reasoning systems, possibly with their own sensors and actuators.

One important point concerns here the development of suitable interfaces, which may be difficult for cognitively impaired people. It may be obvious that an interface should be as more humanlike as possible. This is not only because elders often lack technical or computing systems knowledge, but also because they should not consider it as a mechanical device, but as a tool that works similarly and extends their natural and usual communication with people relevant in their life. In this vein, as an example, the development of gesture-based commands is worth to be pursued. It can be used, in part, also as a mediating tool for establishing a real communication system with family.

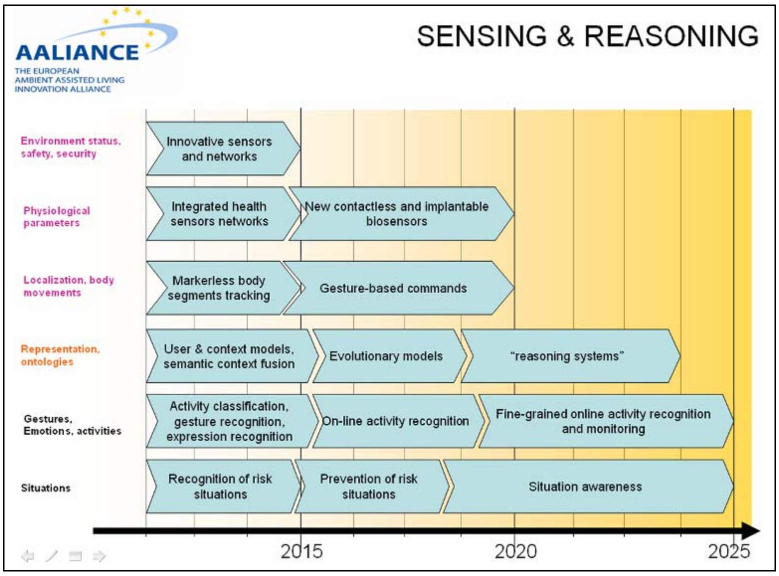


Figure 1: Sensing and Reasoning Roadmap (Van Den Broek, Cavallo, Wehrmann, 2010).

Conclusion

In this paper we have presented the AALIANCE roadmap final document, that addressed some technological trends and established shared guidelines towards building and experimenting realistic Ambient Assisted Living systems, designed to support elders and cognitively impaired persons during their everyday life. We outlined also some basic elements of a cognitive science framework for a future extension toward a cognitive roadmap. This short analysis is, of course, just a starting point, and it does not claim to be systematic and exhaustive. Many details are also being developed in other research groups (e.g. Langdon, Persad & Clarkson, 2010).

The key points we identified here for such an extension are: a methodology that involves users through inquiries aimed at assessing the acceptance of AAL systems, in principle and in some potentially uncomfortable details; the establishing of both a positive attitude in the user and of a sound user model in the system, via interactive learning. We stressed the requirement that such a model be able to autonomously and dynamically detect user inadequacies and to learn user practices. The interaction interface should give the feeling of a natural communication with significant people, like family members. Ultimately, we did not mention about it, but it goes straightforward: a method of performance testing has to be created in order to ascertain the efficacy of the proposed solutions.

Acknowledgments

Authors acknowledge the support of the AALIANCE Project, grant FP7-217050, ECHORD Project, grant FP7-231143, and RITA Project, funded by Region of Tuscany POR FSE 2007-2013.

References

Cavallo F., Aquilano M., Stieger F., Bonaccorsi M., Stieger M., Breznik A., Grasch P., Carrozza M.C., and Dario P. (2010). *ASTROMOBILE Project Assistive SmarT RObotic platform for indoor environments: MOBILity and interaction. Italian Forum on Ambient Assisted Living*,Trento, Italy, 6-8 October 2010.

Dario, P., Anerdi, G. Micera, Carpaneto J. (2006) Biorobotics for Longevity. *51° SIGG Congress*, Florence.

European Commission (2010). *Work Programme 2011 – Cooperation, Theme 3, ICT – Information And Communications Technologies*.

Greco A., Anerdi G., & Rodriguez G. (2009). Acceptance of an Animaloid Robot as a Starting Point for Cognitive Stimulators supporting Elders with Cognitive Impairments. *Revue d'Intelligence Artificielle/Artificial Intelligence Review*, 23, 4, 523-537.

Langdon P., Persad U., Clarkson P.J. (2010). Developing a model of cognitive interaction for analytical inclusive design evaluation. *Interacting with Computers*, 22, 510–529.

Nepa P., Cavallo F., Bonaccorsi M., Aquilano M., Carrozza M.C., Dario P. (2010). Experimental Analysis of RSSI-Based Indoor Location Systems with WLAN Circularly Polarized Antennas. *MobiHealth 2010*, LNICST 55, pp. 176—183.

Newell, A.F., Carmichael, A., Gregor, P., Alm, N., Waller, A. (2008). Information technology for cognitive support. In: Sears, A., Jacko, J.A. (Eds.), *The Human–Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications Human Factors and Ergonomics,* Section 5, Designing for Diversity, 2nd Ed., CDC, Taylor & Francis, pp. 811–882, Chapter 41.

Population projections 2008-2060, Eurostat Press Office.

Van Den Broek G., Cavallo F., Wehrmann C. (2010) AALIANCE Ambient Assisted Living Roadmap. *Ambient Intelligence and Smart Environments*, IOS Press, Volume 6.

The effect of time pressure and task completion on the occurrence of cognitive lockup

Ernestina J. A. Schreuder (elianeschreuder@hotmail.com)
TNO Human Factors Research Institute, kampweg 5
Soesterberg, 3769 ZG The Netherlands

University of Utrecht,
Department of Psychology, Heidelberglaan 1,
Utrecht 3584 CS The Netherlands

Tina Mioch (tina.mioch@tno.nl)
TNO Human Factors Research Institute, kampweg 5
Soesterberg, 3769 ZG The Netherlands

Abstract

Prior studies have suggested that time pressure and task completion play a role in the occurrence of cognitive lockup. However, supportive evidence is only partial. In this study, we conducted an experiment to investigate how both time pressure and task completion influence the occurrence of cognitive lockup, in order to better understand situations that could trigger the phenomenon. We found that if people have almost completed a task, the probability for cognitive lockup increases. We also found that the probability for cognitive lockup decreases, when people execute tasks for the second time. There was no effect of time pressure or an interaction effect found between task completion and time pressure. The results provide further support for the explanation that cognitive lockup up is the result of a decision making bias and that this bias could be triggered by the perception that a task is almost complete.

Introduction

This study is about an inescapable part of action, something all human beings experience to a greater or lesser extent: human error. Human errors can happen in everyday situations with only limited consequences. However, errors can also happen in high-performance environments like in aviation, where they can have tremendous effects and be life threatening. For instance, when pilots forget to enable their landing gear when landing and as a result crash. Hence, it is important to investigate why human errors in aviation are made and how errors can be avoided.

In the past, several cognitive explanations and theories have been proposed to understand why pilots deviate from normative activities (e.g. Wickens and Hollands, 2000; Dekker, 2003). The European project HUMAN (www.human.aero) strives to pave a way for making this knowledge readily available to designers of new cockpit systems, in order to be able to design cockpits that prevent pilots from making errors. They identified cognitive lockup to be among the most relevant mechanisms for modern and future cockpit human machine interfaces. Cognitive lockup is the tendency to deal with disturbances sequentially (Moray and Rotenberg, 1989). As a result of cognitive lockup operators are inclined to focus on the current task and are reluctant to switch to another task, even if that task has a higher priority (Neerincx, 2003). The following flight incident illustrates the possible consequences of cognitive lockup. During landing, the pilot of flight 401 of Eastern Air Lines (1973) was warned about a problem with the landing gear. To win time, the pilot canceled the landing, set the plane in the autopilot mode and started solving the problem with the landing gear. This problem fully occupied the pilot and multiple warnings about a decreasing altitude (a low-altitude alarm, a remark of the air-traffic controller) were ignored. As a consequence, the plane crashed, resulting in the death of most people on board.

Experimental studies exist where cognitive lockup was manifested in the data (Moray and Rotenberg, 1989; Kerstholt, Passenier, Houttuin and Schuffel, 1996), however, an explanation for the occurrence of the phenomenon was not given. The following literature overview provides accounts for the occurrence of the phenomenon.

Reduced situational awareness as trigger for cognitive lockup

A popular concept (Meij, 2004) is the idea that a reduced situational awareness (SA) might cause cognitive lockup. Kerstholt and Passenier (2000) argued that if operators become less aware of the actual situation, for instance, due to automation, they may not be able to understand the links between the various subsystems they have to control. If operators lack the knowledge of the underlying systems cognitive lockup is more likely to occur. Kerstholt and Passenier suggested that knowledge of the underlying systems is important in order to increase SA and prevent cognitive lockup.

Jones and Endsley (1996) investigated flight accidents caused by a reduced SA. They found that a great part of the flight accidents was due to a failure to monitor or observe relevant data that were clearly present in the situation. This type of accidents could also have been the result of cognitive lockup, as important tasks were that were triggered while dealing with another problem, were ignored. Therefore, it could be suggested that cognitive lockup is triggered by a failure to monitor the environment. However, Meij (2004) found in his study on cognitive lockup, that cognitive lockup could not be explained by a neglect of environment. Hence, a reduced SA and specifically the failure to monitor the environment seem not to be an underlying mechanism of cognitive lockup.

High cognitive workload as explanation for cognitive lockup

Cognitive workload refers to the information processing demands imposed by the performance of cognitive tasks (Johnson and Proctor, 2004). In order to predict the cognitive load of a specific task, Neerincx (2003) developed the cognitive task load (CTL) theory. This theory proposed three underlying factors of cognitive task load: (1) time occupied, (2) number of task set switches, which is the number of active tasks in execution or planned to do, and (3) level of information processing. Neerincx (2003) suggested that cognitive lockup would occur when time occupied and the number of task set switches are high. Grootjen, Neerincx and Veltman, (2006) conducted experiments in order to validate the CTL theory. In these real-life experiments participants had to deal with emergencies that appeared on a ship. They found that when all three factors were high people experienced cognitive overload; they did not know what to do. However, no evidence of cognitive lockup was found. Furthermore, Grootjen et al. (2006) found that in the overload situation, participants switched much more between tasks than in the optimal strategy. This result is likely to indicate that a high workload does not influence the occurrence of cognitive lockup.

In the same vein, Meij (2004) investigated whether a lack of cognitive resources could be an explanation for cognitive lockup. He argued that tasks that require a more complex diagnosis process are expected to demand more cognitive resources and thus cause a higher cognitive workload and might cause cognitive lockup. He found, however, that the level of complexity of information processing did not affect the degree of cognitive lockup. Hence, Meij (2004) indicated that cognitive lockup is not caused by the fact that people lack the cognitive resources to switch. This result underscores that it is likely that a high workload does not cause cognitive lockup.

