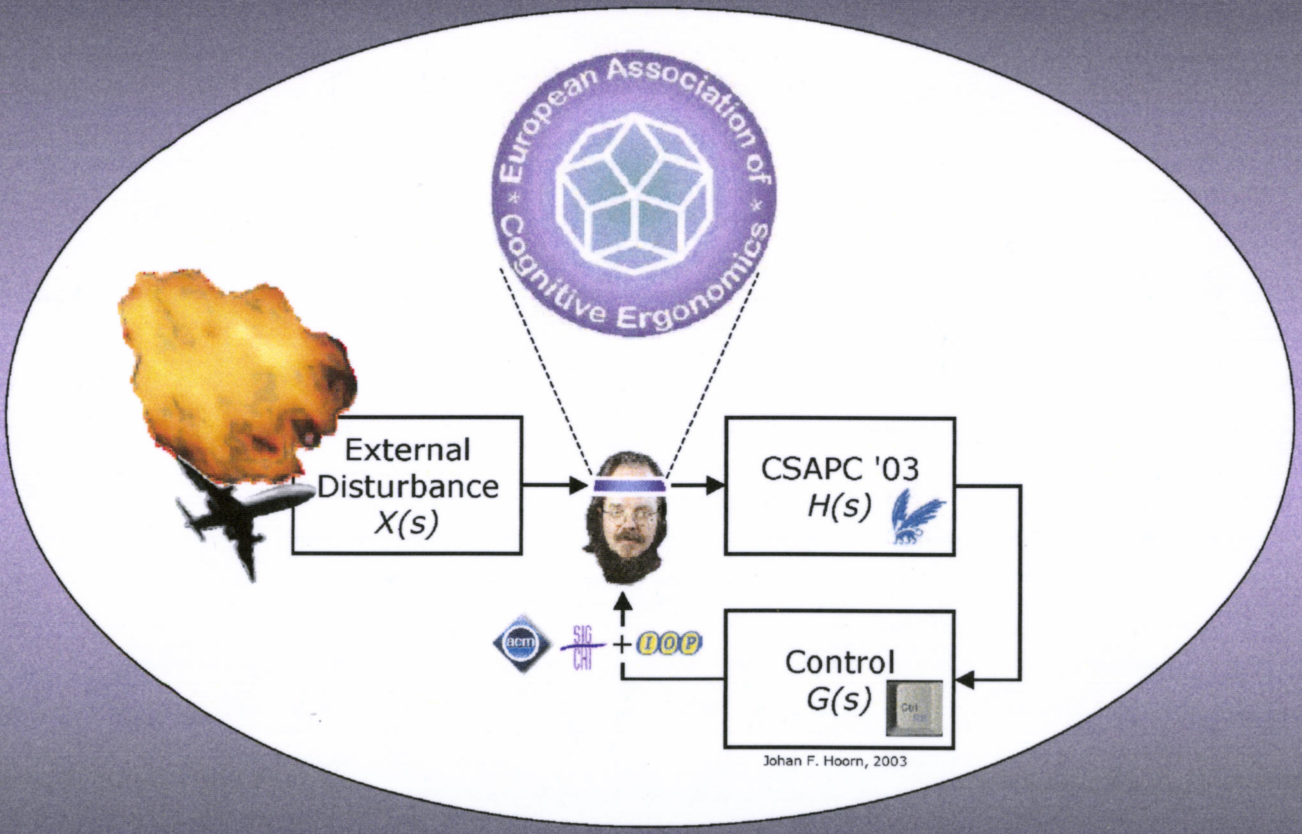
9th European Conference on Cognitive Science Approaches to Process Control



External Disturbance
$X(s)$

CSAPC '03
$H(s)$

Control
$G(s)$

Johan F. Hoorn, 2003

Cognition and Collaboration - Distributed Cognition in Complex Processes

CSAPC '03
9th European Conference on Cognitive Science Approaches to Process Control

Cognition and Collaboration – Distributed Cognition in Complex Processes

16-19 September 2003
Vrije Universiteit, Amsterdam, The Netherlands

Amsterdam, 2003

Proceedings

Gerrit C. van der Veer, Johan F. Hoorn, Editors

An EACE conference Series
European Association of Cognitive Ergonomics

ISBN 90-9017341-2

Printed by the Vrije Universiteit, Amsterdam, The Netherlands

EACE
European Association of Cognitive Ergonomics
Unite de Recherche Inria Rocquencourt
78153 Rocquencourt, Le Chesnay Cedex
France
www.eace.info

Foreword

This volume contains the text of the papers presented at CSAPC'03, the 9th European Conference on Cognitive Science Approaches to Process Control. This conference took place at the Vrije Universiteit, Amsterdam, The Netherlands, September 16-19, 2003.

CSAPC is one of the two conference series organised by EACE (the European Association of Cognitive Ergonomics). It was initiated in 1987 near Paris and held in Denmark, Finland, France, Italy, United Kingdom and Germany. It is a European and biennial conference series aiming at bringing together well-known experts in cognitive psychology and ergonomics, human-machine systems, and artificial intelligence, to discuss their multidisciplinary studies dealing with design and evaluation of complex, dynamic, and risky human-machine systems. The stakes of this kind of research are major in a number of application domains, such as industrial process control (e.g.: nuclear power plant or process industry), aviation (e.g.: cockpit or air traffic control), ship navigation, car driving (e.g.: advanced computer support), medicine (e.g.: anaesthesiology in operation room), etc. Researchers and practitioners studying dynamic situations are used to attend conferences or congress sessions specific to each application domains and each scientific discipline separately. However, common research topics and practical problems spring from these application domains and need to be addressed at a certain level of generality for reaching appropriate solutions in cognitive ergonomics and engineering. This search for similarity and differences across application domains is one of the main objectives of CSAPC.

The theme of CSAPC'03 was related to the increasing possibility to develop "intelligence" and "knowledge" in the systems to be designed. Human operators and other stakeholders will increasingly find themselves in a situation where cognition is distributed, both between human stakeholders, between humans and information technology, and between all these and relevant representations that feature in the work situation. Emerging questions are about the risks of distributed decision making, distributed support for decisions, and the application of multiple types of representations, especially in time- and safety-critical situations. These questions ask for a vision on how to safeguard consistency, mutual understanding, and knowledge management. Ontologies can be expected to increase in importance, as will be tools for managing complexity in unexpected situations and under time constraints.

The presentations during the conference have been structured in five groups:

1. Situational Awareness
2. Error management
3. Human Analytical Tools
4. Technical Analytical Tools and Simulations
5. Process Control Requirements and Collaborative Work: Theory and Method

This conference could not have been implemented without the support our parent-organisation EACE. The Organisation Committee and the International Programme Committee conducted the reviewing process and selected the programme. ACM – SIGCHI collaborated with us in spreading the information and accepted to include the proceedings in the ACM Digital Library. The Vrije Universiteit Amsterdam hosted and sponsored the conference. IOP-MMI, the Dutch innovation-driven research programme of the Ministry of Economic Affairs – section Human-Machine Interaction, supported the social programme, as did the Municipality of Amsterdam. Last but not least, Elly Lammers was indispensable for providing the organisation, the logistics, the financial management, and the communication support for this project.

Gerrit C. van der Veer
Johan F. Hoorn

9th European Conference on
Cognitive Science Approaches to Process Control (CSAPC'03)
Cognition and Collaboration - Distributed Cognition in Complex Processes

16-19 September 2003 - Amsterdam, The Netherlands

Organisation Committee:

Gerrit C. van der Veer (chairman) (Vrije Universiteit, NL)
Erik Hollnagel (University of Linkoping, S)
Jean-Michel Hoc (IRCCyN, F)
Antonio Rizzo (Università degli studi di Siena, I)
Johan F. Hoorn (Vrije Universiteit, NL)
Elly Lammers (Vrije Universiteit, NL)

International Programme Committee:

Sebastiano Bagnara (Politecnico di Milano, I)
Pietro Cacciabue (EC, Joint Research Centre, I)
José J. Cañas (Universidad de Granada, ES)
Bernhard Döring (FKIE, G)
Gunnar Johannsen (University of Kassel, G)
Geert de Haan (Maastricht McLuhan Institute, NL)
Furuta Kazuo (University of Tokyo, J)
Karl-Friedrich Kraiss (RWTH Aachen, G)
Leena Norros (VTT, SF)
Patrick Millot (Université de Valenciennes et du Hainaut Cambresis, F)
John Patrick (Cardiff University, UK)
Bernard Pavard (CNRS, F)
Matthias Rauterberg (Technical University Eindhoven, NL)
Tjerk W. van der Schaaf (Eindhoven Uinversity of Technology, NL)
Oliviero Stock (ITC-IRST, I)
Ola Svenson (Stockholm University, S)
Robert M. Taylor (DSTL, UK)
Bernhard Zimolong (Ruhr University Bochum, F)

CSAPC 2003

CONTENTS

Session 1 Situational Awareness

Session 2 Error Management

Session 3 Human Analytical Tools

Session 4 Technical Analytical Tools and Simulation

Session 5 Process Control Requirements and Collaborative Work: Theory and Method

Session 1 Situational Awareness

Object-Oriented Modelling of Situational Awareness Information for Knowledge-Based User Interfaces

Bernhard Döring Gert Dörfel Helmut Distelmaier
Research Institute for Communication, Information Processing and Ergonomics (FKIE)
Research Establishment for Applied Science (FGAN)
Neuenahrer Str. 20, D-53343 Wachtberg, Germany
doering@fgan.de doerfel@fgan.de distelmaier@fgan.de

ABSTRACT

In nowadays complex and dynamic changing military environments supporting situation awareness (SA) of operators is a prerequisite for situation- and task-adequate decision making. SA refers to information of the operator environment on three levels which represent relevant elements, element patterns describing complex mission situations, and projections of future states and dynamics of elements. A means for supporting SA of operators are adaptive knowledge-based user interfaces. For developing such interfaces information of the three different SA levels which operators need in performing their tasks have to be specified and modelled. One source from which that information can be acquired are scenarios which have to be developed in any case for system design as well as for operator training. For specifying relevant SA information a model of the problem domain has been developed which comprises the True World of scenarios, the Sensed World of detected tracks, and the Deduced World of concluded track information. To uniformly describe these different worlds an object oriented approach has been applied which is based on static and dynamic scenario and track objects which are specified mathematically. Attributes and operations of track objects constitute elements and patterns of relevant SA information to be identified. Additionally, the described mathematical model of track objects constitutes the basis for developing a software specification with the object-oriented Unified Modelling Language UML. Using a Navy anti-air warfare scenario as an example the application of the developed modelling approach is demonstrated in detail.

Keywords

situation awareness, decision support, knowledge acquisition and modelling, knowledge-based user interfaces, design requirements, UML.

1. INTRODUCTION

Making decisions by means of complex technical systems sets high standards to human operators involved, especially in natural settings. Natural settings are particularly specified by uncertainty, changing situation dynamics, time pressure, and ill defined goals where different goals might be competing or even contrary [10]. Complex technical systems provide a multitude of situational data which have to be perceived and interpreted by an operator through individual cognitive abilities like skill and knowledge in order to perform relevant decision making tasks. Managing complex natural situations like medical or plant emergencies as well as complex military situations by means of modern human-machine systems are examples for such tasks. In these situations human operators undergo high mental stress due to the need to respond quickly and accurately or face potentially fatal consequences. Human decision making in such situations is based on a large scale on situation awareness (SA) which is defined as the state of operator knowledge about the external environment resulting from situation perception and situation assessment [5].

Intelligent and adaptive knowledge-based user interfaces are considered to be a viable approach to overcome some of those difficulties decision makers are faced with when having to cope with natural situations. Such interfaces consist of a knowledge-based user assistant (KBUA) and an interactive graphical or multimedia user interface (Figure 1). The KBUA may support decision makers in performing information gathering and processing in all phases of a decision making cycle, i.e., in situation perception and assessment as well as action selection, planning, and execution.

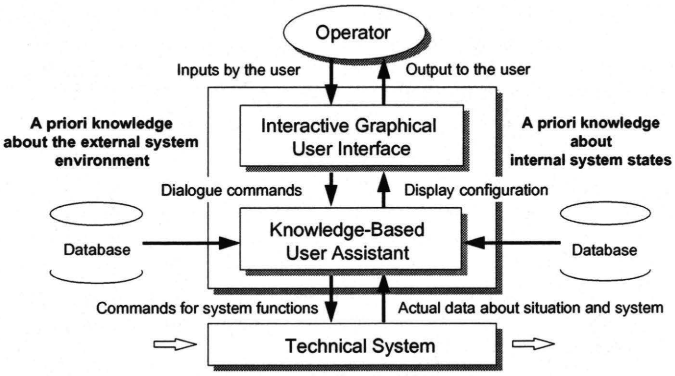


Figure 1: Concept of a knowledge-based user interface

The basic idea of this concept is that an overall automation must not be the objective of system development. The human operator should be involved in the decision making process as far as his abilities and his performance are sufficient. An aid is provided only to exploit human abilities (e.g., in detecting, evaluating, and reacting on complex and critical situations) and to overcome human deficiencies (e.g., when doing mathematical calculations), i.e., to complement individual human performances. For that, actual data about the external system environment and internal system components are provided by the technical system. The KBUA processes these data and presents them to the operator on the interface adaptively depending on, e.g., the mission segment, the situation, and/or the actual user task. For processing these actual data the assistant additionally uses a priori knowledge about possible external situations and internal system states. The human operator is engaged in a co-operative process in which human and KBUA both initiate communication, monitor events, and perform tasks. The KBUA does not act as an interface or layer between the operator and the system. Rather, in parallel to the human operator, the assistant monitors the external as well as the internal situation and, additionally, operator actions. It is called "knowledge-based" because it applies knowledge that the operator normally acquires by learning, training, and experience. If the KBUA encounters critical situations or inappropriate operator behaviour, it may automatically perform some operator-related functions. Faulty behaviour of the operator will be classified, announced, and possibly compensated by the assistant. But in any case the human operator is able to bypass the assistant, so that the responsibility and ultimate decision resides with the operator. Examples of such support systems are described in, e.g., [3], [7], [15], [16].

The overriding problem of designing a knowledge-based operator support still remains eliciting, structuring, formalising, and implementing the knowledge about the complex natural situation as well as about the situation- and mission-relevant decisions. Thus, for decision making in natural settings by means of knowledge-based systems, different types of knowledge have to be considered. These types are: a) actual situational knowledge about a specific problem domain, b) a priori knowledge, i.e., already existing knowledge about comparable problem situations. This presentation is especially focussed on the elicitation, structuring, and formalisation of a priory knowledge about situations of the external system environment. As an example a Navy anti-air warfare situation is used.

2. DECISION MAKING IN COMPLEX SITUATIONS

For modelling decision making in complex situations several approaches have been developed in the past. The very basis model of Wickens [20] discriminates between two steps in decision making. The first step represents a diagnosis or prediction task which an operator has to perform to gather information, generate hypotheses about situations in his environment and to select the most probable one; the second step is a choice task with which the operator determines applicable actions to affect the perceived situation as desired, selects most appropriate actions and executes them. Due to the weaknesses of human decision makers in information gathering, hypothesis generation and selection, as well as in action remembering and selection these activities have to be supported in complex task environments.

To develop a decision support design requirements have to be specified at first. Several modelling approaches have been developed for specifying decision requirements of distinct application areas. For identifying decision requirements for military force management Wohl [22] developed the SHOR-model which defined four elements of a decision process: **S**timulus (data), **H**ypothesis (perception alternatives), **O**ption (response alternatives) and **R**esponse (action). These elements have been related to decision activities like gather, detect, filter, correlate (S), create, evaluate, select (H), create, evaluate, select (O), and plan, organise, execute (R). Similar activities are also contained in the OODA-loop which constitutes the basis of the doctrine publication of the US-Department of the Navy [1]. Loop activities are **O**bserve, **O**rient, **D**ecide, and **A**ct. The loop illustrates the continuous, cyclical process of command and control by which a commander makes decisions and exercises authority over subordinate commanders in accomplishing an assigned mission. On the basis of verbal protocols of experienced power plant operators Rasmussen [11] developed a template, called "decision ladder", for describing operator decision making in the complex environment of process control. The ladder contains information processing activities and resulting states of knowledge in two legs, one upward for the analysis of an actual situation corresponding to the diagnosis task, another downward for planning the proper actions corresponding to the choice task. Between the two legs specific shortcuts are possible. They represent the behaviour of experienced operators. Vicente [18] used this ladder in his cognitive work analysis as a basis for identifying control task requirements for a particular work domain. In contrary to normative models which prescribe how a system should behave and descriptive models which describe how a system actually behaves in practice he entitles this ladder as a formative approach to specify requirements that must be satisfied so that the system could behave in a new, desired way.

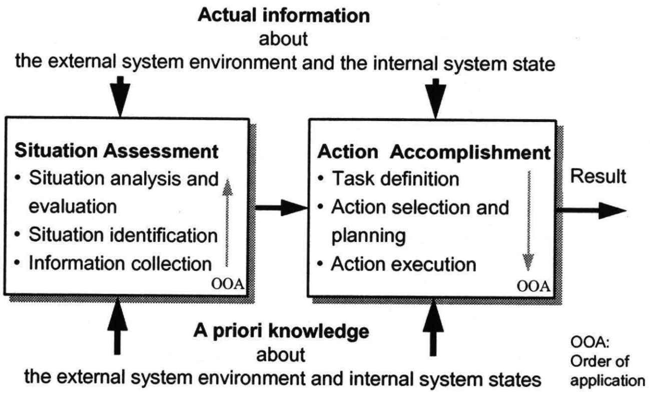


Figure 2: Structure of human decision activities

To determine decision requirements for a Navy command and control (C2) system Rasmussen`s ladder and the two steps of decision making defined in the modelling approach of Wickens [20] have been combined arranging information processing activities correspondingly (Figure 2). Again, the first step represents a diagnosis or prediction task called "situation assessment" comprising activities like information collection, situation identification, and situation analysis and evaluation. These activities are processed in an upward direction. That means, a decision maker first collects stimuli from a number of information sources in the environment. Such sources can be system components with their actual status as well as other relevant systems in the surroundings and their actual status. The collected information is frequently covered by a veil of uncertainty. Sampling and integrating this information a decision maker will usually attempt to formulate a hypothesis about the true situation and use it as a basis for the following choice of actions. If the hypothesis concerns the actual state of affairs information processing represents a diagnosis task, if the hypothesis relates to a future state of the environment a prediction task. In any case, the hypothesising process involves an interaction between long-term and short-term memory [20]. In the long-term memory plausible hypotheses about possible situations of the external system environment and possible states of internal system components are stored after they have been learned. In figure 2 these hypotheses are called a priori knowledge because in decision making it has to be known in advance. In the short-term memory alternative hypotheses are considered, compared, and evaluated against the situation provided by the actual information. The final diagnosis may not be absolute. Instead, it may be a an expression of degree of belief that one hypothesis rather than another is true, but is usually followed by the second step which represents the choice task called "action accomplishment" (Figure 2). With this task downward directed activities are performed like task definition, action selection and planning, and action execution. If according to the actual situation identified a new goal state has been determined a new task will be defined firstly. Then, for this task applicable actions will be identified depending on the available system resources. Finally, the most appropriate actions will be selected. This choice requires the consideration of costs and benefits. Therefor, choice usually involves the evaluation of risk when there is uncertainty about the state of the world that will affect consequences of the choice. Possible actions and the related risk acquired during the learning and/or training phase are stored in the long-term memory of the decision maker [20]. Finally, the execution of selected actions will be planned and carried out.

3. SITUATION AWARENESS IN DECISION MAKING

The diagnosis task "situation assessment" (Figure 2) plays a crucial role in human decision making because it results in the SA of the decision maker and therefore creates the basis for the following selection and planning of actions. SA is the constantly evolving degree of accuracy by which operator´s perception and assessment of the external environment reflects reality. Therefore, SA is the prerequisite for situation- and task-adequate decision making in complex dynamic situations. As being the result of a reliable assessment of the situation in which, e.g., a ship operates, SA is vital for the successful completion of its mission. Therefore, intending to support military operators in decision making in any case and first of all SA has to be supported. According to Endsley [5], operator´s SA is the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their state in the near future. To specify required information Endsley distinguishes three levels of SA: 1.) Perception of the information elements in the environment, i.e., the states, attributes, and dynamics of relevant elements in the operator environment. 2.) Comprehension of the current situation by information processing based on a synthesis of disjoint level 1 elements by putting them together to perform patterns for getting a holistic picture of the environment and an assessment of the current state. 3.) Projection of future states on the basis of actions of elements in the environment. This is achieved through knowledge of the states and dynamics of the elements and comprehension of the situation for both level 1 and level 2.

To establish SA, information about the situation in the environment from sensors or other sources of significant data and conditions must be gathered and processed. This information concerns the external system environment as well as the internal operational state(s) of own combat (sub)system(s) involved. But as already mentioned, not only information about the actual situation is necessary for accomplishing the SA process. For level 2 and level 3 of SA which contain assessment steps pre-existing relevant “a priori” information is necessary to relate actual information elements to already known situational information and patterns. SA occurs as a consequence of integrating a priori information together with actual information by cognitive processing skills that include attention allocation, perception, data extraction, comprehension, and projection [14]. In order to provide a reliable information basis for carrying out missions it will be necessary to assess and reassess the situation on a continuous basis. Therefore, to not only establish but also to maintain SA the continuous extraction of information about a dynamic system and/or environment, the integration of this information with previously acquired knowledge to form a coherent mental picture, and the use of that picture in directing further perception of, anticipation of, and attention to future events is necessary [21].

4. SCENARIO-BASED KNOWLEDGE ACQUISITION

This depiction is focussed especially on the specification of decision requirements for supporting SA of decision makers about the external system environment of a Navy C2 system. Therefore, the acquisition and modelling process of the a priori knowledge about the external system environment stands in the centre of this consideration. One source from which especially information about the external system environment can be acquired are scenarios. Scenario generation and analysis are the first steps in modern software development processes for building solutions for large, complex problems. According to McGraw and Harbison [9] scenarios should comprise the following components:

- Goals and critical success factors.
- Physical (i.e., topology, layout) and logical (i.e., operation circumstances) context.
- Major events or activities that comprise the scenario.
- Performer or participants involved and the events in which they are involved .
- Information and recourses used, including information, products , etc. throughout the scenario.
- Points at which decisions are made, constraints considered, and rules applied.
- Performance problems and opportunities for enhancement.

In a recent study a knowledge-based user interface has been developed for supporting Navy operators in identifying air targets in a surveillance mission [2]. The scenario used in this study has been developed by Navy experts for training operators to perform that task. It includes ownship with different safety and engagement zones, an airway, a transition corridor, a land area with coastal line, neutral and friendly air targets with normal behaviour, and suspect air targets with dubious behaviour. The scenario describes graphically not only a snapshot of a dynamic situation but rather a combination of different static scenes of a dynamically evolving situation with air tracks to be identified. It includes the following situations:

- An approach of two air targets with suspicious behaviour.
- An approach of two friendly air targets which identify themselves by executing a predefined flight pattern.
- An approach of a friendly air target in a transition corridor.
- Neutral air targets flying in an airway.
- An approach of a suspicious air target from inside the airway with a final harassment manoeuvre.

5. MODEL OF THE PROBLEM DOMAIN

To acquire the information necessary for getting SA about the external system environment during the identification of air targets in a surveillance mission a model of the scenario problem domain has been developed. The model comprises three worlds which represent three different views (Figure 3): 1.) The True World stands for the real mission environment of ownship respectively for the developed scenario which represents a model of that environment. 2.) The Sensed World describes the information sphere of track objects inside of ownship acquired to a large extent by ownship sensors. 3.) The Deduced World represents the sphere of deduced information concluded from the Sensed World by means of inference processes. To uniformly describe these different views an object-oriented approach has been applied. This approach also supports an object-oriented problem analysis and facilitates the implementation with modern object-oriented programming languages when later developing a decision support system. To accomplish this approach static and dynamic objects are identified as relevant model elements. These objects can be specified by means of attributes which describe characteristics and states of an object and by operations which characterise its behaviour [12].

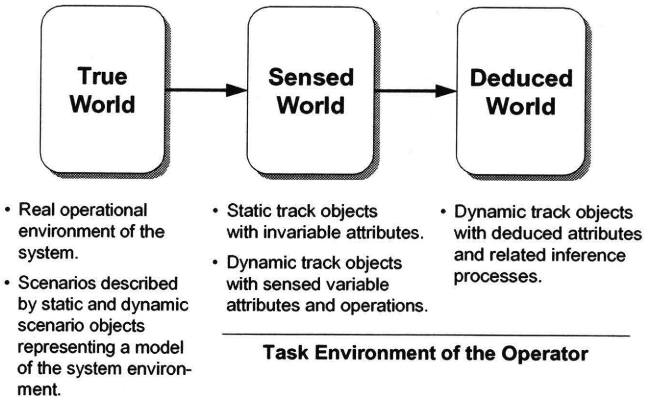


Figure 3: Model of the problem domain

5.1 The True World

Descriptive elements of the True World which is defined by the used scenario are static and dynamic scenario objects. Static scenario objects posses only one state and no operation. Examples of such objects which constitute the static mission environment of the scenario are, for instance, airways, transition corridors, and coastal lines. Dynamic scenario objects correspond to active air, surface, and land targets of the scenario. They are specified by changing attribute values, that means, changing object states caused by object operations. To describe scenario objects mathematically formalisms of the general system theory, e.g., [6], have been applied. Generally, the analysis starts with determining the set SO of all scenario objects so_i. With the index set I_{SO} it counts:

$$SO = \{ so_i : i \in I_{SO} \} . \qquad (1)$$

A scenario object so_i can be described by the set so_i_OP of its operations $so_i\_op_k$, the set so_i_ATT of its attributes $so_i\_att_k$, the value set $so_i\_ATV_k$ of all values $so_i\_att_k(t)$ of an attribute $so_i\_att_k$, with the time set T, the set of the positive real numbers R+, and $t \in T \subset R+ \cup \{ 0 \}$, and the set so_i_S of object states $so_i\_s_k$. Details of the formal description can be found in [4].

As an example, the airway identified in the above mentioned Navy scenario is considered. It represents the static scenario object so_1 which is characterised by the its attributes $so_1\_att_k$ and their related values $so_1\_att_k(t)$. Because o_1 represents a static scenario object there exists no operation and all related attribute values are constant:

$so_1\_object\_identifier(t)$ = airway number ,
$so_1\_reference\_point(t) =$
$(posX(t): a\ [°], posY(t): b\ [°], posZ(t): c\ [ft])$,
$so_1\_width(t) = d\ [nm]$,
$so_1\_length(t) = e\ [nm]$,
$so_1\_height(t) = f\ [ft]$, (2)
$so_1\_direction(t) = g\ [°]$,
$so_1\_region(t) = function\ (so_1\_reference\_point(t), so_1\_width(t), so_1\_length(t), so_1\_height(t), so_1\_direction(t))$,
$so_1\_speed(t) = h\ [kn]$,
$so_1\_flight\_level(t) = (i, j)\ [ft]$.

As another example of the scenario an aircraft inside the airway is considered. It appears to be the dynamic scenario object so_4 which can be specified by attributes, their values, and operations. Some attribute values $so_4\_att_k(t)$ and operations $so_4\_op_k$ are:

$so_4\_aircraft\_identifier(t)$ = registration number ,
$so_4\_position(t) = (posX(t): x\ [°], posY(t): y\ [°], posZ(t) : z\ [ft])$,
$so_4\_altitude(t) = z\ [ft]$,
$so_4\_altitude\_change(t) = 0\ [ft/min]$,
$so_4\_course(t) = g\ [°]$,
$so_4\_course\_change(t) = 0\ [°/sec]$,
$so_4\_speed(t) = h\ [kn]$,
$so_4\_speed\_change(t) = 0\ [kn/min]$, (3)
$so_4\_IFF\_signal(t)$ = Mode 3 ,
$so_4\_emitter(t)$ = Radar R3 ,
$so_4\_role(t) = commercial\_airliner$,
$so_4\_activity(t) = fly\_in\_accordance\_with\_airway$,
$so_4\_identity(t) = neutral$.

$so_4\_op_1$ = IF creation event THEN create object ,
$so_4\_op_2$ = IF extinction event THEN delete object,
$so_4\_op_3$ = IF state change event THEN change state ELSE retain state.

The dynamic scenario object o_4 may be in the actual state $so_4\_s_1 = flying\_inside\_the\_airway$ which is defined by the above attributes. As soon as there occurs an event, for instance, if the value $so_4\_course\_change(t) > x\ [°/sec]$ indicates a course change then the activity takes the value $so_4\_att_2(t) = fly\_not\_in\_accordance\_with\_airway$. The state remains the same.

The dynamic processes of a scenario with all state changes of scenario objects can be simulated and in this way be accessible to an analysis. One possible simulation tool is, e.g., the commercial product STAGE (Scenario Toolkit And Generation Environment). It is a real-time, reconfigurable, extendible simulation framework for military applications [19].

5.2 The Sensed World

The second part of the model constitutes the Sensed World (Figure 3) which describes the information sphere of track objects acquired to a large extent by ownship sensors. Corresponding to scenario objects of the True World there are again static and dynamic track objects in the Sensed World with the same meaning as the scenario objects. Static track objects in the Sensed World correspond for the most part to static scenario objects and, therefore, are known in advance either from the scenario or other geographical data sources like nautical charts. Examples of such static track objects which posses only one state and no operation are again airways, transition corridors, and coastal lines constituting the static mission environment. But there may be certain static track objects which ship sensors may detect, e.g., the wreck of a recently sunk ship which may be detected by the ship sonar but not being registered yet in the corresponding nautical chart. All static objects are stored onboard ownship, for instance, in the central data store of the ship which may contain also a geographical database with nautical chart information. Dynamic track objects correspond again to dynamic scenario objects. They represent active air, surface, and land tracks detected by ownship sensors and stored with attributes and operations in the central data store. That means, that dynamic scenario objects of the True World are transformed into track objects of the Sensed World by considering ship sensor characteristics. Attributes and their values are updated if sensors provide new data. As attributes of dynamic track objects depend on the available sensors on board, only that information can be sensed for which sensors are available.

For instance, if there is a 3D-radar available then the altitude of an air track can be determined as track attribute. If the ship has only a 2D-radar then the altitude cannot be assessed. But there are also track attributes which can be determined from sensed attributes by calculation, e.g., the vertical speed of an air track from the change of its altitude. These attributes are also considered as sensed attributes. With dynamic track objects operations specify processes like creating, updating, and deleting those objects in the data store of the ship.

With TO as the set of all track objects to$_i$ again such an object can be described formally by a set to$_i$ _OP of its operations to$_i$_op$_k$, a set to$_i$ _ATT of its attributes to$_i$_att$_k$, a value set to$_i$_ATV$_k$ of all values to$_i$_att$_k$(t) of attribute to$_i$_att$_k$, with the time set T, and $t \in T$, a set to$_i$_ATV of all value sets to$_i$_ATV$_k$, and a set to$_i$_S of all states to$_i$_s$_k$ of object to$_i$. Details of the formal description can be found in [4].

As an example of a static track object the above mentioned airway which represents in the True World the static scenario object so$_1$ is considered. In the Sensed World it constitutes the static track object to$_1$ with the same attributes and values as so$_1$ specified in equation (2). As example of a dynamic track object the above mentioned aircraft is regarded. In the True World this aircraft has been represented by the dynamic scenario object so$_4$. If this object is in ownship sensor range it will be detected and a dynamic track object, e.g., to$_7$ will be created in the ship´s central data store which represents the Sensed World. Object attributes and values depend on the sensor observation time. It is assumed that attributes like altitude, speed, and course and their alteration can be determined. Assuming sufficient observation time for reaching a stable state values to$_7$_att$_k$(t) of some sensed attributes to$_7$_att$_k$ and operations to$_7$_op$_k$ are listed below:

to$_7$_track_identifier(t) = track number ,
to$_7$_position = (posX(t): x [°] , posY(t): y [°] , posZ(t): z [ft]) ,
to$_7$_altitude(t) $\approx$ z [ft] ,
to$_7$_altitude_change(t) $\approx$ 0 [ft/min] ,
to$_7$_course(t) $\approx$ g [°] , (4)
to$_7$_course_change(t) $\approx$ 0 [°/sec] ,
to$_7$_speed(t) $\approx$ h [kn] ,
to$_7$_speed_change(t) $\approx$ 0 [kn/min] ,
to$_7$_IFF_signal(t) = Mode 3 ,
to$_7$_emitter(t) = Radar R3 ,

to$_7$_op$_1$ = IF detection event THEN create object ,
to$_7$_op$_2$ = IF update event THEN update sensed attributes ,
to$_7$_op$_3$ = IF extinction event THEN delete object.

5.3 The Deduced World

The Deduced World is the third part of the developed model (Figure 3). It is represented by the deduced attributes of dynamic track objects, their values, and by the inference processes necessary for deducing those attributes. Taking again as an example the dynamic track object to$_7$ specified above the following additional deduced attributes to$_7$_d-att$_k$ and operations to$_7$_d-op$_k$ of the object to$_7$ arise in the Deduced World:

to$_7$_d-distance_between_objects,
to$_7$_d-activity,
to$_7$_d-role,
to$_7$_d-category,
to$_7$_d-class,
to$_7$_d-identity,
to$_7$_d-formation,
to$_7$_d-activity sequence,
to$_7$_d-application,
to$_7$_d-type,
to$_7$_d-option, (5)
to$_7$_d-threat.

to$_7$_d-op$_4$ = IF update event THEN update deduced attributes.

Generally, for a deduced attribute to$_i$_d-att$_m$ of track object to$_i$, its values to$_i$_d-att$_m$(t) and the object operation to$_i$_d-op$_k$ which specifies the update and inference process of this attribute it counts:

to$_i$_d-att$_m$ $\in$ to$_i$_ATT ,
to$_i$_d-att$_m$(t) $\in$ to$_i$_d-ATV$_m$ $\in$ to$_i$_ ATV , (6)
to$_i$_d-op$_k$ $\in$ to$_i$_OP .

A deduced attribute of a dynamic track object can be derived from sensed and other already deduced attributes of the same object. Additionally, different attributes of other static and dynamic track objects may be involved in the inference process as well. For modelling that interference process mathematical relations are applied. To deduce an attribute to$_i$_d-att$_m$ of a dynamic track object to$_i$ and its value set to$_i$_d-ATV$_m$ those other value sets to$_i$_ATV$_n$ and to$_i$_d-ATV$_n$ of to$_i$ which contribute to the inference process have to be selected. For describing this selection a set to$_i$_ATVsel which contains all contributing value sets to$_i$_ATV$_n$ and to$_i$_d-ATV$_n$ will be defined. Besides to$_i$ there may be other track objects to$_p$, to$_q$, ... with their attribute value sets to$_p$_ATV, to$_q$_ATV ... contributing as well for deducing the attribute to$_i$_d-att$_m$. The contributing value sets of those objects can be again specified by means of selected value sets to$_p$_ATVsel, to$_q$_ATVsel .... Then, to describe the inference process in detail an inference relation to$_i$_ir$_k$ is defined. With to$_i$_IR as the set of all to$_i$_ir$_k$ it follows:

to$_i$_ATVsel = { to$_i$_ATV$_n$, to$_i$_d-ATV$_{n+k}$: to$_i$_ATV$_n$,
to$_i$_d-ATV$_{n+k}$ $\in$ to$_i$_ATV $\wedge$ to$_i$_ATV$_n$,
to$_i$_d-att$_{n+k}$ are relevant to deduce to$_i$_d-att$_m$ } ,
to$_p$_ATVsel = { to$_p$_ATV$_u$, to$_p$_d-ATV$_{u+k}$: to$_p$_ATV$_u$,
to$_p$_d-ATV$_{u+k}$ $\in$ to$_p$_ATV $\wedge$ to$_p$_ATV$_u$,
to$_p$_d-att$_{u+k}$ are relevant to deduce to$_i$_d-att$_m$ } ,
to$_q$_ATVsel = { to$_q$_ATV$_v$, to$_q$_d-ATV$_{v+k}$: to$_q$_ATV$_v$,
to$_q$_d-ATV$_{v+k}$ $\in$ to$_q$_ATV $\wedge$ to$_q$_ATV$_v$,
to$_q$_d-att$_{v+k}$ are relevant to deduce to$_i$_d-att$_m$ } ,
to$_i$_ir$_k$ $\subset$ to$_i$_d-ATV$_m$ x X to$_i$_ATVsel x X to$_p$_ATVsel x X to$_q$_ATVsel x ....,
to$_i$_IR = { to$_i$_ir$_k$: $k \in I_{toi\_IR}$ } , (7)
IR = { to$_i$_IR : $i \in I_{TO}$ } .

The set to$_i$_IR containing all relations of a dynamic track object to$_i$ specifies the total inference process of that object. This process is part of the object operation to$_i$_d-op$_4$ = “IF update event THEN update deduced attributes” and will be activated if an update event occurs. Moreover, the total Deduced World is represented by the set of all deduced attributes and the set of all relations IR which comprises total inference processes of all considered dynamic track objects.

As an example, again the dynamic track object to$_7$ representing an aircraft is selected. To specify the inference process for deducing the attribute to$_7$_d-activity the value set to$_7$_d-ACTIVITY is interrelated with value sets of relevant sensed attributes to$_7$_(POSX, POSY, POSZ), to$_7$_ALITUDE, to$_7$_COURSE, and to$_7$_SPEED of object to$_7$ (see equation 4). (In the preceding and in the following all value sets of an attribute are named with capital letters.) Additionally, value sets of the static track object to$_1$ representing an airway have to be considered in this inference process. Airway attribute value sets are to$_1$_REGION, to$_1$_DIRECTION, to$_1$_SPEED, and to$_1$_FLIGHT_LEVEL (see equation 2). If the process is specified with equations (7) then the inference relation to$_7$_ir$_1$ results as follows:

$$\mathrm{to_7\_ATVsel} = \{\, \mathrm{to_7\_(POSX, POSY, POSZ), to_7\_ALTITUDE,} \\ \mathrm{to_7\_COURSE, to_7\_SPEED} \,\},$$

$$\mathrm{to_1\_ATVsel} = \{\, \mathrm{to_1\_REGION, to_1\_DIRECTION, to_1\_SPEED,} \\ \mathrm{to_1\_FLIGHT\_LEVEL} \,\}, \qquad (8)$$

$$\mathrm{to_7\_ir_1} \subset \mathrm{to_7\_d\text{-}ACTIVITY} \; x \; X \; \mathrm{to_7\_ATVsel} \; x \; X \; \mathrm{to_1\_ATVsel} \,.$$

For graphically representing this inference process interaction matrices introduced by Sage [13] can be applied. Figure 4 shows the example described with equations (8). In the upper part of the picture the static track object to_1 with its attribute value sets is displayed. The lower part shows attribute value sets of the dynamic track object of interest to_7. The arrows indicate the direction of the inference process. For describing the inference process in detail a table form can be used [4].

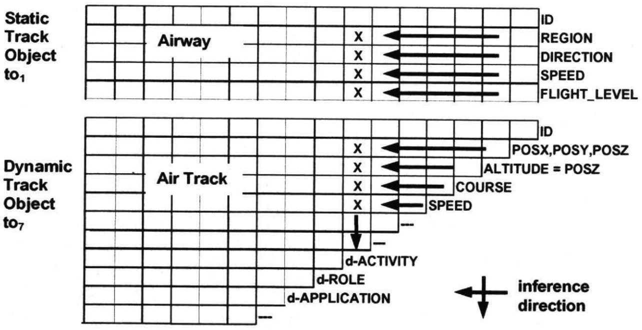


Figure 4: Example of the inference process for the deduced attribute to_7_d-activity

6. THE WORLD MODEL AND SA INFORMATION

To show the relevance of the world model relationships between the three worlds of the model (Figure 3) and the information levels of SA described in the beginning are considered. Because an operator in the combat information centre of a warship does not have a direct contact to the mission environment outside the ship it becomes obviously that there does not exist any direct relationship between the True World and the different SA level information. The True World represents either the real mission environment of a ship or a scenario as a model of this environment. Nevertheless, identified static scenario objects, like air routes, transition corridors, and coastal lines, as well as dynamic scenario objects, like air and surface targets, constitute the starting point of the analysis because from them relevant track objects of the Sensed World and their attributes can be derived.

The Sensed World represents the basis for identifying elements of the SA level 1 which refers to the perception of information elements in the environment of an operator. This environment is represented by characteristics of track objects, i.e., their attributes, states, and behaviour. Therefore, relevant SA information are attributes of static track objects which, as described in detail above, correspond to attributes of static scenario objects of the True World. This information represents a priori knowledge stored on board, for instance, in a geographical data base. Other information of SA level 1 are attributes of dynamic track objects which can be identified by considering available sensors and communication facilities on board the ship and present dynamic scenario objects of the True World. Such track attributes are, e.g., position, course, and speed. This SA information is stored as dynamic track objects with their attributes, e.g., in the central data store of the ship and displayed in any form on consoles of the combat information centre.

The Deduced World contains the same dynamic track objects as the Sensed World but with additionally deduced attributes and their inference processes. These additional attributes and processes constitute information of SA level 2 which are necessary for operator´s comprehension of the current situation. By information processing based on a synthesis of disjoint SA level 1 elements the operator puts these elements together to perform patterns for getting a holistic picture of the environment and an assessment of the current state. As an example, again the dynamic track object to_7 which represents an aircraft can be considered. Deduced attributes of this object (equation 5) belonging to SA level 2 information are, e.g., to_7_d-activity, to_7_d-type, and to_7_d-identity.

But in addition, the Deduced World contains also attributes and processes which reflect information of SA level 3. This elements represent projections of future states of a dynamic track object on the basis of its actual actions and possible actions in future. This is achieved through knowledge of object states and behaviour and comprehension of the situation for both level 1 and level 2. As an example of such a deduced attribute belonging to SA level 3, e.g., the attribute to_7_d-option of the air track to_7 is considered (equation 5). In contrary to the deduced attribute to_7_d-activity which describes the actual observable behaviour of the air track the attribute to_7_d-option portrays the predictable possible object behaviour in the near future (with a maximal prediction time of about five minutes ahead). Both attributes posses the same value sets. Another deduced attribute which belongs to the SA level 3 is the attribute to_7_d-threat with its value set to_7_d-THREAT = {no threat, looming threat, acute threat, critical threat}. This attribute describes possible future threats which depend on the actual activity of the air track and its possible future options. The inference process of this attribute is described in [4].

7. THE UML-MODEL

To support SA of Navy decision makers in identifying air targets the knowledge structures of the sensed and deduced model has to be stored in a knowledge-based user interface. For this implementation the described mathematical model of track objects must be transformed into an object-oriented software specification. To develop this specification the object-oriented Unified Modelling Language (UML) has been applied. UML is a general-purpose visual modelling language that is used to specify, visualise, construct, and document the artefacts of a software system [12]. It captures information about the static structure and the dynamic behaviour of a system. The static structure defines the kind of objects important to a system and its implementation, as well as the relationships among the objects. The dynamic behaviour defines the history of objects over time and the communication among objects to accomplish goals. UML is not a programming language, but tools like TOGETHER [17] are available which provide code generators from UML into different programming languages.

The main constituents of the static view are classes and their relationships [12]. A class is a description of a concept of the application domain. It will be specified by means of attributes and operations. Classes are drawn as rectangles. Relationships among classes are drawn as paths connecting class rectangles. As an example, figure 5 shows upper levels of a class diagram of the modelled track objects containing the class Track Objects with subclasses Static Track Objects and Dynamic Track Objects, and the class Track Processes with subclasses Inference Processes and Control Processes. In this diagram identified processes are dealt with as a separate association class

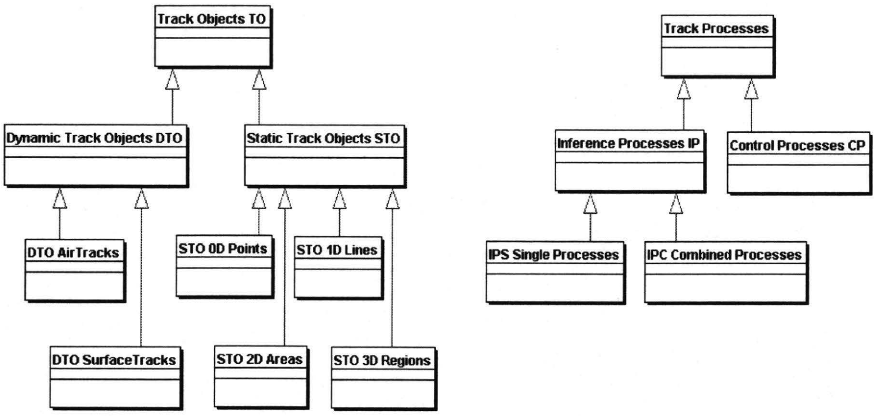


Figure 5: Upper level class diagram of knowledge structure

which is used usually to specify complex association between other classes. Each subclass can be further decomposed into more elementary classes with smaller dimensions, for instance, the class Dynamic Track Objects into classes like Air Tracks, and Surface Tracks. In figure 5, arrows between classes represent generalisation relationships which relates general descriptions of superclasses to more specialised subclasses. Generalisation facilitates the description of inheritance mechanisms which transmit attributes and operations from super- to subclasses.

Besides generalisation links figure 6 shows association and aggregation relationships between classes. An association describes a connection between instances of classes [12]. For example, in figure 6 the plain line between classes STO_Airways and IP_ActivityAirways represents such an association. An aggregation is a special association which specifies a whole-part relationship between an aggregate (a whole) and a constituent part. Aggregations are shown as a diamond on the end of an association line at which it connects to the aggregate class. In figure 6 the link between the classes STO_Segments and STO_CourseChangeAreas represents an aggregation. If the aggregate is a composite, then the diamond is filled like in the aggregation between the classes STO_Airways and STO_Segments.

To model the dynamic behaviour with UML communication patterns are used [12]. These patterns display a set of connected objects as they interact to implement behaviour. The interaction view which provides a holistic view of the behaviour of those objects is modelled by collaborations which describes a society of cooperating objects to carry out some purpose. The interaction between these objects is specified by messages which represent one-way communications, i.e., flows of control with information between objects. Graphically, the interaction is shown by means of sequence diagrams.

A sequence diagram (Figure 7) displays an interaction as a two-dimensional chart. The vertical dimension is the time axis where the time proceeds downward. The horizontal dimension shows those individual objects which participate in the collaboration. For each object a lifeline drawn as a dashed line in a vertical column indicates the life span of the object. During the time an activation of a procedure on the object is active, the lifeline is drawn as double line. A message is shown as an arrow from the lifeline of one object to that of another [12]. Figure 7 shows a sequence diagram of those objects which are involved in determining the activity (object1) of an air track (object3) in reference to an airway (object2) which consists of segments (object4) with course change areas (object5).

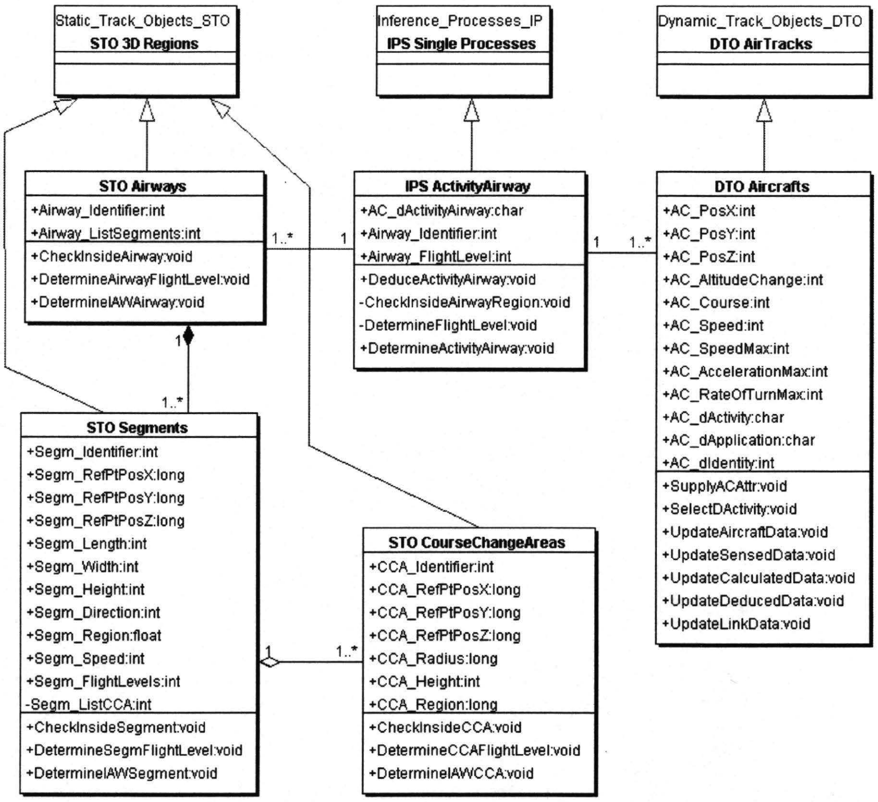


Figure 6: Example of association and aggregation relationships between classes

Another way to model the dynamic behaviour is the life history of one object as it interacts with the rest of the world. This object behaviour is described by means of a state transition diagram which shows the response of the object to events based on its current state, the performance of actions as part of its response, and the transition to a new state [12]. Due to the limited space available this type of diagram is not shown here.

8. REFERENCES

[1] Department of the Navy (1995): Naval Doctrine Publication (NDP) 6, Chapter 2. Naval Command and Control. Washington DC.

[2] Distelmaier, H., Dörfel, G., Döring, B. (2000): A Concept for Knowledge-Based User Support in Naval Environments. RTO Meeting Proceedings 57 on Usability of In-

formation in Battle Management Operations. NATO RTO. Neuilly-sur-Seine Cedex, France; 17.1-17.8

[3] DLR (1999): Intelligent Pilot Assistant. Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Flugführung. Braunschweig, Germany. http://www.dlr.de/fl/Do-kumente/Intelligent_Pilot_Assistant.pdf

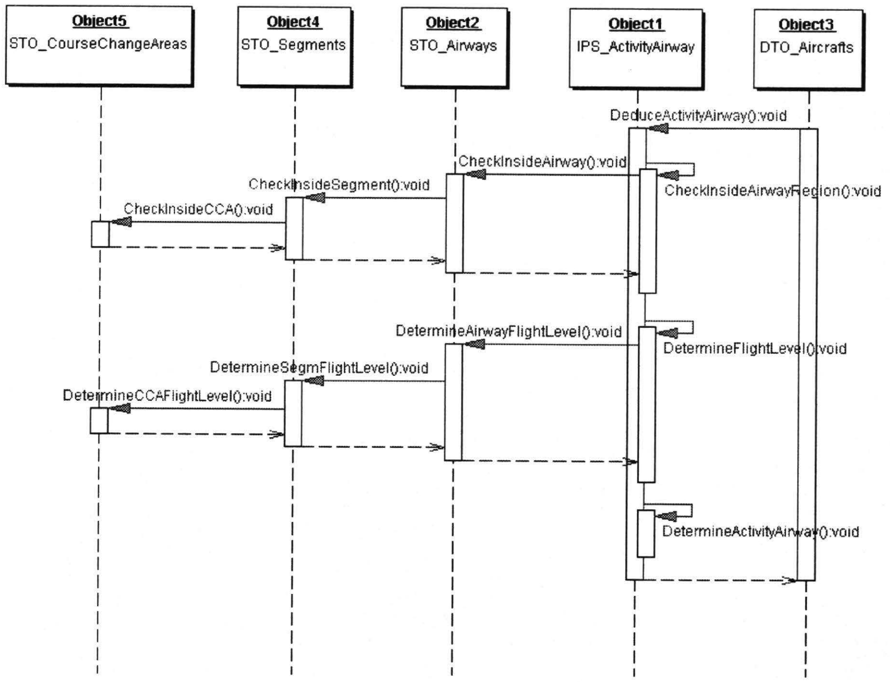


Figure 7: Example of a sequence diagramm

[4] Döring, B., Dörfel, G., Distelmaier, H. (2002): Modelling Situation Awareness Informatio for Naval Decision Support Design. Proceedings of the NATO RTO HFM Panel Symposium on The Role of Humans in Intelligent and Automated Systems. Warsaw, Poland, 7-9 October 2002.

[5] Endsley, M.R. (2000): Theoretical Underpinnings of Situational Awareness: A Critical Review in Dynamic Systems. In M.R. Endsley and D.J. Garland (Eds.), Situation Awareness Analysis and Measurement. Lawrence Erlbaum Ass. Inc. Mahwah, NJ. 3-32

[6] Klir, G. J. (1969): An Approach to a General System Theory. New York, London: Von Nostrand Reinhold Company.

[7] Kopf, M., Simon, J. (2001). A Concept for a Learn-Adaptive Advanced Driver Assistance System. DGLR Bericht 2001-04, Proc. CSAPC01, 24 – 26 Sept. 2001, Neubiberg; Germany. 81-91.

[9] McGraw, K.L., Harbison, K. (1997): User-Centered Requierements: The Scenario-Based Engineering Process. Lawrence Erlbaum Ass., Publisher. Mahwah, NJ.

[10] Orasanu, J., Connolly, T. (1993): The reinvention of decision making. In: G. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), Decision making in action: Models and methods. Ablex Publishing Corp. Norwood, NJ. 3-20

[11] Rasmussen, J. (1976): Outlines of a hybrid model of the process plant operator. In: Sheridan, T.B., Johannsen, G. (Eds.) Monitoring behavior and supervisory control. Plenum Press, London, 371-382.

[12] Rumbaugh J., Jacobson, I., Booch, G. (1999): The Unified Modeling Language Reference Manual. Addison-Wesley, Reading, Massachusetts, Bonn, Amsterdam.