Too high switching costs as explanation for cognitive lockup

When people switch between tasks, people are slower and less accurate than when they repeat tasks (Jersild, 1927; Monsell, 2003) and these switching costs have been attributed to a variety of processes. Pashler (1994) suggested that switching costs arise because of a cognitive bottleneck to process or select information. This means that when a cognitive process is devoted to a primary task, this process can not start for a second task. This second task has to wait, yielding switching costs. Schumacher (1999) and Meyer and Kieras (1997a, 1997b) argued that switching costs arise due to an executive control mechanism. They proposed a class of adaptive executive control models in which it is assumed that people have flexible control over the course of secondary task processing stages. They argued that the fact that switching costs arise is not due to a cognitive bottleneck but is rather optional and strategic. The reconfiguration to another task takes time and thus, switching costs arise.

Meij (2004) proposed that people might decide to stick to their current task because the switching costs that accompany task switching are perceived as too high. He found that cognitive lockup was reduced when it was obvious that the benefits of a switch to another task were higher than the costs of a switch. He argued that the participants were biased in their decision-making process, as the costs of switching to another task had to be disproportionally low before participants decided to abandon the current task. Although the results showed that the tendency for cognitive lockup was considerably reduced when the costs to switch were low, the tendency for cognitive lockup was still present. Therefore, Meij suggested that besides switching costs, other factors that trigger cognitive lockup are involved.

A decision making bias as trigger for cognitive lockup

A decision making bias refers to the fact that decision making can be influenced by a prejudice or 'one-sided' perspective. A bias can be unconscious or conscious in awareness. Meij (2004) believed that cognitive lockup is due to a decision making bias. When dealing with a task people decide to switch or not to switch to another task when triggered. This decision might be biased due to for instance a misperception of expected benefits. As a result, people could decide not to deal with an additional more urgent task until the ongoing task is dealt with. To find out whether cognitive lockup results from a decision making bias, Meij (2004) conducted several experiments with a fire control task. In this task participants had to extinguish fires on a ship in a computer simulation. When a fire appeared the participants had to detect the fire by clicking on the fire. After detection, participants had to extinguish the fire by selecting the right treatment. He used this task because in his first experiment he successfully demonstrated that cognitive lockup could be found using this task.

Meij (2004) proposed that both task completion and prior investments, such as money, time and effort, might bias the decision to switch to another task. Meij found that when prior investments are high and the task was near completion (high task completion), the probability for cognitive lockup increased. Remarkably, he also found that when prior investments are high and task completion is low, people have the tendency to abandon their task. Hence, the probability for cognitive lockup decreased. Meij argued that in the high prior investments condition perceived time pressure may be higher than in the low prior investments condition. This is because the available time in relation to invested time is lower in the high prior investments condition. Therefore, he attributed the effect of prior investments to the perception of time pressure. He suggested that when time pressure is perceived as high and the ongoing task is almost completed, people are more likely to stick to the ongoing task than in situations where time pressure is high and the ongoing task needs considerable activities in order to complete it. Thus, the results of Meij's experiments (2004) are likely to indicate that cognitive lockup is due to a decision making bias. This decision making bias seems to be triggered when time pressure and task completion are high.

Current study: the effect of time pressure and task completion on cognitive lockup

From the literature overview it seems that cognitive lockup is the result of a biased (un)conscious decision to focus on the current task and ignore others. Time pressure and task completion seem to influence this biased decision. The aim of this study is to investigate how both time pressure and task completion influence the occurrence of cognitive lockup. Therefore, this study extends the results found by Meij (2004). Furthermore, the aim is to identify critical situations in cockpit environments that allow for designing cockpit systems that help pilots avoid critical situations and decrease the probability for cognitive lockup. In the following subsections we translate the research question into hypotheses.

Time pressure. Time pressure is the perception that time is scarce. According to Beevis (1999) people experience time pressure when the time required to execute tasks is more than 70% of the total time available for the tasks. Beevis (1999) suggested that people experience high time pressure when 85% of the available time is required to execute the tasks. In this case performance is often impaired in that some tasks are not (well) executed. Other researchers (e.g. van der Kleij, 2009; De Dreu, 2003; Durham et al., 2000; Karau and Kelly, 1992; Kelly and Loving, 2004) indicated the following consequences of time pressure. At the individual level, time pressure leads to (1) faster performance rates, because people stop considering multiple alternatives, (2) lower performance quality, due to the engagement in superficial rather than thorough and systematic processing of information, and (3) more heuristic information processing, as a result of refraining from critical probing of a given seemingly adequate solution or judgment. At group level, increasing levels of time pressure narrows team members' focus on a limited range of task-salient cues in both team interaction patterns and team task performance. This narrow focus of attention that often manifests as a restricted information exchange is due to a filtering process (Kelly and Loving, 2004). Groups working under time pressure attend to all of the information available but then selectively discuss only information that seems especially relevant (Kelly and Loving, 2004). They also found that under high time pressure group members see task completion as their main interaction objective, and the group attempts to reach consensus and complete the task as quickly as possible, but at the sacrifice of quality. Groups under mild or no time pressure can, in contrast, consider a wider set of task features, devote their resources to performing on the task as well as possible, and tend to employ more effortful systematic information processing that gives serious considerations to all possible alternative solutions for a task (Kelly and Loving, 2004). Interestingly, DeDonno and Demaree (2008) found that the mere perception of time pressure as well as real time pressure impair performance.

Thus, time pressure can trigger heuristic information processing that make people focus on an ongoing task (van der Kleij, 2009; De Dreu, 2003; Durham et al., 2000; Karau and Kelly, 1992; Kelly and Loving, 2004). As a result, time pressure might influence the tendency to stick to the ongoing task and influence the occurrence of cognitive lockup. However, in situations where time pressure[1] was high Grootjen et al. (2006) found a high cognitive workload, but they found no relation to cognitive lockup. Therefore, we expect that the effect of time pressure alone is not strong enough to capture people in their current task. We propose the following hypothesis:

1. **Time pressure** has no effect on cognitive lockup. That means, in case people deal with a task, and another more urgent task is triggered, people switch to the more urgent task just as often under time pressure as when there is no time pressure.

Task completion. Task completion literature (Garland and Colon, 1993; Boehne and Paese, 2000; Humphrey S.E., Moon, H., Conlon, D.E., Hofmann D.A., 2004) shows that people have the tendency to complete a task even if it is economically unwise to do so. Garland and Colon (1993) and Boehne and Paese (2000) found that this tendency is strongest when 90% of the task is completed compared to 10% or 50% completion of a task. Meij (2004) found an effect of task completion on cognitive lockup. People tend to complete a task when they are almost finished (high task completion) even when a more urgent task is triggered. We expect that the results of Meij (2004) are replicated in this study. Therefore, we propose the following hypotheses:

2. **Task completion** has an effect on cognitive lockup.
    a. When task completion is high, the probability for cognitive lockup increases. That means, in case people deal with a task, and another more urgent task is triggered, people tend to stick to the current task, when they have almost completed this task.
    b. When task completion is low, the probability for cognitive lockup decreases. That means, in case people deal with a task, and another more urgent task is triggered, people tend to switch to the more urgent task when the first task is not nearly completed.

[1] In the research of Grootjen et al. (2006) time pressure was high when the percentage of the available time for a task that people are occupied with the task was high.

Interaction between time pressure and task completion. Meij (2004) suggested that when task completion and time pressure are high the tendency for cognitive lockup increases. When time pressure is high but task completion is low, the tendency for cognitive lockup decreases. In fact, he investigated the interaction effect between prior investments and task completion and attributed the effect of prior investments to the perception of time pressure. Thus, Meij never investigated the effect of time pressure and task completion. Therefore, in this study, we test whether high time pressure and high task completion indeed increase the probability for cognitive lockup and that high time pressure and low task completion decrease this probability, as suggested by Meij. We propose the following hypotheses.

3. There is an interaction effect between **task completion and time pressure** on cognitive lockup.
   a. When time pressure is high and task completion is high, the probability for cognitive lockup is highest compared to all other conditions. That means, in case people deal with a task, and another more urgent task is triggered, people tend to finish the first task before switching to the more urgent when they feel time pressure and have almost completed the first task.
   b. When time pressure is high and task completion is low, the probability for cognitive lockup is lowest compared to all other conditions. That means, in case people deal with a task, and another more urgent task is triggered, people tend to switch to the urgent task before executing the first task when they feel time pressure, but still need to complete many stages to complete the first task.

This study extends the study conducted by Meij (2004) as we expect to provide further evidence for the explanation that cognitive lockup is caused by a decision making bias, and that this bias could be triggered by time pressure and task completion.

Method

Participants

The experiment counted 46 participants. The participants consisted of:

- TNO trainees/employees (15)
- students of the University of Utrecht (20)
- (ex)members of the Hockey Club Rotterdam (6)
- other (5)

All participants were experienced computer users and most of them were highly educated. They were all between the age of 18 and 35 years old. Psychology students received course credits for participation in the experiment. All participants could win 20 Euro when they had the highest score in the experiment.

Apparatus and material

The experiment included two laptop computers with headphones. Java software was installed on the computers to run the experimental task and a training session. Before the experiment, participants received a hardcopy manual printed on paper with A4 format. During the experiment, participants could use a question-tree (see Figure 2) printed on paper with A4 format for reference purposes.

Procedure

The experiment was conducted at TNO in a computer room, at the University of Utrecht in a laboratory, and in a private setting with the use of laptops. The experiment took ca. 30 minutes per participant: 15 minutes for a training (including test scenarios) and 15 minutes for the experiment.

The participants received an information letter upfront the training and an informed consent document after the training. The aim of the training was to familiarize the participants with the experimental task. The training involved reading the training manual, which was accompanied by a verbal instruction and executing test scenarios on the computer. After the training the participants were asked whether they felt comfortable with performing the task. If so, the experiment was started. If not, questions could be asked and the test scenarios could be done again until the participant was comfortable with executing the task

The experimental task was a computer simulation in which participants had to fight fires on a ship. The fire-fighting software was chosen because Meij (2004) already demonstrated that with this fire-fighting task cognitive lockup could be found.

Experimental task. In the experimental task participants had to fight fires on a ship. Two types of fires existed:

- normal fires, which were indicated by a red triangle
- urgent fires, which fires were indicated by a blue triangle in a yellow background

Next to the fire type, fires had specific features. For example, a fire could be an oil fire, a fire could be life threatening, injured people could be involved, and/or smoke could trouble the sight of the firemen. Therefore, each fire required a specific action based on the fire's specific features. To find out the fire specifics, participants could ask four predefined questions. These questions appeared as buttons on the screen. To ask a question, participants had to press the question button. Figure 1 shows the screen that was visible to the participants once a fire was present. When a question was asked, the system closed for four seconds to answer the question with Y (Yes) or N (No) for a normal fire. In case of an urgent fire the system closed for one second to answer the question. This was because an urgent fire was more dangerous for the ship and needed quick handling. Please note that when the system was closed nothing could be done. Based on the answers generated by the system, participants could select the appropriate action to extinguish the fire. Figure 2 shows the question-tree which indicates the appropriate action. Seven predefined actions could be chosen. The action buttons also appeared on the bottom of the screen, once a fire was present (see Figure 1). An appropriate action extinguished the fire; a wrong action shut down the system for seven seconds. Thereafter, a new action could be selected, if time allowed it.

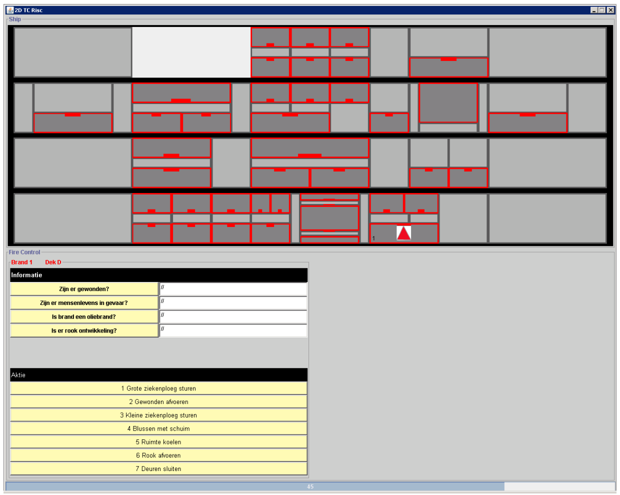


Figure 1: Screen of the ship visible to participants once a fire was present.

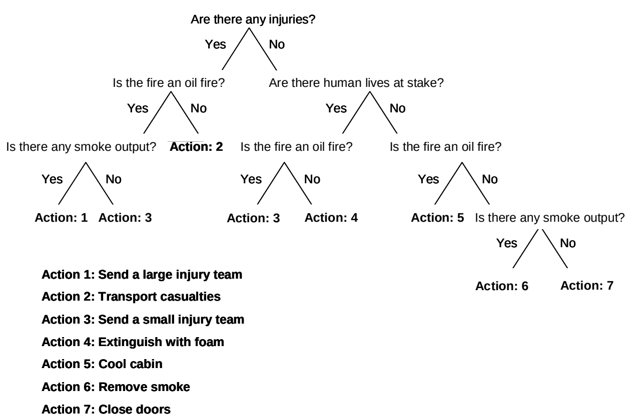


Figure 2: Question-tree and relevant actions to extinguish fires.

Participants knew how much time they had to extinguish the first fire, as this was indicated by the timer at the bottom of the screen. The time to stop a second fire was not indicated. This was done to make the decision to switch to the second fire similar for the different test scenarios (see scenarios). In case the time available would be shown for the second fire as well as, participants could depend their decision to switch on the time available for both fires. In the high task completion condition relatively little time would be left for the first fire and therefore relatively more time would be left to extinguish the second fire, which would give the incentive to finish the first fire. In the low task completion condition relatively much time would be left to extinguish the first fire, and relatively less time would be left to extinguish the second fire. This would give the incentive to switch to the second fire. As a result the switch incentives in these situations would not be comparable. This problem was solved by not showing the time available for the second fire.

If participants extinguished a fire in time they could win points. However, if they did not extinguish the fire in time, points were deducted and a burn down was the result (see Table 1). This was done to emphasize the fact that an urgent fire was more urgent than a normal fire, as more points could be won or lost by respectively extinguishing or missing an urgent fire. If one fire burned down, another fire could still be extinguished if time allowed it.

In the test scenarios (see scenarios) the second fire was always an urgent fire. The time was set in such a way that if a participant finished the first fire before switching to the urgent fire, the urgent fire would burn down. This was also done to emphasize the fact that an urgent fire was urgent. If it was not handled quickly it burned down. This also meant that if participants suffered from cognitive lockup (finishing the first fire instead of extinguishing the urgent fire first), they would have had a lower score than participants who did not suffer cognitive lockup. To ensure the motivation of the participants a reward of 20 Euro was promised to the participant with the highest score. Participants only saw their score at the end of the experiment.

Table 1: Fire-fighting scores.

Action	Points
Extinguish normal fire	1
Extinguish urgent fire	3
Burn down normal fire	-1
Burn down urgent fire	-3

Experimental design

The main goal of this experiment was to investigate how time pressure and task completion influence the occurrence of cognitive lockup. In order to enhance the sensitiveness to find cognitive lockup, we operationalized cognitive lockup in two ways. In the first definition cognitive lockup was only found when participants did not switch to the urgent fire, when dealing with the first fire. In the second definition cognitive lockup was found when participants significantly delayed their switch to the urgent fire. Therefore, the second analysis was more sensitive to find cognitive lockup than the first analysis.

1) Cognitive lockup 1 was defined as completing the first fire before extinguishing the second more urgent fire. Cognitive lockup 1 was measured as the mean percentage of scenarios in which cognitive lockup occur.
2) Cognitive lockup 2 was defined as delayed switching to the second more urgent task once presented, while executing a first task. Cognitive lockup 2 was measured as the action time to switch to the urgent fire, once present (a significantly delayed action time indicates cognitive lockup).

In order to investigate the influence of time pressure and task completion on cognitive lockup, time pressure and task completion were manipulated within-subjects.

Task completion. Task completion was defined as the percentage of the total number of stages of a task that have been completed. People have the tendency to stick to their current task when 90% or more of the total stages of a task have been completed (Boehne and Pease, 2000; Garland and Colon, 1993). We investigated whether people refrain from switching to an urgent second task, when they have almost completed a first task. Therefore, task completion was manipulated by the onset of the urgent fire. The onset depended on the number of questions that had been asked in order to extinguish the first fire. In the high task completion condition the urgent fire appeared after three questions had been asked. In this condition task completion of the first task was 75%, as 75% of the total stages of the task had been completed (three questions answered out of four). This percentage was chosen because this was the pre-final stage for extinguishing the first fire, and closest to 90% task completion as mentioned by Boehne and Pease (2000) and Garland and Colon (1993). In the low task completion condition the urgent fire appeared after one question had been asked. In this case task completion was 25%, as 25% of the total stages of the first task had been completed (one question answered out of four).

Time pressure. Time pressure was defined as the percentage of the available time to execute a task that is required to execute the task. People experience time pressure when 70% or more of the available time is required for the task (Beevis, 1999). People experience high time pressure when 85% of the available time is required for the task (Beevis, 1999). The average time to fight a fire depended on the number of questions that had to be asked to extinguish a fire. Based on the pilot results it appeared that for a normal fire the time required to extinguish a fire based on four questions was 22,1 seconds; based on three questions this was 18,6 seconds and based on two questions this was 14,3 seconds. We investigated whether time pressure on a first task would refrain people from switching to a second more urgent task. Therefore, time pressure was manipulated. This was achieved by increasing or decreasing the time available for fighting the first fire.

In the high time pressure condition the available time to extinguish the first fire was 25 seconds for fires that needed four questions and 20 seconds for fires that needed three questions. Fires that needed two questions were not present in the test scenarios (see scenarios). In this way time pressure was ≥ 88%, as 88% or more of the available time was required for the task. As a result, in the high time pressure and high task completion scenario the available time to extinguish the first fire was almost over when the urgent fire appeared. Therefore, participants had to choose for the urgent fire at the cost of a burn down of the first fire in this condition. Thus, they could never obtain the total payoff of four points in this scenario. They could also choose to extinguish the first fire first, but in that case the urgent fire would burn down as explained before.

In the low time pressure condition the available time to extinguish the first fire was 55 seconds for all fires. In this condition time pressure was ≤ 40%, as 40% or less of the available time was required for the task. The available time of 55 seconds was chosen because perceived time pressure of the first fire could be influenced by the appearance of the second more urgent fire. To make sure that participants would perceive little time pressure in the low time pressure condition, 55 seconds allowed the participants to start with the first fire, and when the urgent fire appeared to switch to the urgent fire and once the urgent fire was extinguished, to switch to the first fire again and extinguish the first fire. Thus, in the low time pressure scenarios the maximum score of four points could be obtained. Table 2 summarizes the test conditions in terms of task completion and time pressure.

Table 2: Test conditions.

Test condition	Time pressure (%)	Task completion (%)
Low-Low	≤ 40	25
Low- High	≤ 40	75
High-Low	≥ 88	25
High-High	≥ 88	75

Scenarios. The experiment consisted of 25 scenarios. The scenarios included 8 test scenarios (2 times all test conditions) and 17 irrelevant scenarios. The irrelevant scenarios were designed in order to accomplish uncertainty, so the participants would not understand the test scenarios. The test scenarios can be described as follows:

1. In the scenario where time pressure was low and task completion was low, participants had 55 seconds to fight the first fire. The urgent fire appeared when they had asked one question of the first fire. The urgent fire needed to be extinguished in 17 seconds. If participants decided to extinguish the first fire first, the available time allowed participants to start with the urgent fire, after they had extinguished the first fire, but they would never be able to extinguish it. In this way the participants would not be demotivated, which would be the case when the urgent fire had already burned down, while still fighting the first fire.
2. In the scenario where time pressure was low and task completion was high, participants had 55 seconds to fight the first fire. The urgent fire appeared when they had asked three questions of the first fire. The urgent fire needed to be extinguished in 13 seconds. In this way it was impossible to extinguish the first fire first, and afterwards extinguish the urgent fire.
3. In the scenario where time pressure was high and task completion was low, participants had 20 seconds to fight the first fire. This was because the first fire could be extinguished after three questions. The urgent fire appeared when they had asked one

question of the first fire. Like scenario 1, the urgent fire needed to be extinguished in 17 seconds.

4. In the scenario where time pressure was high and task completion was high, participants had 25 seconds to fight the first fire. This was because the first fire could only be extinguished after four questions. The urgent fire appeared when they had asked three questions of the first fire. Like scenario 2 the urgent fire needed to be extinguished in 13 seconds.

The scenarios (test and irrelevant scenarios) were presented in random order to avoid order effects. Only the test scenarios were analyzed.

Statistical design

The experimental design was a repeated measures design, as each test condition consisted of two scenarios. Thus, the participants received all test conditions twice. Therefore, we used a 2x2x2 repeated measures ANOVA with attempts, time pressure and task completion as factors to analyze the data.

Before the experiment was executed, we conducted a power analysis in order to examine the number of participants needed for the experimental design. The power analysis for a factorial ANOVA suggested a sample size (N) of 45 to achieve a power of 0.80 for detecting a medium effect size (0.26) and alpha set at 0.05. Hence, this design required a sample size of 45 participants to be able to conduct further statistical analysis. Based on this result, 46 participants were recruited.

Results

Sample data

In the experiment 46 cases with two repeated measures were recorded. Three records were removed as one participant did not follow the experiment instruction correctly and pressed action buttons without asking questions. Therefore, this data could not be analyzed as the urgent fire was not triggered. Table 3 shows the number of valid records per test condition.

Table 3: Number of valid records per test condition.

Test condition		Number of records	
Time Pressure	Task Completion	Attempt 1	Attempt 2
Low	Low	46	46
Low	High	45	45
High	Low	46	46
High	High	45	46

Cognitive lockup 1

Figure 3 shows the mean percentage of scenarios in which cognitive lockup 1 (CL1) was found, taking time pressure and task completion into account. In other words, the Figure shows the mean percentage of scenarios in which the normal fire was extinguished before the participants switched to the more urgent second fire.