[13] Sage, A. P. (1991): Decision Support Systems Engineering. (Wiley series in systems engineering).John Wiley & Sons, Inc. New York

[14] Salas, E., Prince, C., Baker, D. P., Shrestha, L. (1995): Situation Awareness in Team Performance: Implications for Measurement and Training. Human Factors, 37(1), 123-136

[15] Spawar (1992-199): Tactical Decision Making Under Stress (TADMUS) Program. Library for TADMUS Documents. Space and Naval Warfare Systems Center (spawar), San Diego, CA, http://www-tadmus.spawar.navy.mil/

[16] Stütz, P., Onken, R. (2001). Adaptive Pilot Modeling for Cockpit Crew Assistance: Concept, Realisation and Results. DGLR Bericht 2001-04, Proc. of CSAPC01, 24 – 26 Sept. 2001, Neubiberg; Germany. 177-186.

[17] TogetherSoft Corp. (2002): TogetherSoft Corporation, Raleigh, NC. http://www.togethersoft.com/products/index.jsp.

[18] Vicente, K.J. (1999): Cognitive Work Analysis. Lawrence Erlbaum Associates, Publishers. Mahwah, NJ.

[19] Virtual Prototypes (2002): Introduction to Stage. Virtual Prototypes, Montreal, Quebec, H4P 1T7 Canada. http://www1.virtualprototypes.ca/products_solutions/STAGE.htm

[20] Wickens, C. D. (1992): Engineering Psychology and Human Performance, 2. Ed. Harper Collins Publishers Inc., New York, NY.

[21] Wickens, C. D. (1996): Attention and Situation Awareness, NATO AGARD Workshop, Universität der BW, München-Neubiberg.

[22] Wohl, J.G. (1981): Force Management Decision Requirements for Air Force Tactical Command and Control. In: IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-11, No. 9, 618-638.

Cognitive and Functional (COLFUN) Framework for Envisioning and Assessing High-Demand Situations

Mark A. Neerincx
TNO Human Factors
P.O.Box 23
3769 ZG Soesterberg, Netherlands
Neerincx@tm.tno.nl

Jouke Rypkema
TNO Human Factors
P.O.Box 23
3769 ZG Soesterberg, Netherlands
Rypkema@tm.tno.nl

Peter O. Passenier
TNO Human Factors
P.O.Box 23
3769 ZG Soesterberg, Netherlands
Passenier@tm.tno.nl

ABSTRACT

The human role in complex task environments will more and more focus on handling non-routine situations with increasing information velocity and ubiquity. This paper presents a generic Cognitive and Functional (COLFUN) framework for envisioning and assessing such high-demand situations in order to realize an adequate human resource deployment. The framework consists of two models, a cognitive load model and a functional model, that support a coherent scenario analysis of the task demands and information flows. We briefly discuss two example assessments in early development processes: a traffic-control-center analysis and a task analysis for a naval ship's bridge. Both analyses supported the development and refinement of operating procedures, support systems, manning schemes, work organization and training requirements. In general, COLFUN supports the integration of human factors in the iterative development process of complex human-machine systems.

Keywords

Cognitive task analysis, mental load, human-computer interaction, control centers, process control, user-centered design.

1. INTRODUCTION

Addressing Human Factors in the development processes of complex and dynamic human-machine systems is essential to enhance safety in a large set of application domains, such as industrial process control, aviation, ship navigation and motor traffic. Within each domain, research and collaboration initiatives have been developing human-centered analyses and "best practice" guides. In the European project PRISM (Process Industries Safety Management, http://www.prism-network.org), the Focus Group *Human Factors in High-Demand Situations* aims at improving the joint human-system task performance by reduction of the risks for human resource conflicts and cognitive biases that may appear in high demand situations. This Focus Group explicitly aims at knowledge transfer from other domains to the process control domain. For this purpose, we combined approaches from different domains into a generic Cognitive and Functional (COLFUN) framework for envisioning and assessing high-demand situations. This paper presents the framework and will, subsequently, summarize two example applications from different domains. The first example consists of an analysis of the task load on a naval ship's bridge to assess envisioned task allocations and support functions. The second example comprises an assessment of the tasks for the operator in the future control room of a motor-traffic tunnel.

2. THE COLFUN FRAMEWORK

High-demand situations can be defined from different perspectives. On one hand, for instance, high demand can be described in terms of the *functional* setting and corresponding information transfer processes. For example, the production of the plant might deviate from the planned production, in which case urgent action is required to prevent any production losses. On the other hand, high demand situations can be defined in terms of the *workload* of the human task performer. For example, situations may occur in which the number of tasks or time pressure is so high that the operator cannot perform his or her tasks adequately. The work demands do not match the cognitive capacity of the operator, resulting in mental overload. In the COLFUN framework for the analysis of high demand situations, the functional process and human-factors perspectives meet. First, a model for cognitive task load is described that can be used for the (re)design of cognitive tasks and computer support in complex, real-time, task environments. Second, a model is presented that describes generic process-control functions and information transfer processes. Third, we will argue that the integration of both these two models, in combination with a scenario-based design and assessment approach, will help to identify potential critical situations and provide concrete proposals to better handle such situations.

2.1 Cognitive Load Model

Neerincx [4] developed a cognitive load model, distinguishing three load factors that have a substantial effect on task performance and mental effort. The first classical load factor, *percentage time occupied,* has been used to assess workload in practice for time-line assessments. Such assessments are often based on the notion that people should not be occupied more than 70 to 80 percent of the total time available. Secondly, the cognitive load model incorporates the Skill-Rule-Knowledge framework of Rasmussen [9] as an indication of the *level of information processing.* At the skill-based level, information is processed automatically resulting into actions that are hardly cognitively demanding. At the rule-based level, input information triggers routine solutions (i.e. procedures with rules of the type 'if <event/state> then <action>') resulting into efficient problem solving in terms of required cognitive capacities. At the knowledge-based level, the problem is analyzed and solution(s) are planned, in particular to deal with new situations. This type of information processing can involve a high load on the limited capacity of working memory. To address the demands of attention shifts, the cognitive load model distinguishes *task-set switching* as a third load factor. Complex task situations consist of several different tasks, with different goals. These tasks appeal to different sources of human knowledge and capacities and refer to different objects in the environment. We use the term task set to denote the human resources *and* environmental objects with the momentary states, which are involved in the task performance. Table 1 summarizes a number of indicators of possible problems for each load factor.

Table 1. Some risk indicators for each load factor.

Load factor	Indicators of possible problems
Time occupied	Work overtime Work not finished Insufficient interim, brief rests
Task set switches	Interruptions from the environment (e.g. phone calls) Several problems or tasks to be handled "simultaneously"
Level of information processing	Hardly time for concurrent actions like conversation Extensive use of manuals, help systems etc. Need for advise or assistance Occurrence of non-routine situation for which • the critical elements are hard to identify • it is not immediately clear what actions to perform

The combination of the three load factors determines the cognitive task load: the load is high when the percentage time occupied, the level of information processing (i.e. the percentage knowledge-based actions) and the number of task-set switches are high. Figure 1 presents a 3-dimensional "load" space in which human activities can be projected with regions indicating the cognitive demands that the activity imposes on the operator.

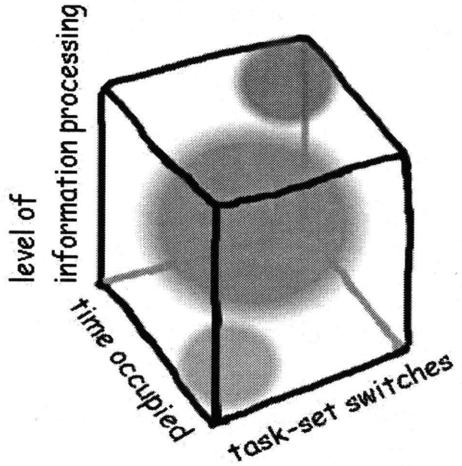


Figure 1. Schematic representation of the task load model.

The middle area represents the area in which task load matches the operator's mental capacity in a certain task setting. In the top area task load is too high. The bottom area represents the area in which performance is not optimal due to underload. The load factors represent task demands that affect human operator performance and effort. When the time occupied is high, and level of information processing and number of task-set switches are low, *vigilance* problems can appear [7]. When the time occupied and the number of task-set switches are high, *cognitive lock-up* can appear (i.e., the tendency of people to focus on single faults, ignoring the other subsystems to be controlled; [3]). The cognitive load model has been used in different domains for task-reallocation *and* design of support functions [4].

Based on the theory and our method for cognitive task analysis, we developed 4 support concepts and for each high-level design principles [6] (table 2):

The Information Handler filters and integrates information to improve situation awareness, i.e. knowledge of the state of the system and its environment, and reduces the *time occupied.* Due to the increasing availability of information, situation awareness can deteriorate without support. Correct information should be presented at the right time, at the right abstraction level, and compatible with the human cognitive processing capacity.

The Rule Provider provides normative procedures for solving (a part of) the current problem and affects the *level of information processing.* Due to training and experience, people develop and retain procedures for efficient task performance. Performance deficiencies may arise when the task is performed rarely so that procedures will not be learned or will be forgotten, or when the information does not trigger the corresponding procedure in human memory. For these situations, rule provision aims at supplementing human procedural knowledge.

The Diagnosis Guide affects the *level of information processing.* The level of information processing increases when no complete (executable) procedure is available to deal with the current alarms and situation. This support function guides the operator during the diagnosis resulting in an adequate problem-solving strategy for a specific task.

The Scheduler affects the number of *task-set switches* by providing an overall work plan for emergency handling. Task priorities are dynamically set and shown in a task-overview to the operator resulting in effective and efficient switches.

Table 2: Load factors and support concepts.

Cognitive load factor	Support concept
Time occupied	Information Handler
Level of info processing	Rule Provider Diagnosis Guide
Task-set switches	Scheduler

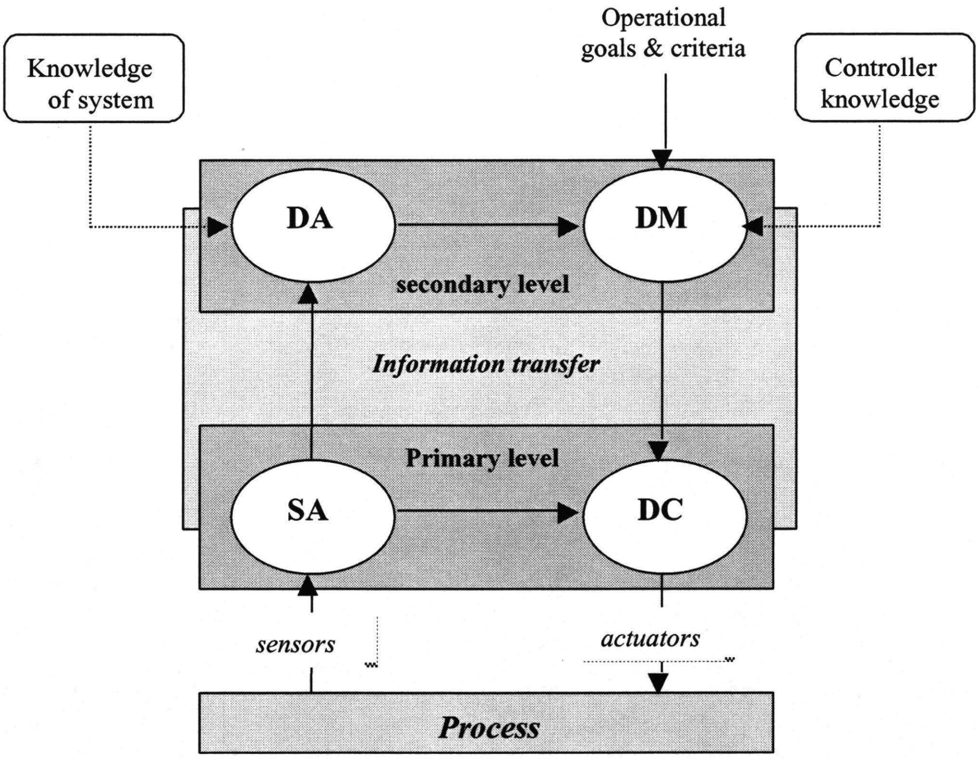


Figure 2. General process-control functions SA, DA, DM and DC, arranged according to two levels of information transfer.

2.2 Functional Model

In general, four generic functions are fulfilled within the control room at two levels of information transfer (figure 2). At the primary level, information provided by sensors is used as input for the crew's *situation awareness (SA)*. Deviations between pre-set values (set points) passed from the secondary level and actual values are directly compensated via the *direction and control (DC)* function. Based on lower-level feedback control loops, adjustments are made, either automatically or assisted by the operator. For example, when the carbon monoxide (CO) level is too high in a tunnel, it will be directly compensated by switching on the ventilation, or when a too-high vehicle approaches, it has to be stopped immediately by the tunnel operator.

At the secondary level, higher-order objectives, determined by the operational goals and criteria for safety and efficiency, are translated into pre-programmed rules for the primary level. Based on the situation awareness and knowledge about the system (e.g. the tunnel) *disturbance assessment (DA)* actions are employed when there are deviations from the planned state. Pre-set goals and criteria, and crew's knowledge are used for *decision making (DM)*. When the goals or criteria cannot be achieved with the current plan, the plan has to be reconsidered. For the tunnel example, when a truck is on fire, the disturbance has to be assessed (e.g., traffic, smoke, and casualties) and adequate decisions have to be made (e.g., announcements in tunnel, resource employment fire control). The functional framework has been used in different domains to identify human and machine tasks, and to improve information transfer in human-machine systems such as naval command center and medical diagnosis [8].

2.3 Scenario Development and Assessment

Cognitive task load can only be analyzed for specific, concrete task contexts. An effective method to create such a context is the use of *scenarios* [1]. Scenarios presuppose a certain *setting*. Within the setting, roles are played by *actors*. In complex scenarios different actors can be involved, possible interacting with each other. Actors have specific *goals or tasks*. To achieve this goal *actions* have to be taken.

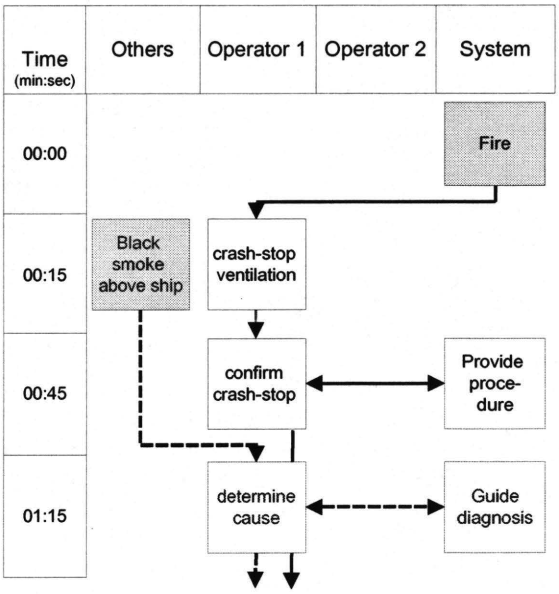


Figure 3: Outline of a Compound Action Sequence (CAS) consisting of 2 Basic Action Sequences (BAS) (i.e. handling of both the fire and the black smoke events).

Neerincx et al. [6] provide a method and description format to systematically create and assess normal *and* critical situations with the corresponding action sequences (figure 3). Such an action sequence displays actions of different actors on a time-line, including the interaction with support systems. The actions can be triggered by events, and are grouped according to their higher-level task (goal).

3. THE MOTOR TRAFFIC EXAMPLE

Rypkema et al. [10] assessed a design of the control room for supervising the traffic in the future Westerscheldetunnel in the Netherlands. The tunnel will measure six kilometers and consist of two separate tubes, each with two driving lanes, connected by corridors. A monitoring and control system (TUBES) has been developed for operating the tunnel, containing 94 cameras, 20 monitors, various sensors and different systems to control the tunnel like traffic lights, speed reduction signs and barriers. The objective of this assessment was to identify possible bottlenecks for the future operator and the envisioned task organization.

Function analysis. First, an inventory of operator tasks was made based on the four generic functions of the process control model (Figure 2). The primary level functions involved mainly tasks that are related to the monitoring and control system TUBES (e.g., *SA* tasks like watching monitors, monitoring sensors and communication systems, and *DC* tasks like control of speed signs, traffic lights and barriers). On the secondary level, a distinction was made between the assessment and handling of incidents—caused by road-users who bring themselves or other road-users in danger—and threats—situations that could bring road-users into danger or lead to incidents.

The *scenario* design was based on three variables: frequency, severity and expected mental load. Highly frequent scenarios occurred more than once a week, low frequent scenarios less than once a month. Severity expressed the number of casualties within a scenario. The expected mental load was defined by domain experts during an interview. Considering the coincidence of variables (e.g. highly frequent severe accidents do not occur), five scenarios were generated.

The *cognitive load* model was used to assess the five scenarios. In three scenarios the complexity and the number of task-set switches showed a peak from the start of the incident to the arrival of the emergency response teams. After their arrival, these teams become responsible and take over a major part of the tasks, so that the operators task load decreases to a lower level.

Results. During serious incidents the mental load is very high just after the incident occurs, especially when there is a fire and the operator has to evacuate the people out of the tunnel. The overload was due to the large number of tasks, the task complexity and in some cases the large number of task-set switches. Also the sudden change from low to high mental load and the operator's responsibilities are burdensome. It was recommended to improve procedures and clustering of tasks. Besides that, it was recommended to support the operator during incidents by deploying a second person who is able to assist the operator within a short period of time (e.g. someone who is working in the same building). Finally, it was recommended to use a simulator that creates a dynamic task environment for selection, training and freshing-up courses.

4. THE SHIP'S BRIDGE EXAMPLE

Van Veenendaal [12] assessed alternative designs for the naval ship's bridge, comprising different task allocations and support functions for navigation and platform supervision.

The *function analysis* resulted in an inventory of operator tasks (i.e. a task hierarchy) and corresponding information needs. It provided insight in the contextual factors that affect the information transfer, in particular for the communication of information about the tactical situation. Furthermore, the functional model helped to define the role of the Officer of the Watch on a naval ship's bridge.

Normal and critical *scenarios* were specified with domain experts, according to the method of Neerincx et al. [6]. Furthermore, for every scenario, support functions were specified and included in the action sequence specifications (i.e. information handler, rule provider, diagnosis guide and task scheduler). The action sequences have been validated with domain experts.

The *cognitive load* model was used to assess these action sequences, with and without the four support functions. First, the three load factors were calculated per 6 minutes task performance, showing the dynamic load fluctuations in the 3-dimensional load cube of figure 1. Subsequently, via questionnaires experts assessed the action sequences to acquire subjective load measures and estimations of the effects of the support functions.

Results. The analysis showed that the task of the Officer of the Watch can be extended with platform control tasks under normal conditions. The support functions will complement the knowledge and experience of the bridge crew to realize an adequate performance level. In critical situations, extra, technical personnel have to be called up. This study provided the first indicators for implementing such a dynamic task allocation.

5. DISCUSSION

Schraagen et al [11] describe individual Cognitive Task Analysis (CTA) approaches and methods for (1) individual training, performance assessment, and selection; (2) the design of human-system interaction; and teamwork situations. They aimed at generic task taxonomy, but concluded that current CTA approaches are diverse and differ on a number of dimensions such as scope, theoretical and empirical foundation, and utility. Consequently, deriving a generic taxonomy is hardly possible.

The COLFUN framework supports a specific type of Cognitive Task Analysis (CTA) that has some similarities and differences with the cognitive work analysis approaches of Rasmussen [9] and Vicente [13]. First of all, our CTA applies the Skill-Rule-Knowledge framework of Rasmussen to determine the level of information processing. Further, both COLFUN and the cognitive work analysis provide a functional view on information processing although at a different level of detail. The main difference seems to be that we do a rather extensive analysis of scenarios and normative procedures, whereas Vicente [13] focuses on supporting knowledge workers in adapting to change and novelty. He challenges the theoretical predisposition of the instruction-based task analyses. The demands of the task domain should be the focus of analysis; such as in the constraint based approaches to work analysis. For training, these approaches help to develop understanding of the task domain, as opposed to learning procedures, holding the promise of flexible response to novel situations. However, they may not result in the required (fast) generation of actions. The more closed a system is, the more amenable it is to instruction-based forms of task analysis. For open systems, workers must adapt online in real time to disturbances that cannot possibly be foreseen by analysis. For this, constraint-based analyses are suitable (although they can be applied when the precise goals cannot be predicted).

In our view, however, design can aim at supporting procedural ("instruction based") task performance, while still enabling adaptive problem solving processes (e.g. by application of the abstraction hierarchy for virtual control panels according to the principles of ecological interface design; [14]). However, the effects of implementing "instruction-based" performances should be well evaluated and should prove to enable adaptive problem solving processes. For example, Grootjen et al. [2] designed a user interface prototype for a ship's bridge that provided the four support functions of table 2. Subsequently, they conducted an experiment to test the effects of the support functions, under high and low task load, on task performance, mental effort and possible side effects (such as operator's loss of situation awareness). In this experiment, 50 RNlN students had to solve damage control problems with the prototype interface. The support proved to result in substantial effectiveness and efficiency profits, i.e. the use of support functions leaded to a substantial improvement of task performance, especially at high task load. Possible costs of being "out of the loop", like not reacting on an implemented wrong advice or a decrease in understanding of performed actions, could not be found.

In process control and related domains, such as aviation and space, improved procedure support can have a major impact on the mission performance (Neerincx et al. [5]). In order to do early, cost-effective assessments, we need instantiations of future work conditions and contexts. COLFUN seems to be a good starting point to realize such a human factors integration in iterative system development processes.

6. CONCLUSIONS

The prevention of and the response to incidents highly depend on the performance of the human task performer. However, during critical situations the task load might become too high for adequate task performance. As a result, incidents may be handled improperly and might escalate. The method described in this chapter provides a framework for the identification of these critical situations through analysis of the functional demands and cognitive load. Both the cognitive load and the functional model have been individually used before, e.g. [8], [4], [2]. Combining the two has resulted in a tool for a structural analysis of operator tasks and information flows, while at the same time the cognitive task load can be measured systematically. The COLFUN framework supported the identification of critical situations and provided concrete proposal for improvement:

- *Procedures*. Transformation of highly complex knowledge-based tasks into less complex rule-based tasks by provision of context-specific procedures and diagnosis guidance.
- *Support systems*. For the four process-control functions, the analyses provided proposals for a support system (from cameras and sensors to advanced decision support).
- *Manning*. The analyses showed when dynamic task allocation helps to handle critical situations (e.g. an (extra) employee takes over some of the operator tasks).
- *Organization*. The framework conveyed requirements for efficient information transfer. For example, the communication between the tunnel operator and the emergency response teams should be supported to prevent an overload of communication tasks for the operator.
- *Training*. It was recommended to train and refresh operators for handling of envisioned critical situations (e.g. in a simulator).

7. ACKNOWLEDGMENTS

We would like to thank Johan van Veenendaal and Marc Grootjen for their contribution to the Cognitive Task Analyses for naval ships. The PRISM Thematic Network is supported by the European Union.

8. REFERENCES

[1] Carroll, J.M. (2000). Making use – scenario-based design of human-computer interactions. Cambridge: MIT press.

[2] Grootjen, M., Passenier, P.O. & Neerincx M.A. (2002). ATOMOS IV Task 3.3.2 Results of additional usability study on dialogue management. Report A403.32.10.052.001. Soesterberg: TNO Human Factors.

[3] Kerstholt, J.H., & Passenier, P.O. (2000). Fault management in supervisory control: the effect of false alarms and support. *Ergonomics, 43*(9), 1371-1389.

[4] Neerincx, M.A. (2003). Cognitive task load design: model, methods and examples. In: E. Hollnagel (ed.), *Handbook of Cognitive Task Design*. Chapter 13. Mahwah, NJ: Lawrence Erlbaum Associates.

[5] Neerincx, M.A., Ruijsendaal, M. & Wolff, M. (2001). Usability Engineering Guide for Integrated Operation Support in Space Station Payloads, *International Journal of Cognitive Ergonomics,* 5(3), 187-198.

[6] Neerincx, M.A., van Doorne, H. & Ruijsendaal, M. (2000). Attuning computer-supported work to human knowledge and processing capacities in Ship Control Centres. In Schraagen, J.M.C., Chipman, S.E. & Shalin, V.L. (Eds.), *Cognitive Task Analysis*. Mahwah, NJ: Erlbaum.

[7] Parasuraman, R. (1986). Vigilance, monitoring, and search. In Boff, K.R., Kaufman, L. & Thomas, J.P. (eds.), *Handbook of perception and human performance*, volume 2: cognitive processes and performance, chapter 43. New York, Wiley.

[8] Passenier, P.O. & Delft, J.H. van (1996). Human Factors in Process Control: designing the Human Interface, *Journal A*, Vol. 37, nr 1, pp 10-14.

[9] Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. Amsterdam: Elsevier.

[10] Rypkema, J.A., Ruijsendaal, M., Bosch, K van den, Schraagen, J.M.C. & Holewijn, M. (2002). *Mentale belasting operator Westerscheldetunnel* [Mental load operator Westerscheldetunnel] (Rapport TNO TM-02-C039). Soesterberg: TNO Human Factors.

[11] Schraagen, J.M., Chipman, S.F. & Shalin, V.L. (2000) (eds.). *Cognitive task analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

[12] Veenendaal, J. van (2002). *Mentale belasting Officier van de Wacht: Een verkennend onderzoek naar de mogelijkheid tot taakuitbreiding en computerondersteuning op de brug* [Mental load Officer of the Watch: An explorative study on the possibilities to add tasks and computer support on the bridge]. Master's thesis. Royal Institute for the Navy, Den Helder, The Netherlands.

[13] Vicente, K.J. Work domain analysis and task analysis: a difference that matters. In: Schraagen, J.M., Chipman, S.F. & Shalin, V.L. (2000) (eds.). *Cognitive task analysis*. Mahwah, NJ: Lawrence Erlbaum Associates. Chapter 7 (pp. 101-118)

[14] Vicente, K.J. (2002). Ecological Interface Design: Progress and Challenges. *Human Factors*, 44, pp. 62-78.

Cooperation between Drivers and In-car Automatic Driving Assistance

Jean-Michel Hoc
CNRS – University of Nantes
IRCCyN, PsyCoTec, B.P. 92101
F-44321 NANTES CEDEX 3
Jean-Michel.Hoc@irccyn.ec-nantes.fr

Jean-Marc Blosseville
INRETS - LCPC
LIVIC, Route de la Minière, Satory
F-78000 VERSAILLES
Blosseville@inrets.fr

ABSTRACT

This paper proposes to use a theoretical framework of Human-Machine Cooperation (HMC) [8] in order to review the main research results on HMC in car-driving, and to identify some psychological problems underlying the introduction of in-car automation as a means of increasing car-driving safety. The use of such a framework is justified by the fact that both driver and automation can be considered as separate agents pursuing their own goals, likely to interfere positively (e.g.: mutual control) or negatively (e.g.: mutual trouble), and trying to facilitate the other's activity. It enables us to describe the cooperative stakes of diverse cooperative modes currently in use or under study within the car-driving context. Finally, some important research questions are outlined, which are partly addressed by a French research program (ARCOS) supported by the Ministries of Research, Transportation, and Industry.

Keywords

Human-Machine Cooperation. Automation. Safety. Car-Driving.

1. INTRODUCTION

In-car automation is on the increase generally, but with some differences between countries. For example, in Japan, more attention is paid to the road's contribution to safety and comfort than is the case in France, especially in terms of telecommunication (information obtained by the car from the road). Roughly, one can distinguish between two kinds of in-car assistance — driving and navigation assistance. This distinction corresponds to two parallel driver's tasks (although many others could be considered, such as conversing with passengers, operating the car radio, etc.). Driving is more concerned with short- and medium-term trajectory management. Navigation is related to the choice of long-term directions. In this paper, we will focus on driving assistance.

For a number of reasons, from marketing to comfort and safety, research and development are making rapid advances, resulting in the introduction of various automatic devices that are supposed to assist with the driving task [5] [25]. The assistance function could be anything from an information provider (e.g., future road adhesion) to an automatic controller (e.g., ABS — Antilock Brake System— aimed at avoiding skids). The perspective taken can look like that found in aviation, but it is, in fact, very different [21]. Certainly, many problems arising from relations between humans and automation have already been studied from within the aviation domain. However, one must be cautious before transferring this framework to car-driving.

There are at least two main differences between airline pilots and car-drivers. First of all, pilots are professionals; which means that they share some standard and well-learned expertise. On the other hand, very few car drivers are professionals. Consequently, in-car automation has to be understandable and usable by a wide range of persons (young, old, nervous, quiet, etc.). Long training periods are not conceivable in the present context. Second, although full automation could be seen as a solution to the problem of increasing road traffic over the next few decades or as a response to some emergency situations, it cannot be implemented in the near future. The drivers themselves must remain as the main car controllers. They cannot be transformed into car automation supervisors.

This creates a situation of cooperation between drivers and automation where both are able to control the vehicle at the same time; a situation which can result in possible interferences. The framework introduced by Hoc [8] to approach human-machine cooperation in industrial situations and in aviation can be one means of clarifying some of the problems posed by relations between car drivers and in-car automation. This paper aims to formulate these problems as research targets and design concerns within the growing domain of car automation, and to sum up the main results in the domain. As a concrete reference, we will take the target of a French research program (ARCOS) which is mainly oriented toward safety improvement [2]. Some existing devices, related to these improvements, will be described. Then, after a brief overview of the theoretical framework for cooperation proposed by Hoc [8], we will try to identify the psychological status of diverse cooperation modes, currently in use or under study. Finally, we will derive some implications of this point of view for research and design within this context.

2. IN-CAR AUTOMATION FOR SAFETY

Although there are some exceptions, such as airbags, in-car safety equipment is seldom sold purely as a safety device, but mostly as a device capable of improving the performance (e.g. speed) of the driver-car system. Within this context, the French research and development program, ARCOS (French acronym for "Research Action for Safe Driving") aims to functionally specify safety devices. The following four safety functions are considered to have a high potential impact on fatal accidents.

? Inter-vehicle distance regulation

This function aims to prevent collisions between vehicles where there are small speed differences. ACC (Adaptive Cruise Control) is an example of the regulation of inter-vehicle distances. However, generally speaking, it is not a safety device as such. Although the general application of such a device to all vehicles could result in preventing some collision risks, ACC, for the moment at least, is not effective in critical situations (e.g., high deceleration of the lead vehicle).

? Stationary or slow-moving obstacle collision avoidance

In this case, the speed differences are very high and the vehicle must stop within a very short distance. ABS (Antilock Brake

System) makes a very partial contribution to this aim by trying to avoid a skid when the brakes are applied strongly.

? Road departure prevention
More generally, this is the problem of lane departure and of the control of the lateral dimension of the trajectory. ESP (Electronic Stability Program) is a partial solution to the problem, helping to avoid spin when the steering angle is too great.

? Upstream accident/ incident alert
This function implies that telecommunication exists between cars, and between cars and the road, to provide information for upstream drivers. It will not be considered within this paper.

3. A FUNCTIONAL APPROACH TO HUMAN-MACHINE COOPERATION

The framework proposed by Hoc [8] is based on a functional rather than structural approach and is aimed at describing relational structures between cooperative agents. Its main objectives are to identify, analyse, implement and support cooperative activities. These are considered to be added to private activities when moving from an isolated individual activity to a collective activity. For example, function allocation is a cooperative activity and would be unnecessary if an agent was working alone.

The framework proposes that cooperation is an activity of interference management between non-independent tasks distributed among several agents. Interference is managed in order to facilitate the individual tasks or the collective task as it stands. Different agents can cooperate without a common task; they can perform different tasks, interfering only with the sharing of common resources. For example, car-drivers pursue different goals, but they cooperate because they share the same road. Cooperation does not always imply a perfect symmetry between the diverse agents. Sometimes there are strong reasons for giving priority to the facilitation of one particular agent's task. In-car HMC is supposed to give priority to the driver's task, at least in normal circumstances.

In order to manage interference, the framework decomposes the cooperative activities involved into three (abstraction and temporal span) levels: action (short-term and local interference resolution); plan (medium-term and resolution by planning) and meta (long-term and resolution by the means of high level models, especially of the agents).

This framework has mainly been applied to cooperation at the symbolic level (e.g. the processing of verbal representation). It could, however, be extended to address cooperation at the subsymbolic level (e.g. sensorimotor coordination). This transposition to car-driving is possible where human-machine cooperation (with automatic devices like those evoked above) mainly relies on sensorimotor coordination.

3.1 Action Level

At the action level, interference management is restricted to the short term, with a minimal anticipation of the agents' goals. The positive or negative features of interference are highly relative. Interference occurs when the tasks are not independent. This means that the tasks can be in precondition relations (one being necessary to perform another one), in interaction relations (the two), or in redundancy relations (the same goal can be reached by any of the agents). Interference can take the form of mutual control when an agent checks another agent's activity to give back an evaluation. So, interference is not only negative, it can be deliberately created to improve effectiveness. After its appearance at the action level, interference can also be managed at the other cooperation levels.

As expected, ergonomics has been more sensitive to the negative aspects of cooperation than to the positive ones [13]. As is the case in other domains, automation can have undesirable and unexpected effects [15]. In order to reduce human-machine interference, considered to be negative, the driver may unconsciously circumvent automation, thus reducing its efficiency. For example, drivers are likely to occupy the left lane more with ACC than without this device [14] [17]. A variety of behavioural adaptation which goes against safety has often been described as risk homeostasis [26]. Automation can give the driver the feeling that there is no danger in critical situations because it suppresses anxiety signals and encourages the driver to take risks beyond those accepted in non-assisted situations. This phenomenon has frequently been described, for example, with vision enhancement devices when driving in fog (increase in speed: [22]), or with ABS (increase in speed or decrease in THW — time headway —: [7] [9] [18]).

3.2 Plan Level

At the plan level, a common frame of reference (COFOR) is managed in order to facilitate the activities situated at the action level. COFOR not only includes representations of the environment (team situation awareness), but also of the team's activity (e.g. common plans and goals, function allocation, etc.). It is easy to create conditions for shared awareness of the external situation, although the information format must be adequate. On the other hand, it is much more difficult to maintain a shared representation of the team's activity; for example, intention recognition by a machine.

COFOR is maintained or elaborated at this plan level, but the by-products of the action level activities can also enrich it covertly. COFOR does not necessarily mean the sharing of identical representations, which can only be compatible. For example, if a physical variable like road adhesion underlies the automatic device functioning, this does not imply that the driver must use this measure as such. Sensorial information coherent with the adhesion value may be a more efficient way to influence the driver's activity.

This cooperation level has not yet been directly addressed within the car-driving context. In the years ahead, the unexpected benefits it may yield could make it one of the most promising areas yet to be researched. For example, if automation exerts a mutual control on the driver's activity, it will be understood and accepted by the latter only if the two agents share the same situation analysis in terms of risks.

3.3 Meta Level

At the meta level, the experience of cooperation within the team is exploited to facilitate the activities situated at the previous levels; for example, the use of models of the other agents and of oneself. At this level, trust in automation and in one's relations with automation, together with self confidence, can be developed through the use of models elaborated by experience.

Although there is not an abundance of research on this question, those few studies that are available show the importance of this level. Within this context, where the shift between manual and automatic control is not as sharply defined as within other domains, the main subject of trust is neither the machine nor the driver, but rather the interaction between the two. Rajaonah [16] has shown the role played by the driver's elaboration of a model of interaction in the development of trust, when driving with ACC. The variability in results concerning the effects of ABS in accidents allows us to think that risk homeostasis is not the only phenomenon that needs to be considered. The driver's ABS model should also be taken into account. For example, Broughton and Baugham [3] have shown that ABS reduces the number of accidents involving young drivers, but not those involving women and older drivers. They suggest that the latter group of drivers do not operate the device properly because of an incorrect model of the device. Mollenhauer *et al.* [12] have positively tested the effect of a brief training period using images and texts similar to those available for presenting safety instructions in aircraft. Stanton and Young [23] have shown an increase in the number of accidents in an ACC simulator study where the lead vehicle stops suddenly. They account for the phenomenon with the fact that drivers tend to assume, incorrectly, that ACC includes a collision avoidance function.

4. COOPERATION MODES BETWEEN DRIVER AND AUTOMATION

For a long time, cognitive engineering has elaborated diverse typologies of human-machine function allocation modes, mainly covering supervision activities (but also teleoperation: [19]). Obviously, the cooperation issues at stake concern cooperation modes between the fully manual mode and the fully automatic mode. Developing safety devices within a context where the driver should remain the main entity in charge of driving means that automation should intervene sparingly. Before looking for highly invasive interventions in the fully automatic mode (for example in emergency situations), lighter interventions should first be envisioned. These intervention modes are related to HMC and may be classified, following the theoretical framework presented above, in terms of cooperation modes in order to make their psychological implications more salient. At the same time, the modes will remain readable from an engineering point of view. They will be presented from the least to the most invasive.

4.1 Perception mode

The machine is utilised as an extension of the sensorial organs. In terms of engineering, it is the instrumented mode. Although the production of a physical measure is a well-defined task for an engineer, the usefulness of this measure for the driver is questionable. At the symbolic level, drivers are used to consulting their speedometer, which they interpret in terms of regulation or action. However, where adhesion is concerned, one doubts that drivers can interpret the physical measure as easily as they do the speedometer. Inasmuch as car-driving involves a strong sensorimotor component, the main question to consider is what function this kind of information has in the sensorimotor loop (from perception to action, and from action to feedback). That is why the "perceptive" mode is a more suitable way of expressing this than the "instrumented" mode.

Above all, the perceptive mode produces sensorial information. However, two distinct cognitive control levels may be triggered when processing such information. At the symbolic level, the sensorial information (form) alone is not sufficient for processing it (e.g., "50"). An interpretation activity is needed to reach the relevant meaning or content, which is not always perceivable (e.g., 50 Km/h has not the same meaning as 50 degrees). Symbolic processing is serial and, therefore, very costly in terms of attentional resources. In addition, the information that is processed is discrete and not continuous, and this is not compatible with smoothness of action. Subsymbolic processing only deals with perceptions, without needing interpretation. For example, it has been shown that the speed travelled at whilst negotiating a curve is regulated in order to maintain lateral acceleration at an acceptable level. Subsymbolic processing is parallel, much less costly than symbolic processing in terms of attentional resources and response time, and can include continuous information.

In any case, the perceptive mode must be considered as a cooperation mode since it is designed to interfere with the driver's activity. It is mainly a question of choosing the best Human-Machine Interface (HMI), either in terms of an appropriate code (form/content) to support symbolic processing or an efficient sensorial modality to easily trigger action. Vision and audition have been favoured within the domain, whereas the haptic modality, which integrates tactile information along time, has been neglected. Information types may be utilised successively. For example, when approaching a curve, speed advice can be produced at the symbolic level since there is a sufficient amount of time for the driver to plan at this level. On the contrary, when negotiating a curve, there is no time to process this kind of information and it would be more appropriate to simply draw the driver's attention towards information that is crucial for the sensorimotor loop. For example, Land and Lee [10] have shown that the point where the curve reverses on its inner side is critical.

4.2 Mutual control mode

In the mutual control mode, the machine is designed in such a way that it can interpret information in terms of limits to be respected in relation to risk assessment. Thus, it can provide drivers with feedback on their actions (mutual control) in terms of exceeding limits. Four modes can be envisioned, all with different degrees of invasiveness. The warning mode and the action suggestion mode are restricted to (interpreted) information transfer, without any action taken on the vehicle itself. The limit mode, although under the driver's control, introduces more constraints, for example by creating pedal or wheel resistance. A fourth possibility (correction mode) is to let the driver go beyond the limit and then to make the required correction. An appropriate COFOR between the machine and the drivers must be maintained in order to render the machine's mutual control understandable.

4.2.1 Alarm mode

Here, alarm is not taken in the sense of information provided on the car's technical state (e.g., a fault), but as a criticism of the driver's actions. Two interesting experiments [11] have shown the positive effect of an early collision avoidance alarm which did not reduce the response time directly, but did give the driver more time to undertake situation analysis and action preparation. This mode has also been shown to present a positive border effect. Although used only for a brief period of time, an alarm on short THW (Time HeadWay) can durably lengthen THW [1] [20]. However, this positive effect only seems to be observable when it is related to safety, and not simply to road regulation ([4]). Thus, the availability of a COFOR between the driver and the machine is a very important issue. If the risks assessed by the driver and those assessed by the machine are very different, the efficiency of the alarm could be greatly restricted.

4.2.2 Action suggestion mode

When the alarm mode is present on a control (e.g., a pedal or wheel), by using the haptic modality it could become an action suggestion and, therefore, could be more effective in emergency situations.

4.2.3 Limit mode

With the limit mode, the intrusion into the control of a vehicle becomes clear. For example, when approaching the limits of the envelope of acceptable trajectories, any action taken by the driver on the wheel which would lead the vehicle outside the envelope will encounter resistance, as if the vehicle is in the gutter. The calibration of this mode must consider two opposite adverse effects. On the one hand, the resistance must be seen to be unpleasant enough to avoid using the device for comfort rather than for safety assistance. On the other hand, the resistance should not produce excessive stress.

4.2.4 Correction mode

The correction mode goes beyond the limit mode, producing not just resistance but also an action which corrects control. When approaching the limit, an action taken by the driver which would lead the car outside the lane triggers the wheel to turn in the opposite direction in order to return the car into the acceptable envelope.

Within the mutual control mode, current research is only concerned with the alarm mode. However, as the other modes are now attainable, more effort should be devoted to these. All these modes assume that there is a shared understanding of the situation (COFOR) between the driver and the machine. This question needs to be studied more explicitly. In addition, mutual control should not be considered simply as the machine's control of the driver. The reverse should be possible, since more often than not the driver has a wider situation analysis than the device. For example, sometimes priority must be given to face-on collision avoidance rather than to road departure avoidance. In other domains (e.g., aviation), such a symmetry has been suggested to avoid complacency. This has led to the definition of two separate supervision fields — of the machine and the human — which could produce very adverse effects when cooperation is needed.

4.3 Function delegation mode

The cooperation modes considered under the function delegation category go beyond simple mutual control. They correspond to a lasting function delegation from the driver to the machine. In the mediatized mode, the machine takes a control as an order to be implemented using a procedure which covers a certain period of time. In the regulation or prescription mode, the machine regulates a parameter, thus allowing the driver to take charge of the others. The desired value can be decided by the driver (regulation mode) or by the road manager (prescription mode). There is also a need for an efficient COFOR maintenance to be sure that the drivers know the current function allocation and the decisions of the road manager.

4.3.1 Mediatized mode

In this mode, the driver's action on a control will not have a direct effect on the car's behaviour. It will be taken by automation (acting as a mediator) as an order to implement in safe conditions. ABS and ESP (Electronic Stability Program, aimed at avoiding spin when the wheel control is too strong) are typical examples of this mode. As far as ABS is concerned, we have already seen just how important is the driver's understanding of the functioning of the device in order to avoid surprise and to ensure that the device is triggered off adequately. Finally, confusion remains possible between the fully manual control and the mediatized mode, since the same driver's action on a control can have a different status depending on driving conditions.

4.3.2 Control mode

Delegation is more long-lasting in the control mode than in the mediatized mode and its triggering does not use the same control as that used in manual control. There is no risk of confusion, but the driver must know how the device can be engaged or disengaged. ACC is a typical example of the control mode. The driver defines a cruise speed and a THW, and the device controls the longitudinal dimension of driving, leaving the driver in charge of the lateral dimension. The control mode, along with the following phenomena, is closer to conditions found in automation in other domains, such as aviation, than in those already discussed. Stanton and Young [23], amongst others, have anticipated that well-known automation bias could be (and have been) observed, and these are listed below.

? Complacency. This phenomenon has already been evoked in relation to the correction mode as a rigid boundary between the driver's and the machine's supervision fields. When the two fields are closely related (e.g., lateral control and longitudinal control), complacency difficulties can be expected. Only one paper refers to this phenomenon in car-driving automation [24] and this should be confirmed.

? Bypassing. When a driver experiences serious difficulties, behavioural adaptation can lead to the device being bypassed. We have already cited the well-known effect of ACC leading to drivers occupying the left lane for longer periods of time.

? Over-generalisation. The device is utilised outside its validity domain (e.g., ACC assimilated to a collision avoidance device).

? Automation surprise. The driver is surprised by the automation's behaviour. This may be due to a lack of an adequate model of the device functioning. This may also be related to a deficient COFOR and a subsequent gap in situation analysis.

? Difficulty when returning to manual control. For the moment, this difficulty has only been shown with an AS (Active Steering) device aimed at negotiating curves automatically; see § 4.4).

4.3.3 Prescription mode

The prescription mode poses the same problems as the control mode, except that the set-point is imposed by the infrastructure rather than chosen by the driver. Thus, new difficulties are introduced in terms of regulation acceptation and are in addition to the previously cited difficulties. This question has already been raised in relation to the alarm mode, when the alarm refers more to regulation than to risk assessment.

4.4 Fully automatic mode

With the fully automatic mode, automation takes overall charge (at least in terms of guidance) of car-driving. It is reasonable to consider two cases where this mode has relevance. In the first case, the driver is identified as being unable to control the vehicle. A typical example is when emergency braking occurs to avoid a collision with a stationary or slow-moving obstacle. If the TTC is longer than the driver's minimal response time, automatic emergency braking is justified, even if the result would only be to mitigate the crash. In the second case, the risks are seen to be very high. Typical examples are roads through tunnels, mountain roads and road works, where lateral control is crucial. In this case, returning to manual control could pose very serious problems. Firstly, the return to a "normal" situation where the system can return to manual control must be identified. Secondly, information needed by the driver to manually control the car must be defined. The aim is to avoid the well-known "human out of the loop" difficulty. Such information does not only concern the present but also the past and the format could be sensorial. It is related to the maintenance of a COFOR during automatic control. The haptic modality is possibly of interest for such maintenance.

5. RESEARCH IMPLICATIONS

For the moment, it is very difficult to produce a sizeable state-of-the-art study on HMC in in-car automation. However, studies conducted in other contexts as well as in car-driving, together with our theoretical approach, enable us to raise some important research questions for the future. We will suggest seven concerns. Is automation developed for safety or for comfort? Is there any interference between automation and the other tasks performed by the driver? How can mutual control be efficient? How can a COFOR be maintained? What model of automation must be available to the driver? How should function allocation be designed? What are the technical costs of in-car automation?

5.1 Automation for safety or comfort?

Whilst safety devices are considered crucial, the development of devices for comfort is questionable. Designers should not allow the drivers to think that there is no risk when the devices are brought into play. The interpretation of some behaviour in terms of risk homeostasis, even if it looks attractive, should be based on more abundant empirical results. In order to establish a policy based not on comfort, with the designing of automation that looks to reinforce hazard signals rather than erase them, further studies are necessary which will more firmly validate risk homeostasis hypothesis. Other complementary hypotheses, such as Fuller's threat avoidance theory [6], should be considered. However, as soon as safety devices intervene by anticipation (e.g. ACC), without any particular hazard signal, the confusion between safety and regulation imposition may render automation inefficient.

5.2 Interference of automation with the other functions performed by the driver

The complexity of the situation analysis performed by automation is not comparable to that which is performed by the driver. Except in extreme cases, where automation takes charge of most areas of driving, automation will only perform a specific driving function. Thus, the driver will be responsible for many other functions that may interfere with the automated ones. In order to adequately resolve such interference, the question of which is the priority function must be posed. If the priority function is in the charge of the driver, it is convenient to avoid bothering with the performance of this function (e.g., cooperation with another road user). Conversely, if the priority function is performed by the machine (e.g. imminent accident), any possible actions on the part of the driver that are capable of jeopardizing the automation functioning should be avoided. That is why an incremental triggering of cooperation modes can be judicious. For example, in curve negotiation the limit mode may be adequate when there is no certitude that the driver's action is priority. But, when the imminence of an accident can be identified with certainty, the correction mode becomes relevant.

As we saw earlier, cooperation does not only introduce negative interference but also positive interference (e.g., mutual control). However, in order to ensure the expected positive effect, two conditions need to be satisfied, to which too few studies are devoted — information temporality and format. The choice of temporality is related to the cognitive control modality which is concerned with information within the "origin of the data" dimension needed for either: anticipative control (with a prominence of internal data) or reactive control (more importance given to external data). In the first case, automation intervention must occur early on. For example, in curve negotiation, the intervention could be an action suggestion in terms of approach speed. In the second case, intervention should take place in real time, providing the driver with sensorial information compatible with human routines. Whereas interference with a reactive control must only be subsymbolic, in the case of anticipative control it may be both symbolic and subsymbolic. For example, an announcement of future road adhesion whilst negotiating a curve may be more efficient if it is translated into a sensorial stimulus.

5.3 Mutual control

The studies cited on mutual control through alarms clearly question the respective efficiency of critics based on a shared representation (or compatible representations) of risks or on some regulation that is not considered by the driver to be related to risk. For example, if critics are entirely out of touch with hazard signals, they might be inefficient because they are likely to be interpreted as regulation imposition rather than the means to improve driving safety. The respective effects of mutual control modes (alarms, action suggestions, limits and correction) need to be examined more closely, as do the types of contexts in which negative interference could be minimal.

5.4 Common frame of reference (COFOR)

On several occasions, we have stressed the importance of the elaboration and maintenance of a common frame of reference (COFOR). As already shown, it groups together shared situation awareness and, at the same time, awareness of the available resources engaged by the agents. This sharing does not mean a close similarity between the individual frames of reference (Current Representation of the situation and of the resources: CR). They must, however, be compatible at the very least. Any automation intervention is likely to be rejected if the agents' CRs are too far from each other. It may be quite easy to transmit the automation's CR to the driver, although the information format must be chosen carefully to make the driver aware of a certain risk level. On the other hand, it still seems too difficult a task to enable drivers to transmit elements of their CR to automation. Apart from real time collection and the interpretation of a driver's behaviour, such an explicit transmission could introduce an overload.

5.5 Model of automation

Some training solutions have been tested (e.g., ABS). However, the benefits and formats of this type of intervention need to be more widely investigated. Our aim is to facilitate the construction of a minimal and relevant model of automation by the driver. Nevertheless, we must go ahead with enabling the driver to elaborate a model of the human-machine interaction. It is the price to be paid for an adequate calibration of trust; that is to say, one that is based on some experience of the interaction, enabling the driver to distinguish between situations where trust is justified and those where it is not.

5.6 Function allocation

The function allocation modes have not been widely studied. Thus, problems that stem from choosing the entity in charge of the allocation (the driver or the machine) have not been addressed, as has been the case in the field of aviation. Explicit delegation by the driver and its cancellation may introduce an overload likely to affect driving if it is not integrated at a sufficiently low regulation level for driving activity. In addition, the frequency of delegation should be considered when developing safety devices. Function allocation modifications can be frequent, but must be completely integrated into human activity. In this context, the allocation is likely to be seldom modified. At the same time, automation triggering may occur infrequently, which could be an obstacle to an "expert" management of the interaction between the driver and the automation device.

6. CONCLUSION

In summary, whilst benefiting from knowledge accumulated in other domains, such as aviation, HMC in car-driving safety automation still raises a number of questions. For some of these, knowledge already available in other domains or in car-driving itself will enable us to provide designers with reasonable answers. However, there are too few empirical results to validate some of the hypotheses proposed in this paper. Other questions have not been examined before; for example, familiarization with safety automation that is seldom used whilst driving under ordinary conditions. For this reason, the HCM research domain will be enriched by further development and resolution of studies on this question in the car-driving domain.

7. REFERENCES

[1] Ben-Yaacov, A., Maltz, M., & Shinar, D. (2002). Effects of an in-vehicle collision avoidance warning system on short- and long-term driving performance. *Human Factors, 44,*335-342.

[2] Blosseville, J.M. (2002). ARCOS, un projet fédérateur [ARCOS, a federation project]. *Revue Autoroute et Mobilité,* n°9.

[3] Broughton, J., & Baughan, C. (2002). The effectiveness of antilock braking systems in reducing accidents in Great Britain. *Accident Analysis and Prevention, 34,* 347-355.

[4] de Waard, D., van der Hulst, M., & Brookhuis, K.A. (1999). Elderly and young drivers' reaction to an in-car enforcement and tutoring system. *Applied Ergonomics, 30,* 147-157.

[5] Dingus, T.A., Gellatly, A.W., & Reinach, S.J. (1997). Human computer interaction applications for intelligent transportation systems. In M. Helander, T.K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (Rev. ed.). (pp. 1259-1282). Amsterdam: Elsevier.

[6] Fuller, R. (1984). A conceptualization of driving behaviour as threat avoidance. *Ergonomics, 27,* 1139-1155.

[7] Heino, , A., Rothengatter, J.A., & van der Hulst, M. (1995). *Collision avoidance systems safety evaluation* (DRIVE II Project V2002, Deliverable Report 33, Workpackage 0016). Groningen: University of Groningen.

[8] Hoc, J.M (2001). Towards a cognitive approach to human-machine cooperation in dynamic situations. *International Journal of Human-Computer Studies, 54,* 509-540.

[9] Hoedemaeker, M., & Brookhuis, K.A. (1998). Behavioural adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F,* 95-106.

[10] Land, M.F., & Lee, D.N. (1994). Where we look when we steer. *Nature, 369,* 742-744.

[11] Lee, J.D., McGehee, D.V., Brown, T.L., & Reyes, M.L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors, 44,* 314-334.

[12] Mollenhauer, M.A., Dingus, T.A., Carney, C., Hankey, J.M., & Jahns, S. (1997). Anti-lock brake systems: an assessment of training on driver effectiveness. *Accident Analysis and Prevention, 29,* 97-108.

[13] Neboit, M. (1995). *Advanced support to car-driving* (Special Issue). *Le Travail Humain, 58*(2).

[14] Nilsson, L. (1995). Safety effects of adaptive cruise controls in critical traffic situations. *Proceedings 2nd World Congress Intelligent Transport Systems* (pp. 1254-1259). Tokyo: VERTIS.