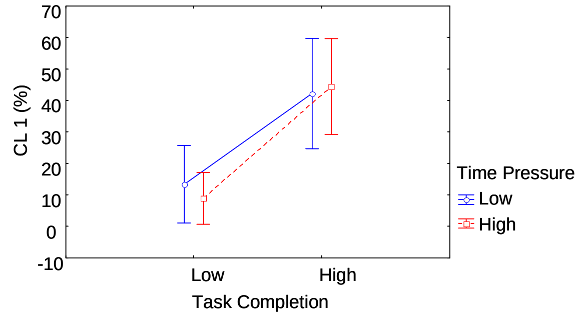


Figure 3: The mean percentage of scenarios in which CL1 was found.

The main effect of task completion on CL1 was significant, $F(1,44) = 36.857$, $p < .001$. This means that the percentage of scenarios in which CL1 was found was significantly higher in the high condition (Mean = 43%) compared to the low condition (Mean = 11%). In other words when participants had almost extinguished the first fire (one question to go to find out the correct action) more people finished the first fire before switching to the urgent fire than when the participants still had three questions to go to find out the correct action. There was no effect of time pressure or an interaction effect found between task completion and time pressure on CL1. Next to these results, a significant main effect was found for the factor attempts (not shown in Figure 3). The participants received all test conditions twice. The results show that the mean percentage of scenarios in which CL1 was found was significantly higher in the first attempt (Mean = 34%) compared to the second attempt (Mean = 21%) $F(1,44) = 10.203$ $p < .003$. This indicates a learning effect.

Figure 4 shows the results for attempt 1 and 2 separately. For both attempt 1 and 2, a significant main effect for task completion was found $F(1,44) = 26.362$, $p < .001$ and $F(1,44) = 24.750$, $p < .001$, respectively. No effect was found for time pressure. In addition, no interaction effect between time pressure and task completion was found.

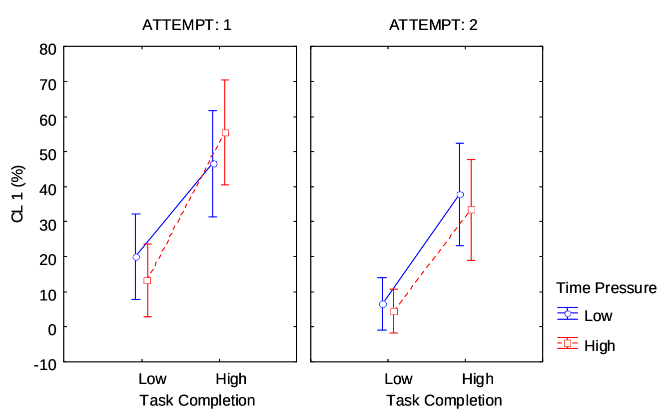


Figure 4: The mean percentage of scenarios in which CL1 was found for attempt 1 and 2.

Cognitive lockup 2

Figure 5 shows the mean reaction times of participants to switch to the second more urgent fire, while fighting the first fire. Cognitive lockup 2 (CL2) was found when the reaction times were significantly longer in a specific condition.

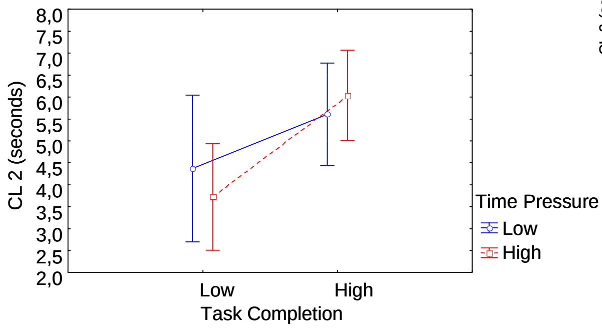


Figure 5: Time needed in seconds to switch to the urgent fire.

The main effect of task completion on CL2 was significant, F(1,44) = 15.182, p < *.001*. This means that the participants needed more time to switch to the urgent fire in the high condition (Mean = 5,9 seconds) compared to the low condition (Mean = 4,1 seconds). In other words, when participants had already asked three questions of the first fire, it took significantly longer to switch to the urgent fire compared to the situation where they had only asked one question. The average time to 'switch' to the first fire, once it was present, was 1,1 second. There was no effect of time pressure or an interaction effect found between task completion and time pressure on CL2. Next to these results, a significant main effect was found for the factor attempt (not shown in Figure 5). The participants received all test conditions twice. The results show that it took longer to switch to the urgent fire in the first attempt (Mean = 5,8 seconds) compared to the second attempt (Mean = 4,1 seconds; F(1,44) = 15,444 p < *.001*). This indicates a learning effect.

Figure 6 shows the results for attempt 1 and attempt 2 separately. For both attempt 1 and 2, a significant main effect for task completion was found F(1,44) = 5.922, p < *.019* and F(1,44)= 14.404, p < *.001* respectively. No effect was found for time pressure. In addition, no interaction effect between time pressure and task completion was found.

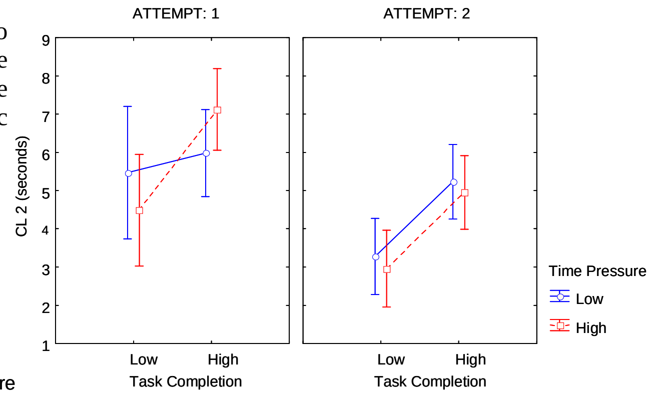


Figure 6: Time needed in seconds to switch to the urgent fire for attempt 1 and 2.

General Discussions

The main goal of this experiment was to investigate how both time pressure and task completion influence the occurrence of cognitive lockup. Firstly, we hypothesized that time pressure alone would not have an effect on cognitive lockup (hypothesis 1). The results of the experiment support this hypothesis. In case people deal with a task, and another more urgent task is triggered, people switch to the more urgent task just as often and just as fast under time pressure as when there is no time pressure. This result implies that although time pressure can trigger heuristic information processing and thereby influence decision making (e.g. van der Kleij, 2009; De Dreu, 2003; Durham et al., 2000; Karau and Kelly, 1992; Kelly and Loving, 2004), people are able to assess the priority of different tasks while dealing with a task, and switch to the most important task if necessary. They are not captured in their current task when facing time pressure.

We should mention that in the high time pressure and high task completion condition there was an incentive to extinguish the normal fire first, before switching to the urgent fire. As a result, this incentive stimulated the chance to find cognitive lockup. The incentive existed because a choice had to be made between the normal fire and the urgent fire. While dealing with the normal fire, participants could see that if they switched to the urgent fire, they would miss the normal fire. However, if they extinguished the normal fire, there was only a chance that they would miss the urgent fire. The participants did not know that the maximum number of points could never be obtained, as they would indeed miss the urgent fire, when they extinguished the normal fire first. Participants with a risk avoiding strategy would switch to the urgent fire as this fire could be missed if they did not switch immediately (two points versus minus two points). However, participants that were very confident with extinguishing the fires and who had a risky result maximizing strategy could try to extinguish both fires instead of one (four points versus two points). They would extinguish the normal fire first. In this way, there was an incentive to finish the normal fire first in the high task completion and high time pressure condition. While this incentive was present, nevertheless participants decided to switch to the urgent fire in the high time pressure and high task completion condition as often as in the low time pressure and high task completion condition. This underscores our finding that time pressure has no influence on the occurrence of cognitive lockup.

Furthermore, we should notice that we used a static time deadline to manipulate time pressure. Other ways exist to induce time pressure. For example, a more dynamic task can be used in which deadlines evolve with different speed. This type of dynamic time pressure might have different effects on performance (Kerstholt and Willems, 1993). As a consequence, the results of this study only apply to settings in which time pressure is induced by a static deadline. Further research is needed to analyze the effect on behavior when time pressure is dynamic.

Secondly, we hypothesized that task completion would have an effect on cognitive lockup (hypothesis 2). The results of the experiment support this hypothesis. The results show that there is a main effect of task completion on cognitive lockup (CL1 and CL2). People that have almost completed a task tend to finish this task even when a more urgent task is triggered. In other words, when task completion is high the probability for cognitive lockup increases (hypotheses 2a). People that still need to complete many stages before a task is completed tend to switch to the more urgent task, when triggered. Thus, when task completion is low the probability for cognitive lockup decreases (hypotheses 2b). These results were found despite a learning effect. We found that cognitive lockup was less present during the second attempt of a test scenario compared to the first attempt. We believe that this was due to a learning effect. People learned from the feedback they received in the first attempt of a scenario, and if needed they changed their strategy in the second attempt. However, the task completion effect remained significant in the second attempt. Thus, although participants lost points when they completed the first fire and as a result missed the urgent fire in the first attempt of the high task completion scenario, they did not change their strategy when the scenario was executed again.

It could be argued that the participants might not have perceived the urge of the urgent fires and as a result stayed with the normal fires. Although the urgent fires had a very different icon, behaved differently (system closure of one second instead of four after a button had been pressed) and generated more (less) points when extinguished (burned down) than normal fires, the categorization might not have been meaningful enough for the participants. However, participants showed in the high time pressure and high task completion condition, in which they had to choose between the urgent and normal fire, that they switched to the urgent fire as often as in the low time pressure and high task completion condition. Such behavior would not have been expected when categorization and consequences were not clear.

This study replicates the results of Meij (2004) as he also found an effect of task completion on cognitive lockup. Furthermore, this finding extends the results reported by Boehne and Pease (2000) and Garland and Colon (1993). In their experiments they found the tendency to complete a task when the task has already been completed for 90%. The present study shows that this tendency is already present when a task has been completed for 75%. In our experimental setup this was achieved when three task stages had been completed out of a maximum number of four stages. It could be argued that participants perceived a higher task completion percentage as only one stage was still required to complete the task.

These results imply that the perception that a task is almost competed could lead to critical situations when another more urgent task is triggered. This urgent task might be ignored as a result of cognitive lockup. In order to avoid cognitive lockup we believe that the tendency to complete a task when it is almost completed should be broken. For instance, this might be done by altering the perception that a task is almost completed or by unlearning this tendency. Further research is needed to investigate how to break the tendency to complete a task when it has almost been completed and a more urgent task is triggered.

Finally, according to the results of Meij (2004) we hypothesized that there would be an interaction effect between time pressure and task completion (hypothesis 3). The results do not support this hypothesis as the interaction effect between time pressure and task completion on cognitive lockup (CL1 and CL2) was not significant. The result implies that time pressure does not enhance the task completion effect, as expected. Thus, when task completion is high the probability for cognitive lockup is not increased when people face time pressure. The present study shows that the interaction effect found by Meij (2004) between prior investments and task completion on cognitive lockup cannot be explained by the perception of time pressure when prior investments are high, as he suggested. An alternative explanation cannot be given as prior investments were not investigated in this experiment. Further research should be done to find an explanation for the effect of prior investments on cognitive lockup and why this factor interacts with task completion.