[15] Parasuraman, R. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

[16] Rajaonah, B. (2001). *Étude de la confiance établie par le conducteur automobile dans sa relation avec un système d'aide intelligent, l'ACC: un cadre théorique et méthodologique.* [Study of car drivers' trust in their relation with an intelligent assistance system]. Mémoire de DEA, Université de Paris 8.

[17] Saad, F., & Villame, T. (1999). Intégration d'un nouveau système d'assistance dans l'activité des conducteurs d'automobile [Integration of a new assistance system in car drivers' activity]. In J.G. Ganascia (Ed.), *Sécurité et cognition* [Safety and cognition] (pp. 105-114). Paris: Hermès.

[18] Sagberg, F., Fosser, S., & Sætermo, I.A. (1997). An investigation of behavioural adaptation to airbags and antilock brakes among taxi drivers. *Accident Analysis and Prevention, 29,* 293-302.

[19] Sheridan, T.B., & Verplanck, W.L. (1978). *Human and computer control of undersea teleoperators.* Cambridge, MA : MIT Man-Machine Systems Laboratory Report.

[20] Shinar, D., & Schechtman, E. (2002). Headway feedback improves intervehicular distance: a field study. *Human Factors, 44,* 474-481.

[21] Stanton, N.A., & Marsden, P. (1996). From fly-by-wire to drive-by-wire: safety implications of automation in vehicules. *Safety Science, 24,* 35-49.

[22] Stanton, N.A., & Pinto, M. (2000). Behavioural compensation by drivers of a simulator when using a vision enhancement system. *Ergonomics,43,* 1359-1370.

[23] Stanton, N.A., & Young, M.S. (1998). Vehicle automation and driving performance. *Ergonomics, 41,* 1014-1028.

[24] Stanton, N.A., Young, M.S., Walker, G.H., Turner, A., & Randle, S. (2001). Automating the driver's control tasks. *International Journal of Cognitive Ergonomics, 5,* 221-236.

[25] Walker, G.H., Stanton, N.A., & Young, M.S. (2001). Where is computing driving cars? *International Journal of Human-Computer Interaction, 13,* 203-229.

[26] Wilde, G.J.S. (1982). Critical issues in risk homeostasis theory. *Risk Analysis, 2,* 349-258

Failure of Distributed Coordination in High Reliability Organizations: The Midair Collision over Germany on July 1, 2002[i]

Paul C. van Fenema
Erasmus University
P.O. Box 1738
3000 DR, Rotterdam, The Netherlands
pfenema@fbk.eur.nl

Steven J. Simon
Mercer University
Atlanta, GA 30341
United States of America
simon_sj@Mercer.EDU

ABSTRACT

High Reliability Organizations (HRO) support vital processes in our society. They operate under conditions that require optimal safety and efficiency, like fire fighting, emergency services, military operations, air traffic control and space operations. Many of these organizations operate geographically dispersed. This contributes to their complexity and risk.

Coordination failures of dispersed HRO have huge consequences for people, machines, and the environment. Organizations focus extensively on avoiding these. In this spirit, understanding of coordination failure is important to academics and practitioners.

This paper reports on a midair collision over Germany on July 1, 2002. It frames the collision as a failure of dispersed coordination. Relevant theories are presented to support this perspective. The case itself is described and analyzed using secondary data that have become available on the disaster. The minutes leading the collision are structured in three episodes.

Our analysis suggests that multiple factors contributed to the collision, such as knowledge gaps between pilots and controllers, knowledge diversity between pilots on the different flights, (partial) transitions between different coordination systems, and the controller's disconnection from time/ space reality. Above all, the case shows that time pressure reduces opportunities for remote coordination. Ex ante training and standardization seem indispensable for ensuring reliable coordination among geographically dispersed actors.

Keywords

Distributed cognition, coordination, remote communications, midair collision, human factors

1. INTRODUCTION

On July 1, 2002 at 21:35:32 hours UTC, a Tupolev TU 154 M (flight BTC 2937) and a Boeing B757-200 (flight DHX 611) collided at approx. FL 350[1] over Germany near the town of Überlingen close to Lake Constance in Germany. All 71 persons (2 crew members on board the Boeing, and 12 crew members and 57 passengers on board the Tupolev) died. Flight DHX 611 flew from Bergamo (Italy) to Brussels (Belgium), and flight BTC 2937 was en route from Moscow (Russia) to Barcelona (Spain).

At the time of the accident, one controller of the Swiss air traffic control (ACC Zurich, operated by a company called Skyguide) was monitoring the sector involved. He worked alone, monitoring two workstations with radar screens while his colleague took a break in accordance with company policies. The ground-based collision warning system (Short Term Conflict Alert, or STCA) was not available due to maintenance. Telephone connections to adjacent control centers did not function for unknown reasons.

The German Federal Bureau of Aircraft Accidents Investigation (Bundesstelle für Flugunfalluntersuchung) commenced its investigation shortly after the accident, resulting in a Status Report in August 2002 (BFU 2002). According to the Bureau, both aircraft and their on-board collision warning devices (Traffic Alert and Collision Avoidance System, or TCAS) functioned normally. The report describes key facts of the accident including a description of actions taking place in the period leading to the collision, visualized flight paths, description of aircraft and site damage, and information on the aircraft, their crew and air traffic control (ATC).

As an initial report, the document offers factual information on the accident but cannot yet offer an interpretive analysis of events leading to the collision. This interpretation of events and the contexts of the disaster is necessary to understand causes and develop possible remedies.

The objective of this research is to improve the understanding of the accident's causes and drivers. We want to contribute to the overall goal of increasing our understanding of the accident from different angles, believing that this is important for designing and implementing strategies that will reduce the risk of future similar accidents. This study extends research on failure in High Reliability Organizations, specifically those operating geographically dispersed (Snook 2000; Orlikowski 2002). Understanding this type of failure is vital for operations in space, fire fighting (Bigley and Roberts 2001), military operations, and emergency and disaster response.

Our research is setup as follows. First, the accident is interpreted from the angle of failure of coordination and collaboration. We propose theories that support an explanation of coordination failure in High Reliability Organizations (HRO) that operate geographically distributed (Schulman 1993; Bigley and Roberts 2001). Second, this

---

[1] Fl 350 means that the airplanes were flying at flight level 35,000 feet, which equals about 6.6 miles or 10.7 kilometers.

framework is used to analyze factual information available on the case. This analysis is organized in a sequential mode (i.e., we follow the events as they occurred on July 1, 2002). We elaborate in detail on the last 5 minutes leading to the collision.

2. THEORETICAL BACKGROUND

This section outlines the theoretical basis of our analysis and includes the following concepts. We perceive the Air Traffic Control system as a High Reliability Organization with specific mechanisms that support coordination. Failure of these organizations has been analyzed in the context of fire fighting, operations of flight deck ships (Rochlin, LaPorte et al. 1987), and friendly fire (Snook 2000). These lessons can be applied in this case.

First, Air traffic control can be interpreted as a High Reliability Organization focused on avoiding failure and breakdown (Schulman 1993; Bigley and Roberts 2001). The ATC system is not a single entity HRO like a flight deck ship, but composed of a (usually national) control organization, and daily evolving contacts with pilots whom are served.

Second, ATC represents a complicated system that includes human actors (pilots, controllers), high tech artifacts (radar, information technology in the cockpit and the controllers' facilities), and elaborate organizational and institutional procedures and standards. This regulated environment relies on coordination mechanisms like communications (between pilots and pilots and controller), rulers, standards, and shared professional knowledge (Van de Ven, Delbecq et al. 1976; Malone and Crowston 1994).

Third, as collective mind theory stresses, common background is not sufficient for avoiding coordination failure (Weick and Roberts 1993; Weick, Sutcliffe et al. 1999). Communications between stakeholders in a risky situation is required to sustain situational awareness and maintain coordination. The combination of shared knowledge and frequent interaction sustains coordinated collective action.

Fourth, failure of HRO organizations has been the subject of a rich research stream. Weick and Roberts (1993) showed that accidents happened on flight deck ships because people did not comply with procedures, or they could not improvise in situations that exceeded coordination capacity (Weick and Roberts 1993). Roberts and Moore (1993) explain the Exxon Valdez disaster from the lack of multi-channel and interactive feedback communications (Roberts and Moore 1993). Weick (1993) suggested that fire fighters became trapped in a situation that was severely underestimated in advance. The fire crew was not prepared to deal with a fire that did not correspond to their normal routines. This was complicated by the combination of time pressure, palpable danger, noise, heat, and smoke (Weick 1993). Apart from two firemen, the crew did not communicate and stay together, resulting in a disintegrating organization. More recently, Snook (2000) sketched the complexity of multi service operations in the military. Different units operate with their own codes and standards. In order to coordinate across these organizations, policies have been committed to. Over time, however, organizations' practices tend to 'drift' according to Snook, meaning that they do not (fully) comply with these policies. Friendly fire becomes the result of multiple interwoven factors, such as lack of attention to standards, lack of feedback communications, and lack of training in visual recognition (Snook 2000)

This theoretical frame is applied to the data on the midair collision after presenting our methods.

3. METHODS

The research methodology relies on data sources like the investigation report (BFU 2002) and transcripts of radio transmissions (-- 2002). The research adopts a process perspective on coordination failure. Since the collision resulted in fatal injuries, no qualitative empirical research is possible as far as related to the crew. Contacting the controller would be difficult for emotional and legal reasons. We relied on articles appearing shortly after the incident in journals (like Aviation Week & Space Technology), magazines and newspapers. These resources triggered our analytical processes and fueled discussions with other researchers and aviation professionals (Habermas 1984). Gradually, this process improved our (still evolving) understanding of the complicated events leading to the midair collision.

4. DESCRIPTION AND ANALYSIS OF EVENTS

The description of events relies on the German investigation report and the transcript of voice recordings during the minutes leading to the crash (-- 2002; BFU 2002). We divide the series of events in 3 episodes. The first episode starts when flight 611 enters the Zurich radar sector, at FL 260 and climbing. The second one starts when flight BTC 2937 enters the same airspace at FL 360, flying level. The third one refers to the final 2 minutes leading to the collision. We describe and analyze each episode, in particular the third episode that captures the final 2 minutes leading to the midair collision.

4.1 Episode 1: Flight 611 enters Zurich radar sector

On July 1, 2002 at 21:20:08 hours UTC flight DHX 611 enters the Zurich south sector radar. The aircraft with cargo and 2 crewmembers on board departed from Bergamo (Italy) to Brussels (Belgium) on 13:30 flying in a northern direction. The pilots contact the air traffic controller working as the Zurich South Sector Radar Executive (abbreviated as S RE), see appendix A. Initially, the controller does not hear or understand what the DHL pilots say. That transcript indicates that he is busy what numerous other aircraft in the area (-- 2002). The pilots report the identity of their flight by the flight number and explain that they are leveling the plane at FL 260. The controller repeats the flight number and reframes their identity by indicating the squawk code2. The pilots repeat that code as well as a portion of the flight number, "611". By now, the controller knows the identity of the aircraft, where it flies, and in which direction, and the pilots know that the controller knows these things (Ayas 1996). A temporary shared understanding of current and future reality connects the controller and pilots. These people probably never collaborated on prior occasions so they do not know each other. Common knowledge and terminology helps them plan the ex ante flight path and actual flight trajectory (Weick and Roberts 1993; Grant 1996). The controller confirms identification of the aircraft on his radar screen and grants permission to climb to FL 320. The pilots repeat this information back in their own words and request permission for FL 360. The controller then confirms his understanding of

---

[2] A Squawk Code is a radio identity code that facilitates identification of the aircraft on the controller's radar screen.

the request and indicates his expectation that the pilots can climb to that level in 4 to 5 minutes. The Swiss controller grants permission at 21:26:36 hours UTC. The DHL pilots again confirm this. Mutual repetition takes a central role in the coordination process between pilots and controller. Repetition concerns the current situation or 'ist' ("flight 611"), a desirable situation ("requesting 360"), and 'soll' situations. The latter include instructions by the controller ("climb flight level 360"). Communications and the feedback loops clarify situations, expectations and request on both sides (Roberts and Moore 1993). They include the remote person in a local physical reality (where you are) or mind reality (plans). This enhances someone's understanding of that remote reality and enables him to align thoughts and actions so that overall coordination is achieved (Thompson 1967). At 21:29:50 hours UTC, flight 611 flies in a northern direction at FL 360 in accordance with the intentions and actions of the pilots and the controller.

4.2 Episode 2: Flight 2937 enters Zurich radar sector

At 21:30:10, 20 seconds later, flight 2937 enters the same sector at FL 360 (same altitude as DHL flight). The pilots identify themselves as flight BTC2937. Again the controller does not hear or understand the pilots' communications. The pilots repeat their identity and also indicate that they are flying at FL 360. The black horizontal arrows, in Appendix B, (➔) indicate that both aircraft fly level at FL 360. The controller repeats back the flight number and – like with flight 611 – adds the squawk code.

Flight 2937 enters the Zurich South radar sector at FL 360 (same altitude as flight 611) and quite close to flight 611. If aircraft are on a collision course, this means that the controller has only a brief period after the new aircraft enters his sector. He must quickly reduce risk by having one or both aircraft descend or climb, change direction, or change speed.

4.3 Episode 3: The final 2 minutes

While both aircraft are flying at FL 360, the ground based warning system[3] at the Upper Area Control Center (UACC) in Karlsruhe Germany issues a warning to local controllers. It means that while the controller and pilots assumed the aircraft flew safely, they were on a collision course. The system in Germany was the one to notice this discrepancy between assumed reality and actual reality as it was unfolding.

The collision alert system in Zurich was not available due to maintenance. STCA would have warned the Zurich controller of the impending collision risk had it been operational. The system contains algorithms that process radar information and calculate possible risks. An audio-visual warning on the radar screens assists a local controller by identifying possible collision risks. Only the German STCA and controller were aware of the risk but they could not reach the Swiss controller by any means.

72 seconds later, the collision warning systems called Traffic Alert and Collision Avoidance System (TCAS) on board both aircraft issued a warning. These systems process information from the other aircraft's transponder, and translate these into audio-visual warnings and (if aircraft get to close) instructions for the pilots.

At 21:34:42, information systems on board and in Germany 'knew' about the collision risk. A few individuals knew too: the controller in Germany, and pilots on both aircraft. In fact, the TCAS systems knew more than the pilots. Communications between these systems enabled alerts on both aircraft. The German warning system functioned as an extended set of senses for the controllers there.

Rather late, the Zurich controller observes the pending risk. He asks flight 2937 to descend to FL 350 because of "crossing traffic". This seems a reasonable solution yet the situation is already critical with less than a minute until collision. The controller's late response has increased the risk and decreased the buffer for solving the pooled dependence. This form of dependence implies here that two aircraft use the same piece of airspace (Van de Ven, Delbecq et al. 1976; Weick 1993).

The TCAS systems now switch to an active role. They instruct the pilots to climb (flight 2937) and descend (flight 611). In fact, the systems become a substitute controller (Hutchins 1990). The adjustment process changes from a human, centralized coordination approach (air traffic controller) to decentralized, machine based control (on board systems that connect). This dual transition is shown in Table 1. The columns depict centralized versus decentralized coordination. This means that a coordinator (man or machine) coordinates for actors or machines, like a traffic policeman coordinating during an accident on the highway, or traffic lights coordinating the flow of traffic.

The rows illustrate the difference between human coordination (pilots or controllers) and machine-based coordination (information systems like STCA or TCAS). Abbreviations indicate the coordination approach. In this case, the TCAS systems issue instructions at 21:34:49 (flight 2937) and a few seconds later to the pilots on board flight 611. At that time, the systems assume an independent coordinating role, implying a shift across the rows and columns in Table 1 (shown as gray cells).

The controller cannot know what the TCAS systems tell the pilots. He is probably not a pilot and does not seem to understand that pilots of aircraft flying so closely will receive instructions from the TCAS systems. Vice versa, the pilots do not seem to realize that the controller does not know that the TCAS issues instructions when aircraft get so close. The Tupolev crew never mentions the TCAS instructions to the controller even though these are contrary to the controller's instructions. The DHL pilots briefly mention the TCAS instruction 23 seconds after the first instruction, only 13 seconds before the collision. They comply with the TCAS instructions as they have been trained to. This means that they work according to the decentralized automated coordination scheme (DAC) as depicted in Table 1. Yet the Tupolev crew ignores the system and follows the controller's instructions (cell CHC). Most Russian crews follow air traffic control instructions quite strictly even though these may conflict with TCAS. The Tupolev pilots confirm and execute the controller's instruction to expedite descend to FL 350. The controller even provides some background information ""ja", we have traffic at your... 2 o'clock now at 360" (this information was incorrect since flight 611 was at the 10 o'clock position of flight 2937).

---

[3] The system is referred to as STCA, or Short Term Conflict Alert System.

The controller still assumes that 611 continues a level flight at FL 360, while 2937 would descend and fly underneath 611, contrary to TCAS instructions. In fact, to the Tupolev crew this did not matter since that crew ignored the system's warning. Yet the DHL crew followed the instructions and confirmed this back to the controller at 21:35:19. Too little time remained for the controller to realize the pending danger. He had instructed 2937 to descend while the TCAS had instructed 611 to do the same. The overall collaborative system appeared too inflexible. Too many knowledge gaps existed that constrained coordination. The TCAS does not have the flexibility to sense the conflict and adjust to pilot behaviors. It is not connected to the controller's systems. The 611 pilots did not realize the danger of the controller's instruction to 2937 to descend. They also informed the controller somewhat late. As a consequence, the coordination system fell apart in the last few minutes. The pilots followed different coordination approaches. The controller was unaware - until it was too late - of the 611 pilots' mindset and course of action (Weick and Roberts 1993). He was partly disconnected from the reality he was responsible for. The Tupolev pilots kept following the controller. They simply ignored the TCAS system and did not even mention the warnings and instructions to the controller. They remained in cell CHC (Table 1).

Table 1. Transition of coordination approaches

	Centralized coordination	**Transition between centralized and decentralized coordination**	**Decentralized coordination**
Human coordination	**CHC** Centralized human coordination: Air traffic controller mediates traffic		**DHC** Decentralized human coordination: Pilots interact directly
Transition between human and automated coordination			
Automated coordination	**CAC** Centralized automated coordination: Air traffic control system would coordinate traffic		**DAC** Decentralized automated coordination: On board TCAS

The DHL pilots on the other hand immediately complied with the TCAS system (cell DAC). They mentioned this 23 seconds afterwards. Until then the controller assumed that 611 flew level since nothing was communicated. This assumption did not match reality. The controller also assumed that 2937 would descend. This happened, but with disastrous consequences. The crews seemed to realize the nearing collision and tried last minutes escapes (cell DHC in Table 1). The night made this strategy unfeasible unlike an earlier near midair collision in Japan in 2001 where the crews successfully avoided each other by adjusting just in time and flying by with only 20 meters in between (ARAIC 2002). Overall, this does not seem a professional coordination strategy when so much is at stake.

5. Key findings

The research found that coordination failure started small and was perceived at a very late stage. The controller was distracted and worked without adequate support from peers and on-site warning systems.

The controller's capacity for handling multiple aircraft and sectors seems limited. The reason for this is not yet clear. Maybe he was tired, or concerned about other aircraft.

As episode 2 illustrates, the second aircraft enters at the same altitude. This reduces opportunities for response considerably. With the speeds of modern aircraft, a few minutes are sufficient for a potential collision.

The second episode also shows that network centric operating depends on connectivity among intelligent actors (Alberts, Garstka et al. 2000). The warning system in Germany was connected to a controller there. But that person could not contact the Zurich controller. Distributed intelligence thus only leads to coordinated performance if there are multiple linkages (Weick and Roberts 1993).

One of the interesting elements is that the pilots depended completely on a controller who *thinks* he is in control. For minutes, there is a discrepancy between assumed reality and time/ space reality. The problem with this dependence and discrepancy is that no one does anything even though the time available for avoiding disaster shrinks every second. This waiting, in turn, increases coupling and complexity with at some point only seconds before the collision (Perrow 1984; Rijpma 1997).

The final few minutes, multiple events complicate the situation so rapidly and severely that coordinated response is virtually impossible. A dispersed HRO in this stage cannot escape deliberately failure. Only an accidental near miss would do (ARAIC 2002). Subsequently, systems on board issue warnings to the pilots, while the controller realizes the pending coordination problem. In fact, machines were much faster than people. Starting with the German warning system, the on board systems 'know' about the pending failure ahead of the pilots.

The complication is that the on board systems temporarily substitute for the controller without him realizing it. This resembles coordination problems between cabin and cockpit crews who do not share training and understand little about each other's jobs (Wiener, Kanki et al. 1993).

Simultaneously, the diversity between the Russian and DHL crews translate into different response patterns. Diversity increases coordination efforts. It requires elaborate communications to identify and resolve differences (Lawrence and Lorsch 1967; Dougherty 1992). Coordination would have to shift from standards and structures towards the collaboration process itself (Mintzberg 1979). No time was left for that to happen.

6. Discussion and conclusion

The objective of this research was to increase our understanding of the midair collision on July 1, 2002. We conceived the collision as a coordination failure in a HRO. Factors that contributed to the failure include knowledge gaps, diverse ways to deal with the TCAS, unawareness on the controller's side of the role of TCAS, (partial) transitions between different coordination systems (Table 1), disconnection from reality, not sharing local realities, and allowing risk to build so that coordination must be achieved last minute.

More research is needed to understand factors contributing to the collision and the vulnerability of the distributed collaborative system of men and machines (Weick 1993). This future research could compare similar incidents (midair collisions), failure in emergency services, friendly fire, and earth to orbit failures (NASA 1999). Distributed coordination appears vital in today's globalizing societies that rely on people collaborating with sophisticated technologies.

7. REFERENCES

[1] -- (2002). Transcript of voice recordings midair collision July 1, 2002, file 'protokoll' available from http://www.sonntagszeitung.ch/pdf/protokoll.pdf.

[2] Alberts, D. S., J. J. Garstka, et al. (2000). *Network Centric Warfare: Developing and Leveraging Information Superiority.* Washington DC, CCRP Publication Series.

[3] ARAIC (2002). Aircraft Accident Investigation Report - Japan Airlines Boeing 747-400, JA8904 (A Near Midair Collision to a Douglas DC-10-40 of Japan Airlines, JA8546) (Report 2002-5). Tokyo, Aircraft and Railway Accidents Investigation Commission (Air Subcommittee), Ministry of Land, Infrastructure and Transport.

[4] Ayas, K. (1996). "Design for learning and innovation." *Long Range Planning* **29**(6): 898-901.

[5] BFU (2002). Status Report (Midair collision 1st July 2002 near Überlingen/ Lake Constance). Braunschweig, Bundesstelle für Flugunfalluntersuchung (BFU). Document # AX001-1/-2/02 (English version). Available from http://www.bfu-web.de.

[6] Bigley, G. A. and K. H. Roberts (2001). "The incident command system: High reliability organizing for complex and volatile task environments." *Academy of Management Journal* **44**(6): 1281-1299.

[7] Dougherty, D. (1992). "Interpretive barriers to successful product innovation in large firms." *Organization Science* **3**(2): 179-202.

[8] Grant, R. M. (1996). "Toward a knowledge-based theory of the firm." *Strategic Management Journal* **17**(Winter): 109-122.

[9] Habermas, J. (1984). *The Theory of Communicative Action.* Cambridge, UK, Polity Press.

[10] Hutchins, E. (1990). The technology of team navigation. *Intellectual teamwork: Social and technological foundations of cooperative work.* J. Galegher, R. E. Kraut and C. Egido. Hillsdale, New Jersey, Lawrence Erlbaum Associates.

[11] Lawrence, P. R. and J. W. Lorsch (1967). *Organization and environment: Managing* differentiation and integration. Boston, MA, Harvard University Press.

[12] Malone, T. W. and K. Crowston (1994). "The interdisciplinary study of coordination." ACM Computing Surveys **26**(1).

[13] Mintzberg, H. (1979). *The Structuring of Organizations.* Englewood Cliffs, N.J., Prentice-Hall.

[14] NASA (1999). Mishap Investigation Board, Mars Climate Orbiter, Phase I Report. Internet: ftp://ftp.hq.nasa.gov/pub/pao/reports/1999/MCO_report.pdf (Downloaded November 1999).

[15] Orlikowski, W. J. (2002). "Knowing in Practice: Enacting a Collective Capability in Distributed Organizing." *Organization Science* **13**(3): 249-273.

[16] Perrow, C. (1984). *Normal Accidents.* New York, Basic Books.

[17] Rijpma, J. A. (1997). "Complexity, tight-coupling and reliability: Connecting Normal Accidents Theory and High Reliability Theory." *Journal of Contingencies and Crisis Management* **5**(1): 15-23.

[18] Roberts, K. H. and W. H. Moore (1993). Bligh reef dead ahead: The grounding of the Exxon Valdez. *New Challenges to Understanding Organizations.* K. H. Roberts. New York, Macmillan.

[19] Rochlin, G. I., T. R. LaPorte, et al. (1987). "The Self-designing High-reliability Organization: Aircraft Carrier Flight Operations at Sea." *Naval War College Review* **40**(4): 76-90.

[20] Schulman, P. R. (1993). The analysis of High Reliability Organizations. *New challenges to understanding organizations.* K. H. Roberts. New York, Macmillan.

[21] Snook, S. A. (2000). *Friendly Fire: The Accidental Shootdown of U.S. Black Hawks over Northern Iraq.* Princeton, NJ, Princeton University Press.

[22] Thompson, J. D. (1967). *Organizations in action,* McGraw-Hill.

[23] Van de Ven, A. H., A. L. Delbecq, et al. (1976). "Determinants of Coordination Modes Within Organizations." *American Sociological Review* **41**(April): 322-338.

[24] Weick, K. E. (1993). "The collapse of sensemaking in organizations: The Mann Gulch disaster." *Administrative Science Quarterly* **38**: 628-652.

[25] Weick, K. E. (1993). The vulnerable system: An analysis of the Tenerife air disaster. *New Challenges to Understanding Organizations.* K. H. Roberts. New York, Macmillan.

[26] Weick, K. E. and K. Roberts (1993). "Collective mind in organizations: Heedful interrelating on flight decks." *Administrative Science Quarterly* **38**: 357-381.

[27] Weick, K. E., K. M. Sutcliffe, et al. (1999). Organizing for high reliability: Processes of

collective mindfulness. *Research in organizational behavior*. B. Staw and L. L. Cummings. Greenwich, CT, JAI Press. **21:** 81-123.

[28] Wiener, E. L., B. G. Kanki, et al. (1993). *Cockpit Resource Management.* San Diego, Academic Press.

---

[i] This paper is exclusively intended to analyze from an academic point of view the midair collision over Germany on July 1, 2002. It contributes from this perspective to the improvement of flight safety. The paper is not supported by and is not tied to any of the parties involved, nor does it substitute for official investigations such as those conducted by the German Bundesstelle für Flugunterfalluntersuchung (BFU 2002). The paper is not intended to play a role in legal or criminal investigations, nor can the authors accept any liability related to any part of this paper or research.