The results of this study provide further support for the explanation that cognitive lockup up is the result of a decision making bias and that this bias could be triggered by the perception that a task is almost completed. This has important implications for the designs of cockpits, as it indicates that decision support tools seem more important in reducing the probability for cognitive lockup than, for example, tools that reduce cognitive workload. The decision support tool should assist pilots to focus on the most urgent task. However, a critical situation exists when the pilot has almost completed a task and a more urgent task is triggered. The decision support tool might be ignored, as a result of cognitive lockup. To avoid critical situations it is important that this decision support tool also helps pilots to break the tendency to complete a task when it is almost finished and another more urgent task is triggered. In this way pilots can act appropriately and deal with the most urgent task.

Training

Although the experimental design was not specifically built to investigate an effect of training, a significant learning effect was found. Participants showed less cognitive lockup (CL1 and CL2) the second time they executed the test scenarios, compared to the first time. As mentioned before, we believe that participants learned from the feedback they received in the first attempt of a scenario. When participants decided to stick to the first fire, it resulted in a burn down of the urgent fire. Their payoff in that scenario would have been minus two. As participants executed the test scenarios twice, they adjusted their strategy in order to improve their payoff. This finding supports the results reported by Kerstholt and Passenier (2000). They suggested that if people understood the underlying system, cognitive lockup was less likely to occur. Training might therefore be a factor that influences cognitive lockup as it increases the knowledge of a system and thereby reduces the probability for cognitive lockup. For the design of cockpits this implies that the decision to automate processes should be done carefully. This is because automation might decrease the understanding of underlying systems as operators are not involved anymore in the normal process (Wickens and Hollands, 2000; Kerstholt and Passenier, 2000). As a result, the likelihood for cognitive lockup might increase. Further research should be done to provide evidence for the suggested effect of training on the occurrence of cognitive lockup.

Task domain and participants

It can be argued that the experiment was conducted in a specific task domain (fire-fighting task) and with a specific set of participants and that generalizing the results should be done carefully. We assume that cognitive lockup is a general cognitive mechanism or heuristic that is domain independent and can happen to every human being. This assumption is based on the fact that heuristics that are used to solve well defined problems are general-purpose or domain independent heuristics, in that they can be applied to a wide range of situations or domains and do not involve specific capabilities (Groom, 2002). Well defined problems are well specified and the knowledge required to find the solution is present in the instructions given, e.g. a puzzle (Groom, 2002). The fire fighting task used in this experiment can be regarded as a well defined problem. As cognitive lockup was found in this problem solving setting, cognitive lockup is assumed to be a general purpose heuristic. This implies that although the results are obtained in a fire-fighting domain, they can be applied to any other domain. The same applies for the sample that has been used. The participants of the experiment were Dutch, between 18-32 years old and most of them were highly educated. As we assume that cognitive lockup does not depend on specific capabilities, we suggest that the results from this sample can be applied to human beings in general.

Acknowledgments

The work described in this paper is funded by the European Commission in the 7th Framework Programme, Transportation under the number FP7 – 211988.

References

Beevis, D., Bost, R., Doering, B., Nordø, E., Oberman, F., Papin, J., et al. (1999). *Analysis techniques for man-machine system design, crew system ergonomics*. Ohio.

Boehne, D.M. & Paese, P.W. (2000). Deciding whether to complete or terminate an unfinished project: a strong test of the project completion hypothesis. *Organizational Behavior and Human Decision Processes*, 2, 178-194.

DeDonno, M.A., Demaree, H.A. (2008). Perceived time pressure and the Iowa Gambling Task. *Judgment and Decision Making*, Vol.3, No. 8 December 2008, pp 636-640.

De Dreu, C.K.W. (2003). Time pressure and closing of the mind in negotiation. *Organizational Behavior and Human Decision Processes* 91, 280–295.

Dekker, S. (2003). Failure to adapt or adoptions that fail: contrasting models on procedures and safety. *Applied Ergonomics*, 34(3), 233–238.

Durham, C.C., Locke, E.A., Poon, J.M.L., McLeod, P.L. (2000). Effects of group goals and time pressure on group efficacy, information-seeking strategy, and performance. *Human Performance* 13, 115–138.

Conlon, D. E., & Garland, H. (1993). The role of project completion information in resource allocation decisions. *Academy of Management Journal*, 36, 402–413.

Groom et al. (2002). *An introduction to cognitive psychology processes and disorders*. Psychology Press Ltd, Hove, UK.

Grootjen, M., Neerincx, M.A., Veltman J.A. (2006). Cognitive task load in a naval ship control centre: from identification to prediction. *Ergonomics*, Vol. 49, Nos. 12-13, 10-22 October 2006, 1238-1264.

Jersild, A. T. (1927). Mental set and shift. *Archives of* Psychology, 14 (Whole No. 89).

Johnson A., Proctor R.W. (2004). *Attention: Theory and Practice*. Sage Publication, Thousand Oaks.

Jones, D.G., and Endsley, M.R. (1996). Sources of situational awareness errors in aviation. *Aviation, Space, and Environmental Medicine*, 67, 507-512.

Humphrey S.E., Moon , H., Conlon, D.E., Hofmann D.A. (2004). Decision-making and behavior fluidity: How focus on completion and emphasis on safety changes over the course of projects. *Organizational Behavior and Human Decision Processes* 93, 14–27.

Karau S.J., and Kelly J.R. (1992). The Effects of Time Scarcity and Time Abundance on Group Performance Quality and Interaction Process. *Journal of Experimental Social Psychology*, 542-571.

Kelly, R.J., and Loving T.J. (2004). Time pressure and group performance: Exploring underlying processes in the Attentional Focus Model. *Journal of Experimental Social Psychology*, 40, 185-198.

Kerstholt, J.H., and Passenier, P.O. (2000). Fault management in supervisory control: the effect of false alarms and support. *Ergonomics*, 43(9), 1371-1389.

Kerstholt, J., Passenier, P., Houttuin, K., & Schuffel, H. (1996). The effect of a priori probability and complexity on decision making in a supervisory control task. *Human Factors*, 38(1), 65–78.

Kerstholt, J.H., Willems, P. (1993) The effect of the time restrictions on information search and information integration in a dynamic task environment, *TNO-Report TM-01-B004,* TNO technische menskunde, Soesterberg.

Kleij, van der R., Lijkwan, J.T.E., Rasker P., De Dreu, C.K.W. (2009). Effects of time pressure and communication environment on team processes and outcomes in dyadic planning. *Int.J. Human-Computer Studies*, 67 (2009) 411-423.

Meij, G. (2004). *Sticking to plans: capacity limitation or decision-making bias?* Doctoral dissertation, Department of Psychology, University of Amsterdam, Amsterdam.

Meyers, D.E. & Kieras, D.E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: part 1. *Basic mechanisms. Psychological Review*, 104, 3-65.

Meyers, D.E. & Kieras, D.E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: part 2. Accounts of psychological refractory-period phenomena. *Psychological Review*, 104, 749-791.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7,134–140.

Moray, N., & Rotenberg, I. (1989). Fault management in process control: eye movements and action. *Ergonomics* 32 (11), 1319-1342.

Neerincx, M.A. (2003). Cognitive task load design: model, methods and examples. In: E. Hollnagel (ed.), Handbook of Cognitive Task Design. Chapter 13 (pp. 283-305). Mahwah, NJ: Lawrence Erlbaum Associates.

NTSB. (1973). Eastern airlines l-1011, miami, florida, december, 29, 1972 (Tech. Rep. No. NTSB-AAR-73-14). Washington, DC: National Transportation Safety Board (NTSB).

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116, 220-244.

Schumacher, E., Lauber, E., Glas, J., Zurbriggen, E., Gmeindl, L., Kieras, D., et al. (1999). Concurrent response-selection processes in dual-task performance: Evidence for adaptive executive control of task scheduling. *Journal of Experimental Psychology*: *Human Perception and Performance*, 25, 791–814.

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

Representing Concepts in Artificial Systems: A Clash of Requirements

Marcello Frixione (mfrixione@unisa.it), **Antonio Lieto (alieto@unisa.it)**
Department of Political, Social and Communication Sciences
University of Salerno, Italy

Abstract

The problem of concept representation is relevant for many subfields of cognitive research, including psychology, philosophy and artificial intelligence. In particular, in recent years, it received great attention within knowledge representation, because of its relevance for knowledge engineering and for ontology-based technologies. However, the notion of concept itself turns out to be highly disputed and problematic. In our opinion, one of the causes of this state of affairs is that the notion of concept is in some sense heterogeneous, and encompasses different cognitive phenomena. This results in a strain between conflicting requirements, such as, for example, compositionality on the one side and the need of representing prototypical information on the other. AI research in some way shows traces of this situation. In this paper we propose an analysis of this state of affairs. Since it is our opinion that a mature methodology to approach knowledge representation and knowledge engineering should take advantage also from the empirical results of cognitive psychology concerning human abilities, we sketch some proposal for concept representation in formal ontologies, which takes into account suggestions coming from psychological research. Our basic assumption is that knowledge representation technologies designed considering evidences coming from experimental psychology (and, therefore, more similar to the humans way of reasoning and organizing information) can have better results in real life applications (e.g. in the field of Semantic Web).

Introduction

Computational representation of concepts is a central problem for the development of ontologies and for knowledge engineering. Concept representation is a multidisciplinary topic of research that involves such different disciplines as Artificial Intelligence, Philosophy, Cognitive Psychology and, more in general, Cognitive Science. However, the notion of concept itself results to be highly disputed and problematic. In our opinion, one of the causes of this state of affairs is that the notion itself of concept is in some sense heterogeneous, and encompasses different cognitive phenomena. This results in a strain between conflicting requirements, such as, for example, compositionality on the one side and the need of representing prototypical information on the other. This has several consequences for the practice of knowledge engineering and for the technology of formal ontologies.

In this paper we propose an analysis of this situation. The paper is organised as follows. In section 2. we point out some differences between the way concepts are conceived in philosophy and in psychology. In section 3. we argue that AI research in some way shows traces of the contradictions individuated in sect. 2. In particular, the requirement of compositional, logical style semantics conflicts with the need of representing concepts in the terms of typical traits that allow for exceptions. In section 4 we review some attempts to resolve this conflict in the field of knowledge representation, with particular attention to description logics. It is our opinion that a mature methodology to approach knowledge representation and knowledge engineering should take advantage from both the empirical results of cognitive psychology that concern human abilities and from philosophical analyses. In this spirit, in section 5 we individuate some possible suggestions coming from different aspects of cognitive research: the distinction between two different types of reasoning processes, developed within the context of the so-called “dual process” accounts of reasoning; the proposal to keep prototypical effects separate from compositional representation of concepts; the possibility to develop hybrid, prototype and exemplar-based representations of concepts. We conclude this article (section 6) with some tentative suggestion to implement the above mentioned proposals within the context of semantic web languages, in terms of the linked data perspective.