A Study of the Link between Trust and Use of Adaptive Cruise Control

B. Rajaonah
CNRS-LAMIH-PERCOTEC
UVHC-Le Mont Houy
F-59313 Valenciennes
Bako.Rajaonah@univ-valenciennes.fr

F. Anceaux
CNRS-LAMIH-PERCOTEC
UVHC-Le Mont Houy
F-59313 Valenciennes
Francoise.Anceaux@univ-valenciennes.fr

S. Espié
INRETS-ARCUEIL
2 Avenue du
G^{al} Malleret Joinville
F-94114 Arcueil
Espie@inrets.fr

J.-M. Hoc
CNRS-IRCCyN
PsyCoTec
1 Rue de la Noë
F-44321 Nantes
Jean-Michel.Hoc@irccyn.ec-nantes.fr

ABSTRACT

In this paper, we describe two studies concerning the link between trust and the use of Adaptive Cruise Control (ACC). The first study employed a questionnaire designed to analyze *a priori* trust, or the trust existing prior to any interaction with ACC. The results of this analysis show that prior to interaction with ACC, *a priori* trust is moderate, with drivers only moderately inclined to use the device. The second study involved experimentation using a driving simulator. The main results of the experimentation indicate that as drivers become familiar with ACC, trust increases significantly, and drivers actually use the device more often than their original intentions would have suggested.

Keywords

Adaptive Cruise Control, driving simulator, human-machine cooperative system, questionnaire, trust, use of automation.

1. INTRODUCTION

The present paper describes two studies organized within the framework of ARCOS[1], a French national ground transportation research program. ARCOS supports research in 10 areas. Our project falls into the sixth category, Human-Machine Systems, which examines the relationships between drivers, automation, vehicles and the environment. Our first study analyzed driver *a priori* trust for automatic control, via questionnaires. Our aim was to determine the level of trust existing prior to interaction with automatic control devices, in our case, adaptive cruise control. The second study, a collaborative effort by LAMIH-PERCOTEC[2] and INRETS-CIR-MSIS[3], re-evaluated trust levels following training on ACC.

2. THEORETICAL BACKGROUND

The concept of a human-machine cooperative system refers to a system comprised of human and artificial agents, in which agent functions are not determined rigidly in advance, but rather are allowed to adapt dynamically to the constraints of complex and uncertain environments [5,10]. The definitive characteristic of such a cooperative system is its adaptability, which allows it to complete tasks when the environmental conditions make the usual procedure impossible [9]. The various agents use information from the system and from the environment to make an endless series of compromises between performance, safety and resources [1,2]. The human operator, however, always retains final authority for overall task organization [5,6]. Given these circumstances, trust is an intermediate variable that must be taken into account because it influences the operator's decision to use, or not to use, the automation system [16,20]. Indeed, the dynamics of the various situations in which human - machine systems evolve prevent operators from knowing, in real time, what they need to know in order to act. Thus, these systems need an internal mechanism that will allow operators to reduce the feeling of uncertainty and risk related to the possible consequences of their decisions [14].

Our definition of trust, based on a review of the existing literature, is adapted to the analysis of decision-making situations encountered during human - machine interactions in which human operators can choose whether or not to delegate function(s) to the automated part of the system. We assume that this choice depends on trust, which we define as the psychological state resulting from knowledge and beliefs about a situation as well as the features of that situation, and which creates positive expectations for human-machine interaction. High levels of trust lead operators to choose to use automation. Trust levels depend on the mental representations that operators create of individual situations, as they balance the perceived advantages and disadvantages of using automation to manage such a situation. The major disadvantage appears to be the risk that the results will not match the expectations, thus obliging the operator to intervene after all.

In this study, the trust referent is the interaction between the human operator and the automated device [3]: trust varies in relation to operator self-confidence, the trust operators have in the automated device, and the degree of competence with which each agent carries out the task. Our premise is that trust in human-machine interaction may be based on:

a) metacognitive knowledge, which concerns the cognitive functions of both the human and machine agents;
b) knowledge about the machine, which includes information about predictability and dependability [15,16] as well as familiarity, understandability and the ability to explain intent [22];
c) situation awareness, which covers awareness of elements, such as the available time and resources and the performance of the human-machine system.

Every time the situation changes, these different considerations come into play as the operators update their representations [7], and decide whether or not to use the automated system. This is one of the reasons why human-machine cooperative systems are so highly adaptive. In turn, trust evolves with the operator's representation of the situation.

Trust evolves on both a long-term and a short-term basis. In the short term, trust evolves according to the local performance of the automated system [15,16] and the operator's performance as well [12,13]. Trust also evolves over a longer term, as the operators elaborate mental models for themselves and the artificial agent [6,24] and learn to interact with that agent [19]. Numan [17] has proposed a model of trust evolution that describes this evolution in both temporal spans. In general, trust ranges from the blind faith that has no base in interactions with the referent, to the absolute certainty that occurs when the operator, perceiving no risk, is sure of what is going to happen. In the short term, for a given time (T1), trust is based upon empirical facts, and either increases or decreases locally depending on the consequences of the events occurring in T2, T3, T4 and so on. Clearly, this trust is fragile and is directly affected by performances of the automated device [16]. However, in the long term, trust tends to increase : according to Numan [17], every individual aims for the level of absolute certainty, and once trust has been established, it never completely disappears, even though it can be altered temporarily.

For Numan [17], the foundation of trust comes from external sources, for example those who already use the system or those who would promote it use. This idea is similar to the one proposed by Kramer [11] who states that, in the absence of direct information from the referent, trust can be based on the opinions, or role, of a third party. Numan [17] distinguishes two types of trust: *trust 1* which is based on indirect information, and *trust 2* which is based on direct information or empirical facts. The first study presented in this paper deals with the first type of trust, which we call *a priori* trust, and the second considers the second type, called *a posteriori* trust.

Adaptive Cruise Control (ACC) is an electronic device with sensors that are positioned on the front of the vehicle. This device calculates the vehicle's relative speed as well as the distance between this vehicle and a target vehicle, which is just ahead in the same lane or is attempting to enter the same lane. Drivers can determine in advance the speed and the distance that they would like to maintain; and can also deactivate the device at any time in order to resume manual control of the vehicle [21,23]. Because driver intervention is necessary in potentially dangerous situations that exceed the ACC's technical capabilities, it becomes important to investigate the environmental and psychological conditions leading to a driver's decision to use the ACC, or not. In our study, the relationship between the driver and the ACC device is considered as cooperation between two agents. We analyzed the connection between the level of driver trust in the ACC and the delegation of speed and distance regulation tasks to the ACC.

3. METHOD

3.1 The first study: a questionnaire concerning *a priori* trust

Two-hundred and fifty-six people participated this study, 192 women and 64 men. Their average age was 25.6 years, with a minimum age of 18, a maximum of 64, and a standard deviation of 8.21.
The questionnaire contained 55 items related to interpersonal confidence, trust in new technologies, the advantages and disadvantages of ACC, and the respondents intentions with regard to using the device. Each participant received a list of the ACC principles prior to completing the questionnaire.
After reading the principles, the participants were asked to position their response to the content of the item by marking a cross on a 10-cm line.

For example: Do you think trust is necessary to use new technologies?

No Yes

3.2 The second study: tests on a driving simulator

This study is a collaborative effort of LAMIH-PERCOTEC[2] and INRETS-CIR-MSIS[3].

3.2.1 *Participants*

The experimentation is not yet complete. Twenty-four people are expected to participate, though not all have actually done so at the moment, so our comments in this article are based on an analysis of the data produced by the first 12 participants. The average age of the participants was 22.9 years, with a minimum age of 19, a maximum of 29, and a standard deviation of 2.9.

3.2.2 *Material*

3.2.2.1 *Driving simulator*

We used a mini driving simulator (L=1m80; l=60cm; H=160cm), adapted for the experiment by INRETS-CIR-MSIS[3], and equipped with the following basic equipment : a starter; clutch , brake and gas pedals; a gear box; a hand brake; the left komodo; and a dashboard with a speedometer (Figure 1). We also used a projection monitor, a projection screen and two computers for the software simulation.

In addition, we used a variety of recording equipment — video cameras, microphones, video recorders — in order to document the position of the drivers' feet and that of the left hand when activating or deactivating the ACC; the simulated traffic projected on the screen; the experimental times appearing on one of the computer screens; and the drivers' spontaneous verbalizations when driving.

Figure 1. A view of the driving simulator.

3.2.2.2 *Adaptive Cruise Control*

The ACC reference speed was set at 130km/h, and the reference Time Headway (THW) at 1.5s. The ACC has a maximal braking capacity of 3-m/s^2. Brake pedal values range from 0 to 255, and the ACC is automatically deactivated when the driver uses the brake pedal above a value of 130. On the other hand, use of the gas pedal does not deactivate the ACC. The driver, however, can consciously deactivate or reactivate the ACC by using the left komodo.

3.2.2.3 Simulation software

The driving circuit consisted of about 51 kilometres of standard highway with 2x2 lanes.

Each participant completed 5 runs :

- A 5mn run to become familiar with the driving simulator ;
- A 5mn run to become familiar with ACC ;
- Two ≅25 mn experimental runs (Run 1 and Run 2) of driving in clear weather. Each run included the same 14 scenarios in the same order (for example, decelerations and accelerations of the lead vehicle or the insertion of a truck in front of the participant's vehicle). A sample scenario is shown in figure 2.
- One ≅25 mn experimental run (Run 3) of driving in fog. This run included the same scenarios as Run 1 and Run 2, in the same order.

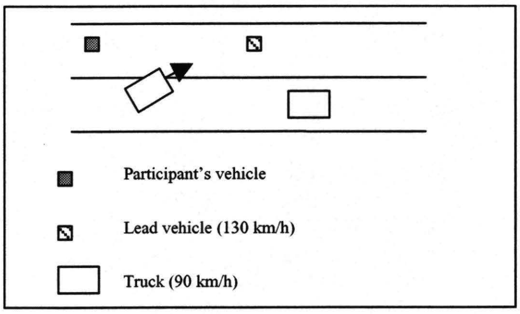


Figure 2. A sample scenario : insertion of a truck in front of the participant's vehicle.

3.2.2.4 The Presentation of ACC to the participants

We created two types of Power Point presentations explaining what an ACC is, using pictures found on the web.

- The first was a detailed presentation of the ACC, with an explanation of the different possible driver manipulations of the device, and of how the ACC works and what its limits are. This presentation included 10 animated and commented slides.
- The second was a non-detailed presentation, describing what an ACC is and what the possible driver manipulations of the device are. This presentation included only 5 of the 10 slides shown in the detailed presentation.

3.2.2.5 Questionnaires

The second study also involved questionnaires. Three questionnaires were used: one to classify the driver personal characteristics and driving habits (QI); one to analyze *a priori* trust before familiarization with ACC (QII - a reduced version of the one used in the first study); and one to analyze the evolution of trust after familiarization with ACC (QIII).

3.2.3 Procedure

Each participant completed the experimental procedure individually.

- Participants began the experiment by responding to questionnaire QI.
- The experimentator started the ACC Power Point presentation. The type of presentation (detailed or non-detailed) was counterbalanced with each participant watching a randomly selected presentation type.
- Participants responded to questionnaire QII.
- Participants completed the two familiarization runs on the mini simulator. On the second run (with ACC), they were told to activate and deactivate the ACC regularly in order to test the device.
- Participants responded to questionnaire QIII.
- Participants completed the three experimental runs, responding to a QIII questionnaire after each one. The ACC was automatically activated at the beginning of each run. The instructions to participants were (1) to follow the lead vehicle and (2) to deactivate and reactivate ACC as desired.

4. RESULTS AND DISCUSSION

The data for the first study comes from the responses to the questionnaire.

The data for the second study comes from:

- The questionnaire responses (questions and scales).
- The spontaneous verbalisations recorded during the experimental runs.
- The driving performance measurements that were recorded automatically every second during the experimental runs: situation number, experimental time, ACC activation state, speed of participant and lead vehicles, the lateral position of the participant and lead vehicles, THW and TTC values, brake and acceleration values (from 0 to 255) for the participant vehicle.

4.1 The first study

The results from the first study were analyzed using XLSTAT 5.2. The position of the cross corresponding to the participant's response was measured from 0 to 10-cm, and the mean of these measurements was calculated. In the two results sections, we report the mean scores and the standard deviations: ms (sd).

4.1.1 The trust referent for use of technology

Two questions were asked concerning the referent of trust for the use of new technology:

> When speaking of trust with regard to the use of a technological object, the phrase which best describes this trust would be:
>
> Q15. Trust in the object: 5,29 (2,54) ;
>
> Q16. Trust in the interaction with the object: 6,38 (2,33).

The mean score for question Q16 is significantly higher than the one for Q15 ($t_{251} = 4{,}784$; $p < 0{,}0001$): trust in the interaction with a technological object thus would appear to better describe participant trust than trust in the object. Even from a general point of view, when we question participants directly without referring to a particular technological object, trust has for referent the interaction between the user and the object and not only the object. Given such referent, it would seem important to analyze other elements in addition to the object properties.

4.1.2 Advantages and disadvantages of ACC

We asked several questions about the perceived advantages and disadvantages of ACC.

4.1.2.1 Perceived advantages

For the participants, the first advantage of an ACC was safer driving: 6,98 (2,35). This was followed by less stressful driving: 5,34 (2,67) and improved driver comfort: 5,25 (2,65). Some participants also indicated that ACC made the driving task easier: 4,65 (2,71) and made the driving task less burdensome because there was less monitoring: 4,27 (2,72).

These results are important in so far as they are the objectives targeted by the designers of this type of device. ACCs are supposed to decrease the driver workload and improve safety [8]. It is probably these advantages which led the participants to recognize the usefulness of ACC. The following are some of the responses to questions concerning the usefulness of an ACC:

- safer driving : r =0,63267 (p<0,00000) ;
- improved driver comfort : r =0,56209 (p<0,00000) ;
- less stressful driving : r =0,45759 (p<0,00000) ;
- an easier driving task : r =0,42606 (p<0,00000) ;
- a less burdensome driving task =0,35179 (p<0,00000).

4.1.2.2 Perceived disadvantages

The perceived disadvantages of an ACC mainly involved the risk of becoming dependant on the ACC: 5,42 (3,00). Others included the risk of reduced driving pleasure: 5,07 (3,18) and risk of interference with regular driving behaviors: 4,90 (2,91). This last result is consistent with previous research done in collaboration with INRETS[3] and Renault [19], in which we clearly showed that ACC could interfere on three levels: (i) ACC could prevent drivers from maintaining a preferred speed, (ii) ACC could hamper the other road users, which could in return (iii) hamper the driver of the ACC-equipped vehicle.

4.1.3 Intent to use an ACC

One question concerning the intent to use an ACC was essential for our purposes.

> Q54. If the vehicle that you drive regularly was equipped with an ACC, would you use it often: 5,09 (2,91).

For many authors [4,12,13,15,16], trust is an essential factor in the use of automatic control, notably for ACC [23]. If we define trust as a psychological state which leads to the decision to use ACC, or not to use it, then trust could be based on the perceived advantages and disadvantages associated with the use of ACC. We analyzed the correlation between the score on question Q54 and the scores on the questions about the perceived advantages and disadvantages of an ACC.

Intent to use is significantly correlated with:

- utility (r=0,60721 ; p <0,00000) ;
- safer driving (r =0,51177 ; p <0,00000)
- improved driver comfort : (r=0,48429 ; p<0,00000) ;
- less stressful driving : (r=0,47945 ; p<0,00000) ;
- an easier driving task : (r=0,41027 ;p<0,00000) ;
- a less burdensome driving task: (r=0,30029 ; p<0,00000) ;

Intent to use is significantly correlated to the following disadvantages:

- interfering with regular driving behavior : (r = - 0,57376 ; p < 0.00000) ;
- reduced driving pleasure : (r = - 0,44842 ; p < 0.00000).

Those results corroborate the link between intent to use and the perceived advantages and disadvantages. We believe that these elements characterize the anticipated performance of the driving task which results from the imagined interaction between the driver and the device. These elements should be distinguished from the properties of ACC.

4.1.4 The characteristics an ACC should have

We asked the question: "What are the qualities that an ACC must have for you to agree to use it often?" The following responses were given.

Q47. Ease of use: 7,72 (1,84);
Q48. Easily understandable technical functions: 7,71 (1,85);
Q49. A guarantee that it will work correctly: 8,05 (1,96);
Q50. Installed by a reputable car manufacturer: 5,06 (2,83).

Unsurprisingly, ease of use, easily understandable technical functions, and a guarantee that the ACC will work correctly are the principal conditions for accepting to use ACC. While the car manufacturer's reputation is not unimportant, it is less important for most participants. In our opinion, the parts of questions Q49 and Q50 concerning guarantees and reputations refer to the external sources highlighted by Numan [17]. Therefore, it is normal that the scores are relatively high for these two items, given that the participants had no experience with ACC. Before interacting with ACC, before testing the limits of its performance, before forming an opinion based on their own subjective feelings, participants could only take into account external information sources: use instructions, third-party explanations, etc. Actual guarantees, or reputations that imply such guarantees, become more important, acting to strengthen preliminary judgments based on the technical descriptions and helping to build trust in ACC.

4.1.5 A priori trust and intent to use an ACC

This section examines the participants' intention to use the ACC often if their regular vehicle is equipped with it (Q54 = 5,09 (2,91).

Despite high correlations with the items concerning the perceived advantages of the ACC, intent to use ACC remains moderate. We divided the participants into three groups:

a) " minimal use " for those with scores ranging from 0 to 3,35 : N=66 ;
b) " moderate use " for those with scores ranging from 3,40 to 6,70 : N=85 ;
c) " heavy use " for those with scores ranging from 6,75 to 10 : N=74.

The difference between the proportions of those with minimal intentions to use ACC and those with moderate intentions is significant (z = - 2,126; p < 0.017), whereas the difference between those with moderate intentions and those who intended to use ACC heavily is not significant (z = 1,234; ns).

For the 66 participants in the "minimal" category, *a priori* trust appears to have been insufficient, and thus participants did not intend to use ACC. For the 74 participants in the "heavy use" group, the disadvantages seem to weigh less than the advantages. For these participants, *a priori* trust appears sufficient for them to intend to use ACC often. For the 85 participants in the "moderate use" group, two possible explanations come to mind, though there may be others: (1) the advantages and disadvantages have the same weight in the decision to use the ACC, and thus participants were sure that their use would be moderate; (2) participants were hesitant because the information they had about ACC was insufficient for making a decision concerning the use of ACC.

Even if the participants of the three groups had no information coming from practice with ACC, for those in the minimal and heavy use groups, the theoretical information seems to have been sufficient for them to know what they would do if their cars were equipped with an ACC. Nonetheless, many authors judge that only familiarization with automatic control allows trust to be established [15,16,24], and this appears to be the case for the "moderate use" group . They were ready to try ACC, but *a priori* trust was insufficient for them to decide to use it often.

Clearly, it will be *a posteriori* trust, established with practice, that will have the most influence on ACC use frequency. The trust of those individuals who become familiar with ACC by trial and error is the most likely to be modified. Those who will not try it, however, will have more difficulty changing their attitudes toward ACC, given that they will receive no feedback allowing them to modify their current representations. The second study presented here concerns this leap from *a priori* trust to *a posteriori* trust - or *trust 2* for Numan [17].

4.2 The second study

For that data analysis in this study, we are using the statistical software, XLSTAT 6.0. Because the analysis is just beginning, we can only present partial results here.

The hypotheses for the analysis presented in this paper are as follows:

- There is a difference between *a priori* trust and trust after practice with ACC.
- The intent to use ACC is different from the actual use of it.
- Trust evolves with experience on an ACC.
- The type of weather conditions (without or with fog) has an effect on trust and on the use of ACC.
- The type of introductory presentation about ACC has an effect on trust.

4.2.1 From a priori trust to a posteriori trust

Just after the introductory presentation about ACC, participants responded to a questionnaire about *a priori* trust (QII). The question, "What level of trust would you accord to an ACC?" (QII-16), was added to the questionnaire used in the first study. As for the first study, the mean scores and standard deviation are presented in cm (from 0 to 10). The mean score for this question was 5,19 (1,58), reflecting a moderate level of trust. In order to discover how trust in ACC evolved with practice on the device, we asked the following question four times, once after the test run with ACC and once after each of the three experimental runs: "Do you feel confident when you drive with an ACC?" (QIII-22).

Even though QII-16 and QIII-22 are not formulated in the same way, the responses show the evolution of trust, by situating trust levels before and after practice with the device. We present these results in Table 1 and Figure 3.

Table 1. Mean scores of trust

	QII16	QIII1-22	QIII2-22	QIII3-22	QIII4-22
	After ACC presentation	After familiarization with ACC	After Run 1	After Run 2	After Run 3
Mean score (-cm)	5,19	5,58	6,07	6,4	6,73
Standard deviation	1,58	2,02	2,18	2,72	3,03

According to the trust evolution curve in Figure 3, trust has a slight tendency to increase with ACC practice, even though Friedman's ANOVA doesn't show significant differences between the runs.

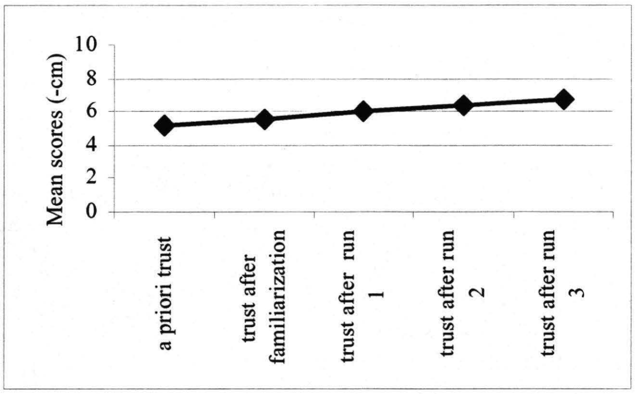


Figure 3. Trust evolution curve.

4.2.2 From intent to use and actual use of ACC

As in the first study, we asked a question about the intent to use ACC. The mean score for this question (QII-38) is 5,00 (1,25).

For each run, we measured the ratio (-%) between the time during which the ACC was activated and the total duration of the run. The data for each experimental run was automatically collected every second, and stored in data files that could be analyzed using Excel. By transforming the mean score of QII-38 into a percentage, we were able to gather all the results in a single table, Table 2, which is represented graphically in Figure 4.

Table 2. Mean percentages of intent to use and actual use of an ACC

	QII-38	%ACC Run 1	%ACC Run 2	%ACC Run 3
	Inttent to use after ACC presentation			
Mean percentage	47,88	81,78	89,57	82,59
Standard deviation	12,2	22,24	6,23	29,61

Temporarily considering "intent to use" to be the same dependant variable as "actual use of ACC", we were able to use Wilcoxon's signed test. This test shows that use of ACC during the first experimental run is significantly higher than the declared intent to use the device (one-tailed p<0.002), and there is no significant difference between use during the three experimental runs: run 1, run 2 and run 3.

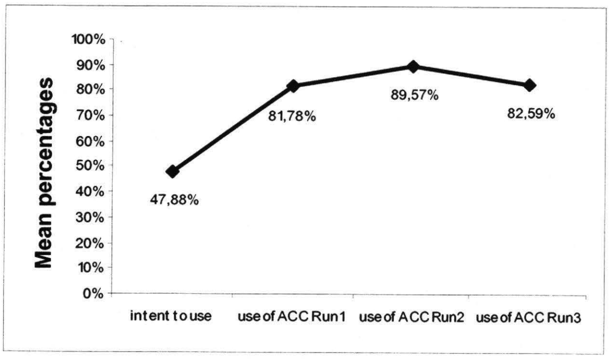


Figure 4. Evolution of use of ACC.

It is important to remember that the participants were asked to test the ACC by activating and deactivating the ACC frequently during the familiarization run, though during the experimental runs, they were free to activate or deactivate ACC as desired. The familiarization run was thus very important because it forced participants to test the device.

Even though the use of ACC seems to be slightly less important when driving in fog (figure 4), figure 3 would seem to indicate that trust doesn't decrease. Two participants said in the spontaneous verbalizations that they were very stressed, that they didn't like to drive in fog and that they preferred to concentrate on the driving task. One of these two participants spent 1,15% of the time on the run with fog using ACC, and the other spent 44,69%. When examining their respective trust scores after this run, we can see that they did not feel confident when driving with ACC (5,40-cm and 0,05-cm). These scores corroborate Muir and Moray's [16] observations about the relationship between trust in automated devices and the monitoring of them: the less the participants trust ACC, the more they want to monitor it, so they prefer to deactivate it in order to avoid having to monitor it.

4.2.3 *Effect of ACC presentation on trust*

Participants were divided into two groups, depending on which introductory ACC presentation they saw: detailed (D) or non-detailed (ND). Figure 5 presents the tendency curves of their respective trust evolution.

Even though Mann-Whitney test shows no significant difference between the two groups, figure 5 illustrates that the participants who watched a detailed presentation tended to be slightly more confident than those who watched the non-detailed presentation. This difference is reduced with practice on the ACC.

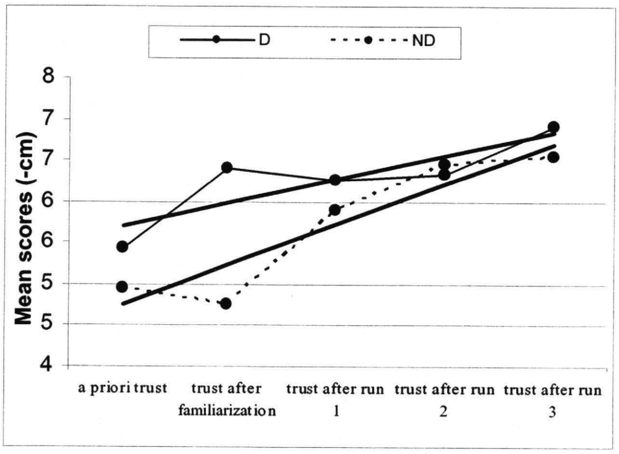


Figure 5. Effect of ACC presentation on trust.

4.2.4 *Intention of use and use of ACC real driving situation*

It appears that the intent to use ACC prior to practice on the system doesn't allow predictions to be made concerning the actual use of ACC in simulated driving situations. We wondered what the intent to use ACC in real driving situations would be after practice on the ACC. The questionnaire QIII, completed after the familiarization and experimental runs, repeated the QII question about the intent to use an ACC if the vehicle was equipped with it (QIII-26). As figure 6 demonstrates, intent to use ACC in real driving situations remains as moderate as in our first study.

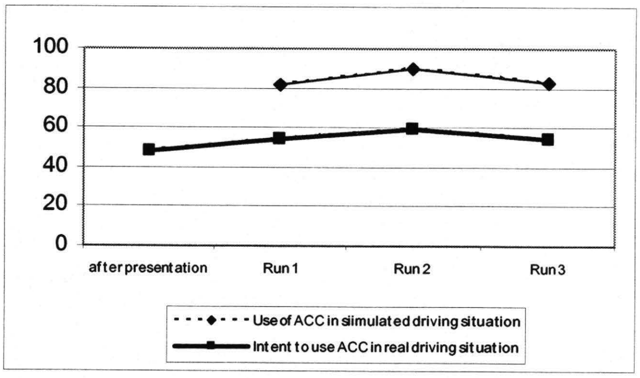


Figure 6. Use of ACC in simulated driving situations and intent to use ACC in real driving situations.

5. CONCLUSION

The two studies presented in this paper try to show how trust is built and how it evolves. Our results show that trust has a tendency to increase over the long term even if specific events can reduce it in the short term. Because increasing trust seems to be a long-term project, it is important to let drivers build trust by allowing them to test ACC.

This work is not perfect, and many questions still have no response. For example: what is the correlation between trust and ACC use? We hope to provide a response when all the data from the second study has been completely analyzed. Still, it will be necessary to experiment with other scenarios and other ACC attributes in both simulated and real driving situations.

6. ACKNOWLEDGMENTS

This work was done within the framework of the ARCOS driving safety program with the financial support of the French Ministry of Transportation.

Our thanks to Fabrice Vienne and Isabelle Aillerie (INRETS-CIR-MSIS^3) for the scenario module, the 3D-imaging module and the data recorder module.

Our thanks to Marie-Pierre Pacaux-Lemoine and Philippe Simon (LAMIH-SHM^4) for their kindness and for installing the experimental equipment.

ARCOS^1 is the French acronym for Program for Research into Driving Safety.

LAMIH^2 is the French acronym for the Laboratory of Industrial and Human Automation, Mechanics and Computer Science.

PERCOTEC^2 is the French acronym for Cognitive Psychology and Ergonomics.

INRETS^3 is the French acronym for the National Institute for Transportation and Safety Research.

SHM^4 is the French acronym for Human Machine Systems.

IRCCYN^5 is the French acronym for the Institute of Cybernetics and Communication Research.

7. REFERENCES

[1] Amalberti, R. (1996). *La conduite des systèmes à risques.* Paris : Presses Universitaires de France.

[2] Amalberti, R. (2001). La maîtrise des situations dynamiques. *Psychologie Française, 46,* 107-118.

[3] Cohen, M.S., Parasuraman, R., Freeman, J.T. (1998). *Trust in decisions aids : a model and its training implication.*http://www.dodccrp.org/Proceedings/DOCS/wcd00000/wcd000de.htm.

[4] Dassonville, I., Jolly, D., Desodt, A.M. (1996). Trust between man and machine in a teleoperation system. *Reliability Engineering and System Safety, 53,* 319-325.

[5] Hoc, J.-M. (2000). From human-machine interaction to human-machine cooperation. *Ergonomics, 43,* 833-843.

[6] Hoc, J.-M. (2001). Towards a cognitive approach to human-machine cooperation in dynamic situations. *International Journal of Human-Computer Studies, 54,* 509-540.

[7] Hoc, J.-M., & Amalberti, R. (1995). Diagnosis: some theorical questions raised by the applied research. *Current Psychology of Cognition, 14,* 73-101.

[8] Hoedemaeker, M., Brookhuis, K.A. (1998).Behavioural adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F, 1,* 95-106.

[9] Hollnagel, E. (1993). *Human reliability analysis : context and control.* London : Academic Press.

[10] Hollnagel, E., & Woods, D.D. (1983). Cognitive systems engineering : new wine in new bottles. *International Journal of Huhuman-machine Studies, 18,* 583-600.

[11] Kramer, R.M. (1999). Trust and distrust in organizations : emerging perspectives, enduring questions. *Annual Review of Psychology, 50,* 569-598.

[12] Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in huhuhuman-machine systems. *Ergonomics, 35,* 1243-1270.

[13] Lee, J., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40,* 153-184.

[14] Luhmann, N. (2000). Familiarity, confidence, trust : problems and alternatives. In D. Gambetta (Ed.), *Trust – Making and breaking cooperative relation*, electronic edition, Department of Sociology, University of Oxford, chapter 6, pp. 94-107, http://www.sociology.ox.ac.uk/papers/luhmann94-107.pdf.

[15] Muir, B.N. (1994). Trust in automation : Part I. Theorical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37,* 1905-1922.

[16] Muir, B.M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39,* 429-460.

[17] Numan, J.H. (1998). *Knowledge-based systems as companions. Trust, human computer interaction and complex systems*, electronic edition, University Library Groningen, pp. 30-56, http://www.ub.rug.nl/eldoc/dis/management/j.h.numan/

[18] Parasuraman, R., & Riley, V. (1997). Humans and automation : use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

[19] Rajaonah, B. (2001). La relation de confiance entre le conducteur et un système automatique de régulation de vitesse. *Communication présentée à EPIQUE 2001: Premières Journées d'Étude en Psychologie Ergonomique.* Nantes, France, Octobre

[20] Riley, V. (1996). Operator reliance on automation : theory and data. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance. Theory and applications* (pp. 19-35). New Jersey : Lawrence Erlbaum Associates.

[21] Saad, F., & Villame, T. (1999). Intégration d'un nouveau système d'assistance dans l'activité des conducteurs automobile. In J.G. Ganascia (Ed.), *Sécurité et cognition.* Paris : Hermès.

[22] Sheridan, T.B. (1992). *Telerobotics, automation, and human supervisory.* The MIT Press. Cambridge, Massachusetts.

[23] Stanton, N.A., Young, M., & McCaulder, B. (1997). Drive-by-wire : the case of driver workload and reclaiming control with adaptive cruise control. *Safety Science, 27,* 149-159.

[24] Zuboff, S. (1988). *In the age of the smart machine. The future of work and power.* New York : Basic Books.

Safety of a Human-Technology-Organization System: Mobile Phones and Driver Cognition Disturbance

Ola Svenson
Risk Analysis Social and Decision
Research Unit, Department of
Psychology, Stockholm University
S- 106 98 Stockholm Sweden
+46-8-163680
osn@psychology.su.se

Christopher Patten
Risk Analysis Social and Decision
Research Unit, Department of
Psychology, Stockholm University
S- 106 98 Stockholm Sweden
+46 -243-75845
christopher.patten@vv.se

ABSTRACT

Like any human-technology-organization system, the road traffic system includes people as operators, supervisors and regulators. Drivers interact and have to share cognitions, communications and rules of behavior to achieve safety in traffic. When new means of communication are introduced into a system their effects on the system should be estimated. Rather recently IT systems have been introduced in the private car. The systems are exemplified by navigation systems and mobile phones The present study reviews research on the effects of using one of these systems, the mobile phone when driving. The study starts with a perceptual-motor cognitive model of the driver and continues with an account of the different effects on driving of using a mobile phone. First, it is should be pointed out that the availability of a mobile phone in a car is of great value in emergencies and accidents. However, the results from some 80 studies show that using a mobile phone in a car while driving impairs driving performance significantly. That is because, a driver's attention to traffic and traffic information is impaired and the control of the car becomes less precise and smooth when talking over a phone. The conversation in itself and, in particular, demanding communications impair both attention and maneuvering performance significantly as well as the motor activities needed for phoning. This implies that hands-free mobile phones will not solve the safety problem of phoning and driving. Analyses of accidents have shown that the impairment of driving while phoning leads to an increased risk for having an accident for both hand-held and hands-free phones. One important characteristic of a phone conversation in relation to most other in-car activities is that the pace and content of the phone conversation cannot be controlled as well by the driver. This makes a phone conversation more disturbing than other equally demanding in-car activities that can be distributed in time and adapted to current traffic and driving conditions.

General Terms

Management, Performance, Security, Human Factors.

Keywords

Human factors, traffic safety, mobile phones

1. INTRODUCTION

People in the road traffic system include drivers, traffic controllers and regulators. The present study will focus on the interaction between drivers and the road traffic system. Drivers have to share cognitions, communications and rules of behavior to achieve safety in traffic. When new means of communication are introduced into a system their effects on the system should be estimated. Rather recently, IT systems have been introduced in the private car. The systems are exemplified by automatic speed control, GSR navigation systems and mobile phones. The present study reviews research on the effects of using one of these systems, namely the mobile phone when driving and was based on a comprehensive review by the authors (Svenson & Patten, 2002).

2. MODELS OF THE DRIVER, DRIVING AND THE USE OF A MOBILE PHONE

When studying safety and the use of information technology equipment in cars while driving, it is of great value to refer to a model of the driver (e.g., Ranney, 1994; Wickens & Hollands, 2000), a model of the task of driving a car (e.g., Ranney, 1994) and a model of a person communicating through IT systems, such as, mobile phones.

2.1 Model of the driver as a cognitive information perceptual psychomotor system

Humans need *energetic mental resources* for interpreting, monitoring, processing and transmitting information. Energetic resources are available up to a certain limit at a given moment in time. Information *load* is the amount of information processed per unit time needed to manage the tasks performed. *Effort* stands for the energetic resources per unit time needed to manage the tasks performed. Effort therefore varies over time. Processing *capacity* is the maximum amount of information that can be processed per unit time.

Capacity depends on input characteristics (e.g., visual accuracy, signal characteristics), energetic resources available for input channel (e.g., the attention allocated to a certain input, how attentive a driver is to road signs) and overall total energetic resources available (e.g., less when a driver is tired than when he is not). The difference between the capacity available and the load is the *redundant capacity* at the moment and it represents a safety margin in that the demands on information processing can increase and the system is still able to handle this (e.g., when an unexpected event occurs while driving and the driver has capacity to handle the situation). *Psychomotor performance* (e.g., steering a car) depends on information processing and, for example, characteristics of the psychomotor performance, competing tasks and on the environment. Psychomotor performance requires information processing, energetic and motor resources and is integrated in the processing framework.

2.1.1. Attention and processing capacity

Among the major contributing causes to traffic accidents one finds inappropriate speed, attention and perception errors. In addition, the severity of the outcome of an accident is closely related to speed. Attention and perception errors predominate as contributors over response errors (Smiley & Brookhuis, 1977). Therefore studies of attention and human information processing are very important for traffic safety.

Fortunately, the area of attention and human information processing is a prominent area of psychology and there are many reviews of the research in the field (e.g., Kahneman, 1973; Hockey, Gaillard & Coles 1986; Samuelsson & Nilsson, 1996; Pashler, 1999; Wickens & Hollands, 2000). As mentioned above, all information processing requires energetic resources. Different processes require different amounts of effort and energetic resources. Automated information processes typically require less effort and resources than elaborate consciously controlled processes. If the system has redundant capacity, which is important in car driving, there are resources and capacity left to mobilize extra effort in case of a hazardous situation coming up. If the system is used to its limit, then there is no such possibility and information can be missed, the information processing delayed and/or some ongoing tasks quickly abandoned to release resources for processing information about the hazardous situation.

It is important to point out that there are numerous system interactions between all parts of the perceptual - cognitive psychomotor systems. To illustrate, parallel processing occurs at different levels with input from different channels. There are different views on the issues of parallel processing and a common energetic resource pool for all information processes. The view we will adopt here is that information processing can be viewed as parallel and that although there is a common resource pool setting limits to processing, there are energetic resources available and used for specific purposes (e.g., visual processing), but only to a certain extent (Wickens & Hollands, 2000). When the common energetic resources are taxed this affects all processing. Time sharing and buffering of information (e.g., for output) is also assumed to be a characteristic of human information processing. To elaborate, if one modality (e.g., hearing) is used for two different tasks, this typically requires more energetic effort than if two equally difficult tasks were divided across two modalities (e.g., vision and hearing). Still, the two tasks would require more effort than one of the subtasks alone.

In the following, we shall make some simplifying distinctions of human information processing when dividing them into input, central and output system components.

2.1.2. External sensory input of information

The most important input when driving a car is *visual.* Driving a car without visual information is impossible. Therefore, the available visual attention and redundant processing capacity available are of great importance. However, *auditory* input is also very important when driving, for example, in order to alert the driver about possible dangers and for giving feed back from the driving activity. *Tactile* input is also crucial in the driving situation because it provides feedback to the psychomotor activity of using the controls of the car. Finally, *kinesthetic* input gives information about acceleration in different directions and provides information about, e.g., speed.

2.1.3. Central processing – memory and cognition

The human brain is active generating and processing information during all the time. This is true also for a driver when he or she drives a car. Thoughts that are unrelated to driving come and go at the same time as external inputs are processed while driving. In addition, planning a route to follow when driving through a city is also an example of an internal cognitive activity that generates and processes information. This can affect the level of attention to external sensory inputs and the processing of the corresponding information.

The information of the sensory input or internally generated information (e.g., thoughts) is fed into the human who processes the information on different levels from the automatic to a very elaborated and conscious cognitive level. Pre-processing of information related to the input channel (e.g., vision) occurs first, followed by integration of information from the different channels. This integration process means that information from one sensory input may affect the capacity to process information from another channel (e.g., listening intensely to something can affect visual attention and a visual input may affect auditory input). As mentioned above, there is a common pool of energetic resources from which resources can be allocated to different inputs through conscious or non-conscious attention processes. Based on the information available, decisions are made and actions prepared under the prevailing conditions.

For example, Shallice (1988) describes actions and the allocation of energetic resources as hierarchically controlled. There are corresponding models of the driver in a car (e.g., Allen, Lunefeldt, & Alexander, 1971; Michon, 1985). In such a hierarchical supervisory model a general-purpose system can use representations of the environment and intentions and abilities of the driver. The system selects higher- level schemas, which attenuate lower-level schemas in turn controlling specific subsystems. To exemplify, when allocating focal attention over time a certain schema is selected (e.g., looking out the windscreen to control the car) provided its activation level exceeds a certain level (e.g., the car starts moving which activates the schema with a certain intensity above the threshold). When, some other (lower-level) schema is activated (e.g., by a phone signal), the original schema may be retained, but allowed a certain time interval to try to reach the goal of the lower-level schema (e.g., to answer the phone). However the higher-level schema is still active and only a certain limited time interval is normally allowed to carry out the lower-level schema (of answering the phone). But, a lower-level schema like answering a phone may seriously disrupt the higher-level schema of driving a car safely. Most of the schemas in driving are automatic and only partly available to conscious control, in particular lower level schemas on the operational level.

2.1.4.Output – response execution

Outputs from information processes include symbolic (e.g., speech) and psychomotor actions (e.g., manually interacting with a control such as a radio or a light switch). When driving there are different manual output actions, such as, turning the wheel and braking. Most of these actions are psychomotor actions to control the car and some give signals to fellow drivers. All outputs require coordination, mental energetic and psychomotor resources. The output quality can be negatively affected through e.g., inadequate central processing and timing of actions.

2.2. Driving a car (primary task)

Driving means the control of a car in a dynamic traffic environment. Time is a fundamental variable when driving and the speed of the car determines the time windows available for the driver to attend to, process and respond to the environment. Driving a car also means that a driver should be able to use a great repertoire of psychomotor skills. Sensory inputs provide information that is processed in order to produce outputs compatible with the demands of the driving situation. The driver's possibility to perform the necessary driving task varies with her or his own condition and the demands posed by road and traffic conditions, the car and other tasks going on at the same time (talking on a mobile phone, tuning the radio etc.).

A driver uses sensory inputs to regulate the speed and position of his or her vehicle. Vision is used all the time and ideally, the eyes of the driver should never leave the exterior view (but of course, e.g., speed information is provided inside the car and needs visual attention in today's vehicles). Hearing is used for drawing attention (e.g., to warnings) and as an input to cognitive activities (e.g., listening to traffic reports on the radio). Hearing also provides feed back about the state of the car (e.g., speed, mechanical problems). Finally, kinesthetic input provides feed back from controls (e.g., steering wheel) and the road (e.g., side position) etc.

Most driver responses are psychomotor responses directed toward driver-vehicle interfaces. Only a few (e.g., facial expression as signal to another driver) are directed to other recipients. Most operational (e.g., to brake) and some tactical decisions (e.g., to change lanes) are more or less automatic while strategic decisions (e.g., to choose a new route) are not automatic (cf., Alm & Nilsson, 2001). As mentioned above, the more automatic a decision is the less energetic resources are needed in general. This means that strategic decisions are more disrupted than operational decisions when the cognitive load of a driver approaches his or her capacity limit.

The physical psychomotor character of operative control of a mobile phone engages parts of the body (e.g., one hand for dialing a number) and therefore the execution of simultaneously activated operational decisions (e.g., to signal a turn) may suffer in timing and quality.

2.3 Using IT equipment, such as a mobile phone (extra task)

The mobile phone is a real time interactive IT system making it possible for the driver to interact with another (distant) person a capacity, which makes it different from many other modern systems in the car. When considering the effects of mobile phones it is important to emphasize that the mobile phone should be seen in the context of other systems in the car and not in an isolated way. in the present context, we will first focus on the task of making and receiving mobile phone calls in a car without much consideration of interactions with other IT systems.

Calling up someone or receiving a call on a mobile phone involves both psychomotor and cognitive activities, which may interfere with the driving task. There are different kinds of mobile phones and ways of characterizing them. Making a phone call and receiving a call are different in some important ways (c.f., Alm and Nilsson, 2001 who presented an analysis of a phone calling task and possible interferences with driving). The most important difference is the one between making a phone call and receiving a call. When calling up, the driver has control over when she or he should start to call and in this way it is possible to place the call when the traffic situation is favorable. But, when receiving a call there is no such control and the call arrives uncoordinated with the driving task.

Making a call means that a number has to be dialed and this is a perceptual cognitive psychomotor activity that may interfere with the driving task. Retrieving the number from memory requires some energetic resources and if the number has to be accessed in a phone book (in the phone or separately) further visual attention resources are required. Pressing the button(s) on the phone to call up requires visual attention and psychomotor activity.

During a call the conversation needs attention, cognitive activities and energetic resources that vary depending on the content of the conversation. To exemplify, taking notes interferes with visual attention, cognitive activities and psychomotor activities needed for driving a car. Having a conversation over the phone requires mental processing and energetic resources etc.

The introduction of hands free mobile phones has relieved the driver from the psychomotor control of the phone, which earlier engaged one of his or her two hands preferably needed for driving. However, mental processing and energetic resources are still needed. Salvucci and Macuga (2002) presented a cognitive process model of a person using a mobile phone with manual, speed (dial a single preprogrammed number) and voice control (with press and hold) routines for calling up someone. The researchers combined this model with a model of driving. The model simulation output was compared with participants driving in a simulator. In terms of correlation, the results indicted reasonable validity concerning speed and lateral deviation. However, the absolute levels of speed and lateral deviation were generally underestimated. The model overestimated the effects of using the manual phone mode on these variables (about twice as big).

3. PERSPECTIVES ON ASSESSMENTS OF DRIVING SAFETY

The previous overview of the driving task and the driver provided a perspective on how it is possible to assess the effects of IT and mobile phone use on the quality of driving. The general philosophy behind most of the methods to assess driving safety, is to explore to what extent redundant driver capacity is available. It is assumed that that the more redundant capacity there is, the safer the driving. As mentioned earlier, this is because when, for example, unexpected stimuli or events appear a driver has extra resources to handle them in a quick operational action. Also, tactical and strategic decisions are generally improved when cognitive load and time pressure decrease (Svenson & Maule, 1993).

One way of measuring driving safety while talking over a mobile phone is to monitor, so called "primary task" measures (e.g., speed, lateral position) on the road or in a simulator. Other methods include assessing the redundant capacity of different stages in information processing, decision making and action. To illustrate, processing capacity available when the primary task (e.g., driving) and extra task(s) are managed by a driver can be assessed through a ***secondary task*** (e.g., the detection of a light signal). The performance on the secondary task in conjunction with the primary and extra tasks, provide information about energetic resources available and how they are allocated. It is important to measure performance on all tasks, because a researcher has limited control over the priorities a participant assigns to different tasks.

The above measures relate to the "over stimulated" driver with too much sensory input. However, there are also situations, in which there is a risk of "under stimulation" leading to unsafe driving, for example, on monotonous straight motorways without any other traffic. In such situations, the driver can also loose attention capacity, miss important traffic information, forget plans and loose driving ability.

3.1 Assessments of input

Methods in this category typically use secondary tasks to indicate how much of the information processing capacity that is available in addition to the ongoing focal activities. That is, to measure how well a driver can handle other information and action when driving is the primary task. To exemplify, visual attention capacity that is available can be measured by the driver reacting to visual stimuli appearing in the periphery of the visual field. The reaction time and hit rate can be used as a measure of the resources available besides the focal task(s).

However, it is most important to note that a driver/participant's resource allocation can be affected by instructions and/or the participants' own expectations. To illustrate, when driving in a simulator with an extra and a secondary task, the priorities between these tasks can change over conditions in the same experiment. In one condition talking over a phone may have the highest priority, while in another condition it may become the secondary task. Thus, a reallocation of energetic resources from the focal driving task (e.g., decreased visual attention to the road) to the secondary task may give the false impression that a driver has more redundant resources available on top of safe driving than he or she really has. Instead, the driving safety margin resources were used to perform well on the secondary task.

3.2 Assessments of effort

Effort, is the energetic resources spent per unit time to perform a set of activities. One way to assess effort is to ask drivers about the effort needed for a particular driving task. The degree of rated effort is assumed to reflect the effort used to perform a task. There are a number of formalized scales and instruments available to measure cognitive load or effort, for example, the NASA_ TLX scale (Hart & Staveland, 1988), which is an instrument used quite frequently in research on traffic safety and mobile phoning. The NASA-TLX is a multidimensional rating scale providing a global score of mental effort. It consists of six factors: mental demand, physical demand, time pressure, effort, performance and frustration. The ratings are given on bipolar rating scales and concern the experiences during the task as remembered immediately afterwards. However, measures based on subjective ratings are open to different possible biases depending on the ability and willingness of the raters to judge effort.

To the best of the present authors' knowledge, there is no study in which participants were asked to rate how they prioritized and distributed their attention and processing capacity among tasks. Although a number of these processes using energetic resources are not available to consciousness, information about resource allocation may be of value when explaining some of the laboratory and simulator results with mobile phones and driving.

There is also a possibility to collect physiological measures (cardiac, respiratory, skin responses etc.) as measures of effort. These measures are not susceptible to the biases of judged effort, but their validity as measures of effort can be questioned in other ways. According to Pashler (1999, pp. 382 - 383), the physiological measures have not always been shown to correlate with effort measured in other ways (e.g., graded voluntary control, subjective perception of effort).

3.3 Assessments of redundant central processing capacity

A secondary task may involve more complex central processing than just detecting a given stimulus signal as treated above.

Asking a driver to perform standardized cognitive activities, such as adding numbers or process information in other ways while driving demands energetic resources. Therefore, the performance on such tasks can provide measures of how much extra central capacity the driver has when he or she drives a car.

In accordance with what we pointed out earlier, Brown, Tickner and Simmonds (1969) found that overlearned tasks (skill based processes, such as, the control of a car in normal motorway driving) are generally less affected by divided attention than less well learned tasks (rule based and knowledge based processes, such as, planning a route, figuring out when to fill gas). However, they found that some skill based perceptual-motor tasks were negatively affected (e.g., steering through a gap).

3.4 Assessments of output - control of the car

There are many ways of measuring output behavior. To exemplify, the speed of talk in a conversation may vary in a systematic way reflecting the processing load of the driving task and the need of resource allocation from talking to driving. Other measures relate to the control of the car, e.g., headway, time to collision, braking pattern, lateral position on the road.

3.5 Assessments of output – traffic interactions

This refers to driving characteristics that can be normatively evaluated in relation to, e.g., norms, regulations and safety. Examples of this are forgetting to switch to dipped headlights when meeting a car, forgetting to signal direction and not noticing road signs (e.g., speed limits). This gives an indication of how the driver is able to process the extra tasks in parallel with the processes needed for performing the driving task (e.g., reacting to signals and signs). Such measures may also indicate that the driver experiences the situation so effort demanding that some other tasks will be slowed down (e.g., the speed of the car).

When people perceive risks adequately and have a possibility and a wish to compensate for an increased risk, such as the one imposed by talking over a mobile phone, people can reduce the risk. To exemplify, Alm and Nilsson (1995) found that drivers did not adjust for the increased braking time associated with the use of a mobile phone through having a greater headway to a leading car. This can be explained by the drivers not being aware of their increased braking reaction times and/or by the drivers assuming that they already have a long enough headway to ensure safe driving. But, in reality the braking safety margins were insufficient according to Alm and Nilsson despite the fact that talking in a mobile phone often decreases the mean speed (Alm & Nilsson, 1990; Brookhuis, de Vries & De Ward, 1991). Alm and Nilsson discuss the possibility that drivers may not have had the opportunity to learn from experience. One may also question what the possibilities are to learn from experience. Strayer, Drews and Johnston (2002) showed that implicit learning was impaired during a mobile phone conversation. In other words, drivers talking over the phone are not only unaware of their greater risk taking, but they also impair their chances to learn about this fact through distortion of feedback just because they are talking over the phone when they need to learn.

4. LABORATORY, SIMULATOR FIELD AND EPIDEMIOLOGICAL STUDIES

4.1 Effects of mobile phones on driver sensory and perceptual input

When driving while talking on a mobile phone, the duration of each glance on the road increases and the field covered decreases so that a "visual tunneling effect" arises where the central area of the visual field is given proportionally more attention and the peripheral area less. The number of saccades (eye movements) can decrease from about 90 per min in normal driving to about 80 per min when the driver becomes engaged in a phone conversation (Recartes & Nunes, 2000). Inexperienced drivers leave the attention of the central task of driving for longer times than experienced drivers (Wikman, Nieminen & Summala, 1998).

Detection times to traffic targets typically increase with about 50 to about 400 ms and the probability of missing a target altogether increases during a mobile phone conversation (Lamble, Kauranen & Summala, 1999; Tokunaga, Hagiwara, Kagaya, & Ondera, 2001; Strayer & Johnston, 2001). The more demanding the conversation is, the greater the loss in reaction time and target detection probability.

Making a call involving dialing and handling the phone, divides driver attention during about 10 to about 40 sec or more depending on type of phone. Searching for a phone and/or a phone number are tasks seriously competing with driver road attention and these activities affect driving ability quite negatively (Graham & Carter, 2001).

4.2 Effects on central cognition and thought processes

Talking over a mobile phone requires extra mental and psychomotor resources and many studies have shown this in increases in physiological and subjective measures (Parkes, Fairclough, & Ashby, 1993; Brookhuis, de Vries & De Ward, 1991; Cnossen, Rothengatter & Meijman, 2000). The effects of this, is that concurrent thought processes may be disturbed, postponed or eliminated. To exemplify, difficult phone conversations requiring great amounts of mental resources are very destructive for other thought processes using spatial representations, such as, planning or checking one's route, thinking about where to find a resting place, a parking lot and estimating whether the car fits into an open space.