2 Concepts in Philosophy and in Psychology

Within the field of cognitive science, the notion of concept is highly disputed and problematic. Artificial intelligence (from now on AI) and, more in general, the computational approach to cognition reflect this state of affairs. Conceptual representation seems to be constrained by conflicting requirements, such as, for example, compositionality on the one side and the need of representing prototypical information on the other.

A first problem (or, better, a first symptom that some problem exists) consists in the fact that the use of the term “concept” in the philosophical tradition is not homogeneous with the use of the same term in empirical psychology (see e.g. Dell’Anna and Frixione 2010). Briefly[1], we could say that in cognitive psychology a

[1] Things are made more complex by the fact that also within the two fields considered separately this notion is used in a heterogeneous way, as we shall synthetically see in the following. As a consequence, the following characterisation of

concept is essentially intended as the mental representations of a category, and the emphasis is on such processes as categorisation, induction and learning. According to philosophers, concepts are above all the components of thoughts. Even if we leave aside the problem of specifying what thoughts exactly are, this requires a more demanding notion of concept. In other words, some phenomena that are classified as "conceptual" by psychologists turn out to be "nonconceptual" for philosophers. There are, thus, mental representations of categories that philosophers would not consider genuine concepts. For example, according to many philosophers, concept possession involves the ability to make explicit, high level inferences, and sometimes also the ability to justify them (Peacocke 1992; Brandom 1994). This clearly exceeds the possession of the mere mental representation of categories. Moreover, according to some philosophers, concepts can be attributed only to agents who can use natural language (i.e., only adult human beings). On the other hand, a position that can be considered in some sense representative of an "extremist" version of the psychological attitude towards concepts is expressed by Lawrence Barsalou in an article symptomatically entitled "Continuity of the conceptual system across species" (Barsalou 2005). He refers to knowledge of scream situations in macaques, which involves different modality-specific systems (auditory, visual, affective systems, etc.). Barsalou interprets these data in favour of the thesis of a continuity of conceptual representations in different animal species, in particular between humans and non-human mammals: "this same basic architecture for representing knowledge is present in humans. [...] knowledge about a particular category is distributed across the modality-specific systems that process its properties" (p. 309). Therefore, according to Barsalou, a) we can speak of a "conceptual system" also in the case of non human animals; b) also low-level forms of categorisation, that depend on some specific perceptual modality pertain to the conceptual system. Elizabeth Spelke's experiments on infants (see e.g. Spelke 1994; Spelke and Kinzler 2007) are symptomatic of the difference in approach between psychologists and philosophers. Such experiments demonstrate that some extremely general categories are very precocious and presumably innate. According to the author, they show that newborn babies already possess certain *concepts* (e.g., the concept of physical object). But some philosophers interpreted these same data as a paradigmatic example of the existence of *nonconceptual* contents in agents (babies) that had not yet developed a conceptual system.

2.1 Compositionality

The fact that philosophers consider concepts mainly as the components of thoughts brought a great emphasis on *compositionality*, and on related features, such as productivity and systematicity, that are often ignored by psychological treatments of concepts. On the other hand, it is well known that compositionality is at odds with *prototypicality effects*, which are crucial in most psychological characterisations of concepts.

Let us consider first the *compositionality* requirement. In a compositional system of representations we can distinguish between a set of *primitive*, or *atomic symbols*, and a set of *complex symbols*. Complex symbols are generated starting from primitive symbols through the application of a set of suitable recursive syntactic rules (usually, starting from a finite set of primitive symbols, a potentially infinite set of complex symbols can be generated). Natural languages are the paradigmatic example of compositional systems: primitive symbols correspond to the elements of the lexicon (or, better, to morphemes), and complex symbols include the (potentially infinite) set of all sentences.

In compositional systems the meaning of a complex symbol *s* functionally depends on the syntactic structure of *s* and from the meaning of primitive symbols in it. In other words, the meaning of complex symbols can be determined by means of recursive semantic rules that work in parallel with syntactic composition rules. In this consists the so-called *principle of compositionality of meaning*, which Gottlob Frege identified as one of the main features of human natural languages.

In classical cognitive science it is often assumed that mental representations are compositional. One of the most clear and explicit formulation of this assumption is due to Jerry Fodor and Zenon Pylyshyn (1988). They claim that compositionality of mental representations is mandatory in order to explain some fundamental cognitive phenomena. In the first place, human cognition is *generative*: in spite of the fact that human mind is presumably finite, we can conceive and understand an unlimited number of thoughts that we never encountered before. Moreover, also *systematicity* of cognition seems to depend on compositionality: the ability of conceiving certain contents is related in a systematic way to the ability of conceiving other contents. For example, if somebody can understand the sentence *the cat chases a rat*, then she is presumably able to understand also *a rat chases the cat*, in virtue of the fact that the forms of the two sentences are syntactically related. We can conclude that the ability of understanding certain propositional contents systematically depends on the compositional structure of the contents themselves. This can be easily accounted for if we assume that mental representations have a structure similar to a compositional language.

2.2 Against "Classical" Concepts

Compositionality is less important for many psychologists. In the field of psychology, most research on concepts moves from the critiques to the so-called classical theory of concepts, i.e. the traditional point of view according to which concepts can be defined in terms of necessary and sufficient conditions. Rather, empirical evidence favours those approaches to concepts that accounts for prototypical effects. The central claim of the classical theory of concepts (i.e.) is that every concept *c*

the philosophical and psychological points of view is highly schematic.

is defined in terms of a set of features (or conditions) f_1, ..., f_n that are individually necessary and jointly sufficient for the application of *c*. In other words, everything that satisfies features f_1, ..., f_n is a *c*, and if anything is a *c*, then it must satisfy f_1, ..., f_n. For example, the features that define the concept *bachelor* could be *human*, *male*, *adult* and *not married*; the conditions defining *square* could be *regular polygon* and *quadrilateral*. This point of view was unanimously and tacitly accepted by psychologists, philosophers and linguists until the middle of the 20th century.

The first critique to the classical theory is due to a philosopher: in a well known section from the *Philosophical Investigations*, Ludwig Wittgenstein observes that it is impossible to individuate a set of necessary and sufficient conditions to define a concept such as GAME (Wittgenstein, 1953, § 66). Therefore, concepts exist, which cannot be defined according to classical theory, i.e. in terms of necessary and sufficient conditions. Rather, concepts like GAME rest on a complex network of *family resemblances*. Wittgenstein introduces this notion in another passage in the *Investigations*: «I can think of no better expression to characterise these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc.» (*ibid*., § 67).

Wittgenstein's considerations were corroborated by empirical psychological research: starting from the seminal work by Eleanor Rosch, psychological experiments showed that common-sense concepts do not obey to the requirement of the classical theory[2]: usually common-sense concepts cannot be defined in terms of necessary and sufficient conditions (and even if for some concept such a definition is available, subjects do not use it in many cognitive tasks). Rather, concepts exhibit *prototypical effects*: some members of a category are considered better instances than others. For example, a robin is considered a better example of the category of birds than, say, a penguin or an ostrich. More central instances share certain typical features (e.g., the ability of flying for birds, having fur for mammals) that, in general, are not necessary neither sufficient conditions.

Prototypical effects are a well established empirical phenomenon. However, the characterisation of concepts in prototypical terms is difficult to reconcile with the requirement of compositionality. According to a well known argument by Jerry Fodor (1981), prototypes are not compositional (and, since concepts in Fodor's opinion must be compositional, concepts cannot be prototypes). In synthesis, Fodor's argument runs as follows: consider a concept like PET FISH. It results from the composition of the concept PET and of the concept FISH. But the prototype of PET FISH cannot result from the composition of the prototypes of PET and of FISH. For example, a typical PET is furry and warm, a typical FISH is greyish, but a typical PET FISH is not furry and warm neither greyish.

[2] On the empirical inadequacy of the classical theory and on the psychological theories of concepts see (Murphy 2002).

Moreover, things are made more complex by the fact that, also within the two fields of philosophy and psychology considered separately, the situation is not very encouraging. In neither of the two disciplines does a clear, unambiguous and coherent notion of concept seem to emerge. Consider for example psychology. Different positions and theories on the nature of concepts are available (prototype view[3], exemplar view, theory theory), that can hardly be integrated. From this point of view the conclusions of Murphy (2002) are of great significance, since in many respects this book reflects the current status of empirical research on concepts. Murphy contrasts the approaches mentioned above in relation to different classes of problems, including learning, induction, lexical concepts and children's concepts. His conclusions are rather discouraging: the result of comparing the various approaches is that "there is no clear, dominant winner" (ibid., p. 488) and that "[i]n short, concepts are a mess" (p. 492). This situation persuaded some scholars to doubt whether concepts constitute a homogeneous phenomenon from the point of view of a science of the mind (see e.g. Machery 2005 and 2009; Frixione 2007).

3 Concept Representation in Artificial Intelligence

The situation sketched in the section above is in some sense reflected by the state of the art in AI and, more in general, in the field of computational modelling of cognition. This research area seems often to hesitate between different (and hardly compatible) points of view. In AI the representation of concepts is faced mainly within the field of knowledge representation (KR). Symbolic KR systems (KRs) are formalisms whose structure is, in a wide sense, language-like. This usually involves that KRs are assumed to be compositional.

In a first phase of their development (historically corresponding to the end of the 60s and to the 70s) many KRs oriented to conceptual representations tried to keep into account suggestions coming from psychological research. Examples are early semantic networks and frame systems. Frame and semantic networks were originally proposed as alternatives to the use of logic in KR. The notion of frame was developed by Marvin Minsky (1975) as a solution to the problem of representing structured knowledge in AI systems[4]. Both frames and most semantic networks allowed the possibility to characterise concepts in terms of prototypical information.

However, such early KRs where usually characterised in a rather rough and imprecise way. They

[3] Note that the so-called prototype view does not coincide with the acknowledgement of prototypical effects: as said before, prototypical effects are a well established phenomenon that all psychological theories of concepts are bound to explain; the prototype view is a particular attempt to explain empirical facts concerning concepts (including prototypical effects). On these aspects see again Murphy 2002.

[4] Many of the original articles describing these early KRs can be found in (Brachman & Levesque 1985), a collection of classical papers of the field.

lacked a clear formal definition, and the study of their meta-theoretical properties was almost impossible. When AI practitioners tried to provide a stronger formal foundation to concept oriented KRs, it turned out to be difficult to reconcile compositionality and prototypical representations. As a consequence, they often choose to sacrifice the latter.

In particular, this is the solution adopted in a class of concept-oriented KRs which had (and still have) wide diffusion within AI, namely the class of formalisms that stem from the so-called structured inheritance networks and from the KL-ONE system (Brachman and Schmolze 1985). Such systems were subsequently called terminological logics, and today are usually known as *description logics* (DLs) (Baader et al. 2002).

A standard inference mechanism for this kind of networks is *inheritance*. Representation of prototypical information in semantic networks usually takes the form of allowing exceptions to inheritance. Networks in this tradition do not admit exceptions to inheritance, and therefore do not allow the representation of prototypical information. Indeed, representations of exceptions can be hardly accommodated with other types of inference defined on these formalisms, concept classification in the first place (Brachman 1985). Since the representation of prototypical information is not allowed, inferential mechanisms defined on these networks (e.g. inheritance) can be traced back to classical logical inferences.

In more recent years, representation systems in this tradition have been directly formulated as logical formalisms (the above mentioned description logics, Baader et al., 2002), in which Tarskian, compositional semantics is straightly associated to the syntax of the language. Logical formalisms are paradigmatic examples of compositional representation systems. As a consequence, this kind of systems fully satisfy the requirement of compositionality. This has been achieved at the cost of not allowing exceptions to inheritance. By doing this we gave up the possibility of representing concepts in prototypical terms. From this point of view, such formalisms can be seen as a revival of the classical theory of concepts, in spite of its empirical inadequacy in dealing with most common-sense concepts.

Nowadays, DLs are widely adopted within many application fields, in particular within the field of the representation of ontologies. For example, the OWL (Web Ontology Language) system[5] is a formalism in this tradition that has been endorsed by the World Wide Web Consortium for the development of the semantic web.

4 Non-classical concepts in computational ontologies

Of course, within symbolic, logic oriented KR, rigorous approaches exist, that allow to represent exceptions, and that therefore would be, at least in principle, suitable for representing "non-classical" concepts. Examples are fuzzy logics and non-monotonic formalisms. Therefore, the adoption of logic oriented semantics is not necessarily incompatible with prototypical effects. But such approaches pose various theoretical and practical difficulties, and many unsolved problems remain.

In this section we overview some recent proposal of extending concept-oriented KRs, and in particular DLs, in order to represent non-classical concepts.

Recently different methods and techniques have been adopted to represent non-classical concepts within computational ontologies. They are based on extensions of DLs and of standard ontology languages such as OWL. The different proposals that have been advanced can be grouped in three main classes: a) fuzzy approaches, b) probabilistic and Bayesan approaches, c) approaches based on non-monotonic formalisms.

a) Following this direction, for as the integration of *fuzzy logics* in DLs and in ontology oriented formalisms, see for example Gao and Liu 2005, and Calegari and Ciucci 2007, Stoilos et al. (2005) propose a fuzzy extension of OWL, f-OWL, able to capture imprecise and vague knowledge, and a fuzzy reasoning engine that lets f-OWL reason about such knowledge. Bobillo and Staccia (2009) propose a fuzzy extension of OWL 2 for representating vague information in semantic web languages. However, it is well known (Osherson and Smith 1981) that approaches to prototypical effects based on fuzzy logic encounter some difficulty with compositionality.

b) The literature offers also several *probabilistic generalizations* of web ontology languages. Many of these approaches, as pointed out in Lukasiewicz and Straccia (2008), focus on combining the OWL language with probabilistic formalisms based on Bayesian networks. In particular, Da Costa and Laskey (2006) suggest a probabilistic generalization of OWL, called PR-OWL, whose probabilistic semantics is based on multi-entity Bayesian networks (MEBNs); Ding et al. (2006) propose a probabilistic generalization of OWL, called Bayes-OWL, which is based on standard Bayesian networks. Bayes-OWL provides a set of rules and procedures for the direct translation of an OWL ontology into a Bayesian network. A problem here could be represented by the "translation" from one form of "semantics" (OWL based) to another one.

c) The role of *non-monotonic reasoning* in the context of formalisms for the ontologies is actually a debated problem. According to many KR researches, non-monotonic logics are expected to play an important role for the improvement of the reasoning capabilities of ontologies and of the Semantic Web applications. In the field of non-monotonic extensions of DLs, Baader and Hollunder (1995) propose an extension of ALCF system based on Reiter's default logic[6]. The same authors, however, point out both the semantic and computational difficulties of this integration and, for this reason, propose a restricted semantics for open default theories,

[5] http://www.w3.org/TR/owl-features/

[6] The authors pointed out that "Reiter's default rule approach seems to fit well into the philosophy of terminological systems because most of them already provide their users with a form of 'monotonic' rules. These rules can be considered as special default rules where the justifications - which make the behavior of default rules nonmonotonic – are absent".

in which default rules are only applied to individuals explicitly represented in the knowledge base. Because of Reiter's default logic does not provide a direct of modelling inheritance with exceptions, Straccia (1993) proposes an extension of DL H-logics (*Hybrid KL-ONE style logics*) able to perform *default inheritance reasoning* (a kind of default reasoning specifically oriented to reasoning on taxonomies). This proposal is based on the definition of a priority order between default rules. Donini et al. (1998, 2002), propose an extension of DL with two non-monotonic epistemic operators. This extension allows one to encode Reiter's default logic as well as to express epistemic concepts and procedural rules. However, this extension presents a rather complicated semantics, so that the integration with the existing systems requires significant changes to the standard semantics of DLs. Bonatti et al. (2006) propose an extension of DLs with circumscription. One of motivating applications of circumscription is indeed to express prototypical properties with exceptions, and this is done by introducing "abnormality" predicates, whose extension is minimized. Giordano et al. (2007) propose an approach to defeasible inheritance based on the introduction in the *ALC* DL of a typicality operator **T**[7], which allows to reason about prototypical properties and inheritance with exceptions. This approach, given the nonmonotonic character of the **T** operator, encounters the problem of irrelevance (have some difficulties in the management of additional information that could be irrelevant for the reasoning). Katz and Parsia argue that ALCK, a non monotonic DL extended with the epistemic operator K[8] (that can be applied to concepts or roles) could represent a model for a similar non monotonic extension of OWL. In fact, according to the authors, it would be possible to create "local" closed-world assumption conditions, in order the reap the benefits of nonmonotonicity without giving up OWL's open-world semantics in general.

A different approach, investigated by Klinov and Parsia (2008), is based on the use of the OWL 2 annotation properties (APs) in order to represent vague or prototypical, information. The limit of this approach is that APs are not taken into account by the reasoner, and therefore have no effect on the inferential behaviour of the system (Bobillo and Straccia 2009).

5 Some Suggestions from Cognitive Science

Though the presence of a relevant field of research, there isn't, in the scientific community, a common view about the use of non-monotonic and, more in general, non-classical logics in ontologies. For practical applications, systems that are based on classical Tarskian semantics and that do not allow for exceptions (as it is the case of "traditional" DLs), are usually still preferred. Some researchers, as, for example, Pat Hayes (2001), argue that the non monotonic logics (and, therefore, the non monotonic "machine" reasoning for Semantic Web) can be maybe adopted for local uses only or for specific applications because it is "unsafe on the web". Anyway, the question about which "logics" must be used in the Semantic Web (or, at least, until which degree, and in which cases, certain logics could be useful) is still open.

The empirical results from cognitive psychology show that most common-sense concepts cannot be characterised in terms of necessary/sufficient conditions. Classical, monotonic DLs seem to capture the compositional aspects of conceptual knowledge, but are inadequate to represent prototypical knowledge. But a "non classical" alternative, a general DL able to represent concepts in prototypical terms does not still emerge.

As a possible way out, we sketch a tentative proposal that is based on some suggestions coming from cognitive science. Some recent trends of psychological research favour the hypothesis that reasoning is not an unitary cognitive phenomenon. At the same time, empirical data on concepts seem to suggest that prototypical effects could stem from different representation mechanisms. In this spirit, we individuate some hints that, in our opinion, could be useful for the development of artificial representation systems, namely: (i) the distinction between two different types of reasoning processes, which has been developed within the context of the so-called "dual process" accounts of reasoning (sect. 5.1 below); (ii) the proposal to keep prototypical effects separate from compositional representation of concepts (sect. 5.2); and (iii) the possibility to develop hybrid, prototype and exemplar-based representations of concepts (sect. 5.3).

5.1 A "dual process" approach

Cognitive research about concepts seems to suggest that concept representation does not constitute an unitary phenomenon from the cognitive point of view. In this perspective, a possible solution should be inspired by the experimental results of empirical psychology, in particular by the so-called dual process theories of reasoning and rationality (Stanovich and West 2000, Evan and Frankish 2008). In such theories, the existence of two different types of cognitive systems is assumed. The systems of the first type (type 1) are phylogenetically older, unconscious, automatic, associative, parallel and fast. The systems of the type 2 are more recent, conscious, sequential and slow, and are based on explicit rule following. In our opinion, there are good prima facie reasons to believe that, in human subjects, classification, a monotonic form of reasoning which is defined on semantic networks, and which is typical of DL systems, is a task of the type 2 (it is a difficult, slow, sequential task). On the contrary, exceptions play an important role in processes such as categorization and inheritance, which are more likely to be tasks of the type 1: they are fast, automatic, usually do not require particular conscious effort, and so on.

Therefore, a reasonable hypothesis is that a concept representation system should include different "modules": a monotonic module of type 2, involved in

[7] For any concept *C*, **T**(*C*) are the instances of *C* that are considered as "typical" or "normal".

[8] The K operator could be encoded in RDF/XML syntax of OWL as property or as annotation property.

classification and in similar "difficult" tasks, and a non-monotonic module involved in the management of exceptions. This last module should be a "weak" non monotonic system, able to perform only some simple forms of non monotonic inferences (mainly related to categorization and to exceptions inheritance). This solution goes in the direction of a "dual" representation of concepts within the ontologies, and the realization of hybrid reasoning systems (monotonic and non monotonic) on semantic network knowledge bases.

5.2 A "Pseudo-Fodorian" proposal

As seen before (section 2.2), according to Fodor, concepts cannot be prototypical representations, since concepts must be compositional, and prototypes do not compose. On the other hand, in virtue of the criticisms to "classical" theory, concepts cannot be definitions. Therefore, Fodor argues that (most) concepts are atoms, i.e., are symbols with no internal structure. Their content is determined by their relation to the world, and not by their internal structure and/or by their relations with other concepts (Fodor 1987, 1998). Of course, Fodor acknowledges the existence of prototypical effects. However, he claims that prototypical representations are not part of concepts. Prototypical representations allow to individuate the reference of concepts, but they must not be identified with concepts. Consider for example the concept DOG. Of course, in our minds there is some prototypical representation associated to DOG (e.g., that dogs usually have fur, that they typically bark, and so on). But this representation does not the coincide with the concept DOG: DOG is an atomic, unstructured symbol.

We borrow from Fodor the hypothesis that compositional representations and prototypical effects are demanded to different components of the representational architecture. We assume that there is a compositional component of representations, which admits no exceptions and exhibits no prototypical effects, and which can be represented, for example, in the terms of some classical DL knowledge base. In addition, a prototypical representation of categories is responsible for such processes as categorisation, but it does not affect the inferential behaviour of the compositional component.

It must be noted that our present proposal is not entirely "Fodorian", at least in the following three senses:

i. We leave aside the problem of the nature of semantic content of conceptual representations. Fodor endorses a causal, informational theory of meaning, according to which the content of concepts is constituted by some nomic mind-world relation. We are in no way committed with such an account of semantic content. (In any case, the philosophical problem of the nature of the intentional content of representations is largely irrelevant to our present purposes).

ii. Fodor claims that concepts are compositional, and that prototypical representations, in being not compositional, cannot be concepts. We do not take position on which part of the system we propose must be considered as truly "conceptual". Rather, in our opinion the notion of concept is spurious from the cognitive point of view. Both the compositional and the prototypical components contribute to the "conceptual behaviour" of the system (i.e., they have some role in those abilities that we usually describe in terms of possession of concepts).

iii. According to Fodor, the majority of concepts are atomic. In particular, he claims that almost all concepts that correspond to lexical entries have no structure. We maintain that many lexical concepts, even though not definable in the terms classical theory, should exhibit some form of structure, and that such structure can be represented, for example, by means of a DL taxonomy.

5.3 Prototypes and individuals

As we told before (section 2.2), within the field of psychology, different positions and theories on the nature of concepts are available. Usually, they are grouped in three main classes, namely prototype views, exemplar views and theory-theories (see e.g. Murphy 2002, Machery 2009). All of them are assumed to account for (some aspects of) prototypical effects in conceptualisation.

According to the prototype view, knowledge about categories is stored in terms of prototypes, i.e. in terms of some representation of the "best" instances of the category. For example, the concept CAT should coincide with a representation of a prototypical cat. In the simpler versions of this approach, prototypes are represented as (possibly weighted) lists of features.

According to the exemplar view, a given category is mentally represented as set of specific exemplars explicitly stored within memory: the mental representation of the concept CAT is the set of the representations of (some of) the cats we encountered during our lifetime.

Theory-theories approaches adopt some form of holistic point of view about concepts. According to some versions of the theory-theories, concepts are analogous to theoretical terms in a scientific theory. For example, the concept CAT is individuated by the role it plays in our mental theory of zoology. In other version of the approach, concepts themselves are identified with micro-theories of some sort. For example, the concept CAT should be identified with a mentally represented micro-theory about cats.

These approaches turned out to be not mutually exclusive. Rather, they seem to succeed in explaining different classes of cognitive phenomena, and many researchers hold that all of them are needed to explain psychological data. In this perspspective, we propose to integrate some of them in computational representations of concepts. More precisely, we try to combine a prototypical and an exemplar based representation in order to account for category representation and prototypical effects (for a similar, hybrid prototypical and exemplar based proposal, see Gagliardi 2008). We do not take into consideration the theory-theory approach, since it is in some sense more vaguely defined if compared the other two points of view. As a consequence, its computational treatment seems at present to be less feasible.

6. Concluding Remarks: Some Suggestion for Implementation

In the field of web ontology languages, the developments sketched above appear nowadays, technologically possible. Within the Semantic Web research community, in fact, the Linked Data perspective is assuming a prominent position (see Bizer, Heath and Berners-Lee 2009). According to this view, in recent years, one of the main objectives of the Semantic Web community regards the integration of different data representations (often stored in different data sources) within unique, semantically linked, representational frameworks. The main technical result coming from this integration is represented by the possibility of enlarging the answer-space of a query through the realization of “semantic bridges” between different pieces of data (and, often, data sources). Such integration is made possible through constructs provided by Semantic Web languages, such as OWL, SKOS etc.

Consider for example the opposition between exemplar and prototype theories (see sect. 5.3 above). Both theories can be implemented in a representation system using the Linked Data perspective.

Let us consider first the case of prototype theory. A “dual” representation of concepts and reasoning mechanisms appears to be possible trough the following approach: a concept is represented both in a formal ontology (based on a classical, compositional DL system), and in terms a prototypical representation, implemented using the Open Knowledge-Base Connectivity (OKBC) protocol[9]. The knowledge model of the OKBC protocol is supported and implemented in Protegé Frames, an ontology editor that supports the building of the so called Frame Ontologies. Since it is possible to export (without losing the prototypical information) the Frame Ontologies built with Protegé Frames in OWL language, the connection between these two types of representation can be done using the standard formalisms provided by the Semantic Web community within the linked data perspective (e.g. using the owl:sameAs construct)[10].

In a similar way, an exemplar based representation of a given concept can be expressed in a Linked Data format, and connected to a DL ontological representation.

In this way, according to our hypothesis, different types of reasoning processes (e.g., classification and categorization) can follow different paths. For example, classification could involve only the DL ontology, while the non monotonic categorization process could involve the component based on exemplars and prototypical information.

[9] http://www.ai.sri.com/~okbc/

[10] The only constraint is that, at the present state of the art, connecting OWL classes and Frames Ontology classes requires the use of OWL Full.

References

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P., 2003. *The Description Logic Handbook: Theory, Implementations and Applications*. Cambridge University Press.

Baader, F., Hollunder, B., 1995. *Embedding defaults into terminological knowledge representation formalisms.* J. Autom. Reasoning 14(1), 149–180.

Barsalou, L.W., 1985. Continuity of the conceptual system across species. *Trends in Cognitive Science*, 9(7), 305-311.

Bizer, C., Heath, T., Berners-Lee, T., 2009. "Linked Data - The Story So Far", *International Journal on Semantic Web and Information System*, 5 (3), pp. 1-22

Bobillo, F., Straccia, U., 2009. An OWL Ontology for Fuzzy OWL 2. Proceedings of the 18th *International Symposium on Methodologies for Intelligent Systems (ISMIS-09*). *Lecture Notes in Computer Science*. Springer Verlag.

Bonatti, P.A., Lutz, C., Wolter, F., 2006. *Description logics with circumscription*. In: Proc. of KR, pp. 400–410.

Brachman, R., 1985. I lied about the trees. *The AI Magazine*, 3(6), 80-95.

Brachman, R., Schmolze, J.G., 1985. An overview of the KL-ONE knowledge representation system, *Cognitive Science*, 9, 171-216.

Brachman, R., Schmolze, J.G. (eds.), 1985. *Readings in Knowledge Representation*, Los Altos, CA: Morgan Kaufmann.

Brandom, R., 1994. *Making it Explicit*. Cambridge, MA: Harvard University Press.

Calegari, S., Ciucci, D., 2007. Fuzzy Ontology, Fuzzy Description Logics and Fuzzy-OWL. *Proc. WILF 2007*, Vol. 4578 of LNCS.

Da Costa P.C.G., Laskey, K.B., 2006. PR-OWL: A framework for probabilistic ontologies. *Proc. FOIS-2006*, 237–249.

Dell’Anna, A., Frixione, M., 2010. On the advantage (if any) and disadvantage of the conceptual/nonconceptual distinction for cognitive science. *Minds & Machines*, 20, 29-45.

Ding, Z., Peng, Y., Pan, R.. 2006. BayesOWL: Uncertainty modeling in Semantic Web ontologies. In Z. Ma (ed.), *Soft Computing in Ontologies and Semantic Web*, vol. 204 of *Studies in Fuzziness and Soft Computing*, Springer.

Donini, F.M., Lenzerini, M., Nardi, D., Nutt, W., Schaerf, A., 1998. An epistemic operator for description logics. *Artificial Intelligence*, 100(1-2), 225–274.

Donini, F.M., Nardi, D., Rosati, R., 2002. Description logics of minimal knowledge and negation as failure. *ACM Trans. Comput. Log.*, 3(2), 177–225.

Evans, J.S.B.T., Frankish, K. (eds.), 2008. *In Two Minds: Dual Processes and Beyond*. New York, NY: Oxford UP.

Fodor, J., 1981. The present status of the innateness controversy. In J. Fodor, *Representations*, Cambridge, MA: MIT Press.

Fodor, J., 1987,l *Psychosemantics*, Cambridge, MA: The MIT Press/A Bradford Book.

Fodor, J., 1998, *Concepts: Where Cognitive Science Went Wrong*, Oxford, UK: Oxford University Press.

Fodor, J., Pylyshyn, Z., 1988. Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-71.

Frixione, M., 2007. Do concepts exist? A naturalistic point of view. In C. Penco, M. Beaney, M. Vignolo (eds.). *Explaining the Mental*. Cambridge, UK: Cambridge Scholars Publishing.

Gagliardi, F. 2008, A Prototype-Exemplars Hybrid Cognitive Model of “Phenomenon of Typicality” in Categorization: A Case Study in Biological Classification. in *Proc. 30th Annual Conf. of the Cognitive Science Society*, Austin, TX, pp. 1176–1181.

Gao, M., Liu, C., 2005. Extending OWL by fuzzy Description Logic. Proc. 17th *IEEE Int. Conf. on Tools with Artificial Intelligence (ICTAI 2005)*, 562–567. IEEE Computer Society, Los Alamitos.

Giordano, L., Ghiozzi, V., Olivetti, N., Pozzato, G., 2007. Preferential Description Logics. Logic for Programming, In *Artificial Intelligence, and Reasoning, LNCS*, Springer Verlag.

Hayes, P., 2001. Dialogue on rdf-logic. Why must the web be monotonic?. World Wide Web Consortium (W3C). Link: http://lists.w3.org/Archives/public/www-rdf-logic/2001Jul/0067.html

Katz, Y., Parsia, B., 2005. Towards a non monotonic extension to OWL. *Proc. OWL Experiences and Directions*, Galway, November 11-12.

Klinov, P., Parsia, B., 2008. Optimization and evaluation of reasoning in probabilistic description logic: Towards a systematic approach. In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T., Thirunarayan, K. (eds.) *ISWC 2008*. LNCS, vol. 5318, 213–228. Springer, Heidelberg.

Lukasiewicz, L., Straccia, U., 2008. Managing uncertainty and vagueness in description logics for the Semantic Web**.** *Journal of Web Semantics*, 6, 291-308.

Machery, E., 2005. Concepts are not a natural kind. *Philosophy of Science*, 72, 444–467.

Machery, E., 2009. *Doing without Concepts*. Oxford, UK: Oxford University Press.

Minsky, M., 1975. A framework for representing knowledge, in Patrick Winston (a cura di), *The Psychology of Computer Vision*, New York, McGraw-Hill. Also in Brachman & Levesque (2005).

Murphy, G.L., 2002. *The Big Book of Concepts*. Cambridge, MA: The MIT Press.

Osherson, D.N., Smith, E.E., 1981. On the adequacy of prototype theory as a theory of concepts. *Cognition*, 11, 237-262.

Peacocke, C., 1992. *A Study of Concepts*. Cambridge, MA: The MIT Press.

Rosch, E., 1975. Cognitive representation of semantic categories. *Journal of Experimental Psychology*, 104, 573-605.

Spelke, E.S., 1994. Initial knowledge: six suggestions. *Cognition*, 50, 431-445.

Spelke, E.S., Kinzler, K.D., 2007. Core knowledge. *Developmental Science*, 10(1), 89–96.

Stanovich, K.E., West, R., 2000. Individual Differences in Reasoning: Implications for the Rationality Debate?. *The Behavioural and Brain Sciences* 23 (5), 645- 665.

Stoilos, G., Stamou, G., Tzouvaras, V., Pan, J.Z., Horrocks, I., 2005. Fuzzy OWL: Uncertainty and the Semantic Web. *Proc. Workshop on OWL: Experience and Directions (OWLED 2005)*. CEUR Workshop Proceedings, vol. 188.

Straccia, U., 1993. Default inheritance reasoning in hybrid kl-one-style logics, *Proc. IJCAI*, 676–681.

Wittgenstein, L., 1953. *Philosophische Untersuchungen*. Oxford, Blackwell.