4.3 Effects on driver behavioral output

In comparison with just driving, a driver makes more frequent and larger steering movements when engaged in a mobile phone conversation. Lane keeping can be measured by lateral position in relation to mean track. When talking on a mobile phone the standard deviation of lateral position (meaning that about 30% of the lateral positions are outside that position on a trip) typically increase from about 0.2 m to about 0.3 m. Note, however that for straight rural roads the standard deviation has been found not to increase under some conditions. When engaged in a phone call in more complex traffic environments than straight rural roads, the standard deviation of lateral position can be expected always to increase (Brookhuis, DeVries & DeWard, 1991; Alm & Nilsson, 1995; Reed & Green, 1999)

Reaction time to a speed decrease of a car in front of a driver has been found to increase by about 600 ms delaying a proper speed adjustment of the car with that time (Lamble, Kauranen & Summala, 1999). Drivers have been found to brake harder in response to a traffic condition while conducting a mobile phone conversation than when they are not on the phone. The harder braking can compensate (at least in good weather) the longer brake reaction time when talking on a phone. However, it has also been shown that longer braking distances are a result of phoning while driving and that the drivers in those cases are not able to adjust their speeds or braking sufficiently to compensate for the loss in reaction time. Therefore, the result is longer braking distances while talking over the phone. Generally speaking, older drivers react more slowly than younger drivers and are also affected negatively by, in particular, complex phone conversations (but not necessarily by very simple conversations).

Driver psychomotor control activities (steering, shifting gears, changing to dipped headlights etc.) compete with finding a number, dialing it, using a hand held phone, etc. This competition of resources plays a greater disturbing role, for instance, in dense and demanding traffic than on a straight motorway, where the effects usually have been relatively small or nonexistent as reported in the research literature. However, expert drivers (having driven more than 10 000 hours) are less disturbed by extra tasks because driving draws less of their cognitive and motor capacities than it does for other drivers. In addition, many professional drivers also attain expertise in using different extra IT tasks.

4.4 Conclusions

Having a phone conversation over any mobile phone system interferes with driving - with or without the driver's own awareness. Having a demanding mobile phone conversation while driving disturbs or eliminates other thought processes, for instance, the planning ahead of which way to drive.

Making a phone call interferes with both the driver's cognitive and psychomotor processes needed for controlling a car and driving safely. The strength of the psychomotor interference depends on the phone system, but the disturbing effects cannot be completely eliminated through e.g., hands free systems.

The information needed for driving is sampled less frequently and the control of the car becomes less smooth when the driver communicates over a mobile phone. There is a possibility to compensate for loss in attention and control functions during a mobile phone conversation through, e.g., decreasing speed and increasing headway distance to a vehicle in front. However, contemporary research indicates that the compensation cannot be expected to be sufficiently strong to outweigh the decrease in driving performance accompanying a mobile phone conversation - in particular in sudden critical traffic situations.

The conclusions from experimental, simulator and fields studies that using mobile phones while driving impairs driving significantly are validated by results from accident investigation studies showing that mobile phoning is a significantly contributing factor of accidents.

5. REFERENCES

Allen, T.M., Lunefeldt, H. & Alexander, G.J. (1971) Driver information needs. *Highway Redearch Records, 336*, 102 – 115.

Alm, H. & Nilsson, L. (1990) Changes in driver behaviour as a function of hands-free mobile telephones. A simulator study. *Reprint from Drive project V1017 (Bertie), October 1990. VTI, Linköping, Sweden.* VTI särtryck 175/1990.

Alm, H. & Nilsson, L. (1995) The effects of a mobile telephone task on driver behaviour in a car following situation. Accident Analysis and Prevention, 27, 701 – 715.

Alm, H. & Nilsson, L. (2001) The use of car phones and changes in driver behavior. *International Journal of Vehicle Design, 26*, 4 – 11.

Brookhuis, K.A., DeVries & DeWard, D. (1991) The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention, 23*, 309 – 316.

Brown, I. D., Tickner, A. H., & Simmonds, D. C. V. (1969) Interference between concurrent tasks of driving and telephoning. *Journal of Applied Psychology, 5*, 419-424.

Cnossen, F., Rothengatter, T. & Meijman, T. (2000) Strategic changes in task performance in simulated car driving as an adaptive response to task demands. *Transportation Research* Part F, 3, 123 – 140.

Graham, R. & Carter, C. (2001) Voice dialling can reduce the interference between concurrent tasks of driving and phoning. *International Journal of Vehicle Design, 26*, 30 – 47.

Hart, S.G. & Staveland, L.E. (1988) Development of NASA_TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: Elsevier.

Hockey, R.G., Gaillard, A.W.K. & Coles, M.G.H. (1986) *Energetics and human information processing.* Dordrecht: Martinus Nijhoff.

Kahneman, D. (1973) *Attention and effort.* New York: Prentice Hall.

Lamble, D., Kauranen, T. & Summala, H. (1999) Cognitive load and detection thresholds in car following situations: Safety implications for using mobile (cellular) telephones while driving. *Accident Analysis and Prevention, 31*, 617 – 623.

Svenson, O., & Maule, J. (1993) *Time pressure and stress in human judgment and decision making*. New York: Plenum (pp 323-330).

Michon, J.A. (1985) A critical review of driver behaviour models: what do we know, what should we do? In L. Evans & R. (Eds.) *Human behavior and traffic safety*.New York: Plenum

Parkes,, A.M., Fairclough, S.H. &Ashby, M.C. (1993) Carphone use and motorway driving. HUSAT Research Institute, Loughborough, United Kingdom.

Pashler, H.E. (1999) *The psychology of attention.* Cambridge USA: MIT Press.

Ranney, T.A. (1994) Models of driving behavior: A review of their evolution. *Accident Analysis and Prevention, 26*, 733 – 750.

Recarte, :A. & Nunes, L.M. (2000) Effects of verbal and spatial imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied, 6*, 31 – 43.

Reed, M.P. 6 Green, P.A. (1999) Comparison of driving performance on-road and in a low-cost simulator using concurrent telephone dialing task. *Ergonomics, 42*, 1015 – 1037.

Salvucci, D.D. & Macuga, K.L. (2002) Predicting the effects od cellular phone dialing on driver performance. *Cognitive Systems research, 3*, 95 – 102.

Samuelsson, S. & Nilsson, L. (1996) Om möjligheterna att upptäcka och identifiera perifert presenterad information i bilen. Linköping: Väg och Transportforskningsinstitutet, VTI rapport No. 412.

Shallice, T. (1988) *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Smiley, A. & Brookhuis, K.A. (1987) Alcohol, drugs and traffic safety. In J.A. Rothengatter and R.A. de Bruin (Eds.) *Road users and traffic safety*. Assen, The Netherlands: Van Gorcum.

Strayer, D.L. & Johnston, W.A. (2001) Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science, 12*, 462 – 466.

Strayer, D.L., Drews, F.A. & Johnston, W.A. (2002) Cell phone induced failures ov visual awareness during simulated driving. Utah, USA: *University of Utah, Department of Psychology*, manuscript.

Svenson, O. & Patten, C. (2002) Information technology in the car: Mobile phones and traffic safety – A review of contemporary research. Report to the Swedish National Road Administraion 64 pages.

Tokunaga, R.A., Hagiwara, T. Kagaya, S. & Ondera, Y. (2001) Cellular phone conversation while driving: Effects on driver reaction time and subjective workload. *Transportation Research Record, 1724*. Paper no.00-1480.

Wickens, C.D. & Hollands, J.G. (2000) Engineering psychology and human performance. Upper Saddle River, NJ: Prentice Hall.

Wikman, A-S., Nieminen, T & Summala, H. (1998) Driving experience and time-sharing during in-ar tasks on roads of different width. *Ergonomics,41*, 358 – 372.

Session 2 Error Management

Searching for Ways to Recover from Fixation: Proposal for a Different Viewpoint

Machteld van der Vlugt
Delft University of Technology
Department of Mechanical Engineering
Section Man-Machine Systems
Mekelweg 2, 2628 CD Delft, The Netherlands
m.vandervlugt@wbmt.tudelft.nl

Peter A. Wieringa
Delft University of Technology
Department of Mechanical Engineering
Section Man-Machine Systems
Mekelweg 2, 2628 CD Delft, The Netherlands
p.a.wieringa@wbmt.tudelft.nl

ABSTRACT

Despite the efforts to maximize safety and performance of complex dynamic environments such as aviation, there is widespread acceptance that human error and consequently erroneous behavior is inevitable in the end. Fixation is such a form of erroneous behavior which is found hard to prevent, because it occurs unexpectedly. Yet, in many cases it may be possible to recover from fixation before it leads to adverse consequences. An adequate understanding of fixation is required to be able to design interventions to recover from fixation. In this paper a viewpoint is proposed to analyze behavior like fixation which considers behavior as autonomous. This viewpoint adheres to the hypothesis that humans cannot behave inconsistently within their actual frame of reference, which is not necessarily the designed frame of reference. This autonomy viewpoint is applied to a controlled flight into terrain. Analysis of this case indicates that the proposed viewpoint could offer opportunities to detect fixation earlier and may give new perspectives for developing intervention techniques. Still, a further validation of the model is necessary based on new cases.

Keywords

Fixation, human error, erroneous behavior, recovery, Sense of Influence

1 INTRODUCTION

Several studies in complex dynamic environments have noted a typical form of erroneous behavior, which is characterized as a failure to revise plans or diagnoses when they should have changed due to changes in the environment. This form of behavior is in the literature defined as fixation, tunnel vision, or preoccupation ([3], [6], [14], and [15]). Despite the measures that have been taken to prevent a recurrence of fixation, it has been found difficult to rule out fixation completely ([7], [8], and [11]).

In complex dynamic domains a situation can evolve over time into a critical situation without the necessity of any (human) command input. Consequently, critical situations rarely pop up at once full-size, they develop due to a succession of minor failures during (non-)normal situations. Non-normal situations initiated by technical failures, have been anticipated successfully by improved training methods, improved user interfaces and support systems, and improved procedures. However, non-normal situations can still develop into critical situations due to a series of "decision errors" of the crew causing an unanticipated development of the situation. These developments coincide with fixation for both experienced and less experienced crews [7].

Lately the focus of safety has been mainly on prevention of human error. Though unforeseen failures are, as the word unforeseen already implies, impossible to prevent. Yet, in many cases after the occurrence of a failure there are still possibilities to recover from it [2], on the condition that there are time and opportunities to intervene. Characteristic for fixation is that it often results from a succession of minor failures which themselves are anticipated for in the precautions that have been taken. Therefore the aim is to move beyond prevention of fixation and to look for ways to recover from fixation. However, the development of techniques that will force the pilot to recover from fixation asks for an adequate understanding of human behavior.

The model/theory that we have been developing to find ways to force the recovery from fixation is based on Varela and Maturana's theory of autonomous living systems [5]. This theory does not represent humans as systems controlled by inputs, but as autonomous systems only perturbed by influences. In this paper this viewpoint is introduced for studying fixation and applied to a case of the National Transportation Board.

2 VIEWPOINT PROPOSED FOR STUDYING ERRONEOUS BEHAVIOR

According to Varela and Maturana's theory of autonomous living systems an autonomous system cannot behave inconsistently. Behavior that seems inconsistent to the observer is due to a difference between the frame of references of the observer and the subject [5].

2.1 Behavior and Maintenance of Organization

Current behavior models ([1], [9], [10], and [13]) implicitly represent humans as controlled systems, that is, as systems whose output is controlled by its input [12]. By representing humans as controlled systems, the essence of humans and thus

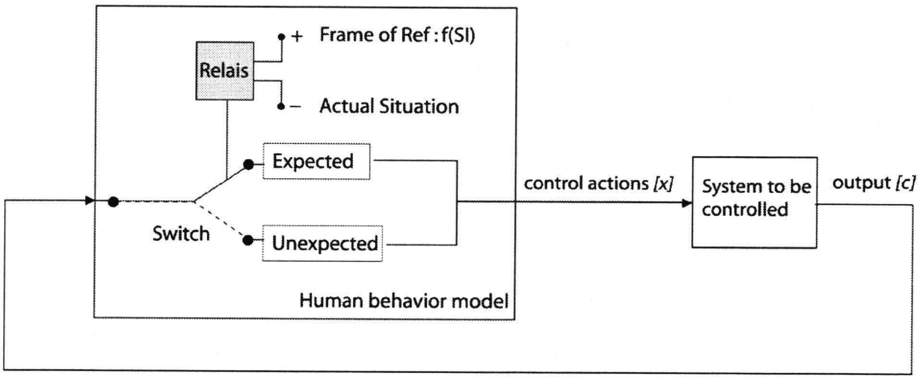


Figure 1: Principle of autonomy: maintaining sense of influence (SI), the frame of reference taken is a function of the critical variable (SI); SI is not determined by the environment but by the subject. When the actual situation matches the subject's expectations, the subject's SI is optimal and the designer's anticipated frame of reference matches the subject's frame of reference: the switch is in the upper position and the subject shows expected behavior; when the actual situation mismatches the subject's expectations, his SI threatens to lower resulting in a tension, a change in frame of reference occurs, though the tension remains, the switch is in the lower position, the subject shows unexpected behavior.

the essence of human behavior is lost, namely its autonomy. Yet, with erroneous behavior such as fixation, autonomy is strikingly present; the output (actions or decisions) is not in accordance with the expectations the observer has taking into account to the observed input based on the controlled representation of human behavior. The model, that we have been developing and that will be presented in this paper, represents humans not as systems that are controlled by inputs (or references), but as systems, which are perturbed by influences from the environment. Outputs of controlled systems are by definition determined by their inputs (and of course the system's structure), whereas outputs of autonomous systems are determined by their organization and not by their inputs from the environment. In case of autonomy, inputs from the environment (influences) are subordinate to the maintenance of this organization, represented by the (internal) variable sense of influence (SI).

Humans are autonomous homeostatic systems of which the fundamental variable is the maintenance of their organization, that is, their survival [5]. The organization, that is, the network of relations, characterizes the system, whereas the components forming a structure realize the concrete system. The organization of a system, therefore, does not specify the properties of the components. An artificial system can be realized by different structures. A thermometer, for example, can be made of a glass tube containing mercury or alcohol or can be made with a thermocouple; the concrete structures differ, but the organization remains the same. Typical for living systems is that they maintain their organization while their concrete structure is continuously changing. A good example of a continuously changing structure is the learning effect of humans. Thus analyzing living systems in terms of their components only won't work because the moment they are modeled they have been changed already. Therefore it is tempting to observe human behavior considering maintenance of organization.

2.2 Inconsistencies in Behavior are Inconsistencies in Context

The logic of the description of a system and hence the logic of the behavior of a system (the subject) is necessarily the logic of the describing system (the observer) [5]. Because behavior of humans is dominated by the maintenance of their organization no inconsistencies in their behavior can possibly exist. From this viewpoint inconsistencies in the subject's behavior as they appear to the observer arise from a change in frame of reference (context) that generates the subject's behavior while the context of the observer does not change accordingly. To render inconsistent behavior (from the observer's viewpoint) into consistent (understandable) behavior, the observer should widen his context such that it embraces the subject's context.

Behavior is an ongoing process of interactions, involving two or more participants. Essential for humans is acknowledgement of their membership of a social system. A group of humans recurrently interacting constitute a social system from which they derive their identity. To get involved as a member of a social system consists in becoming behaviorally coupled to it. To stay involved in a social system requires humans to satisfy the system's unwritten rules and habits that come about by recurrent interactions. These are called social representations.

The key factor regarding acknowledgement is "sense of influence". We hypothesize that a person's sense of influence is at best when the desired effect matches the obtained effect. Sense of influence can be lost when the desired effect in mind mismatches the obtained effect [4]. When it has been lost the subject will optimize it in another context. This may be noticed by the observer as a change in the subject's behavior. In case the observer's context does not change accordingly, the subject's behavior seems inconsistent to the observer.

3 FIXATION REVISITED

Fixation has been defined as "a failure to revise a situation assessment in the face of opportunities to revise" ([14] and [3]). Following this definition, fixation is considered as inconsistent (erroneous) behavior; accordingly it is outside the frame of reference of the observer.

Regarding the impossibility of humans to behave inconsistently, the observer will need to acquire a frame of reference to match the subject's frame of reference. So far fixation has been considered as a result. The encompassed frame of reference is obtained by regarding the phenomenon fixation as a process. If behavior is a continuous process of generating acknowledgement, then fixation, being a subset of behavior, is a continuous process of generating acknowledgement as well.

In case of fixation the subject does not obtain the desired effect from his actions, resulting in a loss of sense of influence within the context in which the subject shows fixated behavior. A fixated subject does not obtain the desired effect from his actions prescribed in the tasks and goals formulated in the procedures and protocols. A tension arises between what should happen and what actually is happening. Failing to obtain the system's desired effect by behaving in accordance with the prescribed procedures, the subject's SI threatens to disappear. However, our axiom is that humans as autonomous beings optimize their SI. This is not a rational process of goals means and choices, it just happens, it is essential for living. Although autonomy may lead to unexpected or unforeseen behavior the effect is unambiguous: optimization of SI. If optimal SI is not realized within the frame of reference foreseen that the designer or observer have foreseen, the subject's frame of reference changes (unforeseen) optimizing his SI.

Thus, instead of viewing fixation as a failure to revise situation assessment, fixation is viewed as the visible effect of a change in frame of reference which optimizes the subject's SI. Failing to revise his situation assessment, the subject's behavior changes in accordance with the applicable social representations of the subject's professional identity. In other words the subject satisfies the (silently) agreed social representations. By behaving in this way, the subject regains sense of influence by satisfying the representations that hold for the social system that he takes part in. He will not act on the concrete situation he cannot handle.

The essence is to find the social representations to which the fixated subject behaves, only then an accurate intervention for a recovery from fixation is possible.

4 APPLICATION OF PROPOSED VIEWPOINT TO A CASE

In this preliminary phase of the investigation, the examination of the proposed viewpoint about fixation will be based on aviation accidents described thoroughly in the public NTSB (National Transportation and Safety Board) reports, in which fixation played an important role. In a later phase of the research, this proposed method will be applied to simulated cases. Firstly, the cases taken from the NTSB reports will be reviewed on the basis of the proposed viewpoint. The case taken in this paper is the controlled flight into terrain of Flight 801 of Korean Airlines. The critical activities and alarm signals concerning fixation will be plotted. With this plot the opportunities to intervene can be detected. Secondly, the social representations to which the subject behaved have to be uncovered. This will be done by looking for the prevailing representations of the social system "aviation". Partly these can be found in the NTSB reports, partly these have to be discovered by interviewing pilots and instructors about this case and other similar cases.

4.1 Short Description of the Case

The case concerns Flight 801 of Korean Air (KAL 801) of August 6, 1997 (see [7]) for a thorough analysis of the whole event). At 0105:00 the event is picked up, at that time the airplane had been cleared to land on runway 6 Left of Guam International Airport (Guam 6L). The approach had to be a nonprecision approach since the glideslope, necessary for a precision approach was unusable. For a precision approach the guidance for the crew is to maintain the glideslope, a lateral and longitudinal guidance path displayed on the flight director. For a nonprecision approach on instruments the guidance is to maintain a required vertical speed, which is displayed at the flight director and which is set manually. In a nonprecision approach the airplane is not allowed to descend below the intermediate safety altitudes as shown in Figure 2.

In case of a nonprecision approach two possibilities exist, a visual approach when vision is sufficient (according to the Visual Flight Rules) or an instrument approach when vision is insufficient. In his approach briefing the captain mainly briefed a visual approach. As backup he briefed some parts of the instrument approach in case the visual approach turned out to be impossible, with the reminder that the glideslope was unusable. After the briefing and a clearance the plane started to descend. Sure enough it turned out to be raining making a visual approach not likely. For about six minutes the crew discussed the weather, they had to switch from a visual approach to instrument approach, about which a discussion about the glideslope status emerged. For the following two minutes the plane was descending on approach. Twice the captain ordered to reset the intermediate altitudes (1,440 and 560 feet). In the last minute of the flight the crew was still looking for the runway to attempt for a visual approach when the ground proximity warning system warned for 500 feet (above ground level) at time 1.42.00. At time 1.42.20 when the runway was still not in sight the first officer requested to make a go around, two seconds later the go around procedure was executed though not aggressive enough to avoid a collision.

4.2 Evolution of the Fixation Patterns

The NTSB concluded that the captain lost awareness of flight 801's position on the approach to the runway at Guam International Airport, as a result of his preoccupation with the glideslope, his failure to cross-check the plane's position and his continuing expectation of a visual approach. According to the NTSB causal to the accident was the captain's early descend below the intermediate altitudes 2,000 and 1,440 feet, together with first officer and flight engineer their failure to properly monitor/challenge the captain's performance.

At 01.11 the captain gives his crew a short briefing on Guam 6L approach procedures, including indications to the captain's expectation of a visual approach and shortened briefing for a localizer-only approach in the event that a visual approach is not available. The Guam non-precision localizer-only approach uses the localizer for lateral guidance to the runway, the DME to identify step-down points and the VOR to identify the final step-down fix to the minimum descent altitude (MDA). The captain's briefing includes a reminder that the glideslope is inoperative, some details of the radio setup, the localizer-only MDA, the missed approach procedure, and the visibility at Guam (stated by the captain to be six miles). He does not note the definitions of the final approach fix (FAF), step-down fixes or their associated crossing altitude restrictions. He also

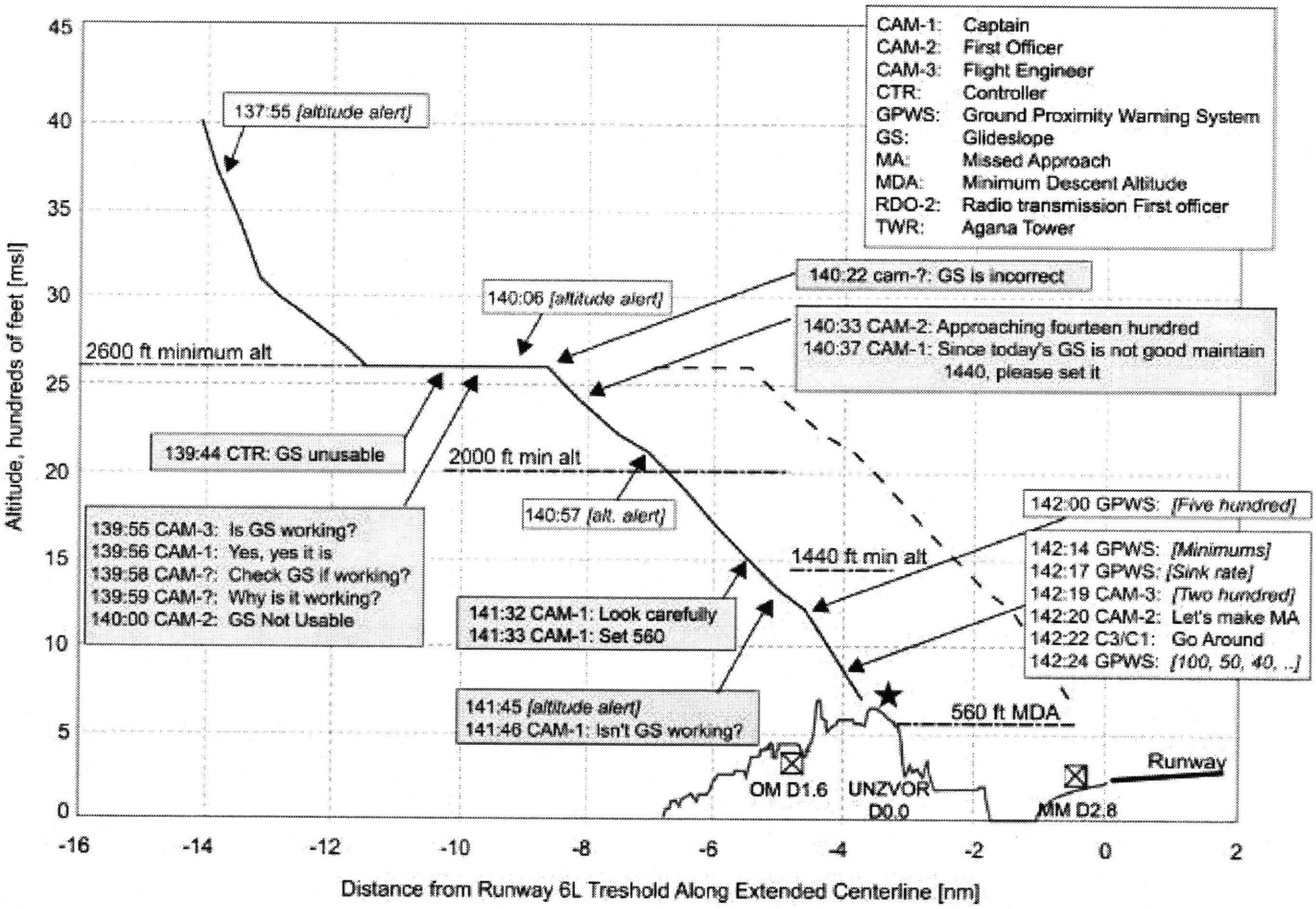


Figure 2: Overview of the last five minutes of the KAL801 incident adapted from NTSB (2000), for a complete CVR transcript is referred to the NTSB report Flight KAL801. The vertical axis shows the altitude above mean sea level and the horizontal axis the distance from the runway. The lower right corner between the –7 miles and 2 miles depicts the terrain elevation. The 2,600 feet, 2,000 feet, and 1,440 feet above mean sea level are the intermediate safety altitudes. The abbreviations are shown in the rectangle in the upper right corner of the plot. The black line is a side view of the descent profile of Flight 801 derived from the FAA radar data and the dashed line is the descent profile of Flight 801 shifted 3.3 miles to the airport.

neglects to brief the first officer on the way he will fly the descent, and does not discuss go-around decision criteria. Despite the captain's reminder in the approach briefing and the remark of the controller at 0139:44 that glideslope was inoperative the CVR recorded a dialogue of the flight crew about the status of the glideslope. The initial uncertainty about the glideslope appeared immediate after the radio transmission of the controller to which the first officer should have acknowledged that they received the message that the glideslope was out (see Figure 2). The flight engineer asked about the status of the glideslope to which the captain answered that it was working. The captain should have had sufficient cues to convince him that glideslope was out because when glideslope is out the captain's glideslope needles are covered by red labels (at least at some times) indicating that the glideslope is out and the glideslope capture indicators are absent.

As the airplane descends through 2,400 feet above mean sea level the first officer called out that they were approaching fourteen hundred feet. Several seconds later the captain directed the first officer to reset the altitude to 1,440 feet accompanied with the statement that today's glideslope was *not good.* By doing so, the captain replaced the intermediate safety altitude of 2,000 feet before the autopilot had captured it or reached the outer marker OM D1.6 (see Figure 2), causing the airplane to descend premature below 2,000 feet. Neither the first officer nor the flight engineer challenged the captain's decision.

Just before the plane reached 1,440 feet the captain asked the crew to look carefully (for runway 6L) and he ordered the first officer to set the altitude to 560 feet, the published minimum descent altitude for a localizer only approach. For the second time, the airplane descended too early through the restricted altitude, again neither the first officer nor the flight engineer challenged the captain's premature reset, or notified the captain of the premature descent.

The GPWS issues a radio callout, 1,000 feet (above ground level). Again the captain wonders if the glideslope was not working to which the first officer nor the flight engineer answered according to the CVR recordings. The flight engineer is astonished but no action was undertaken according to the data of the flight data recorder. The crew began the landing checklist. The GPWS annunciated "minimums, minimums" followed by "sink rate". At which the first officer stated, "Sink rate okay". Only after the flight engineer stated, "two hundred [feet]" the first officer said, "let's make a missed approach." One second later the flight engineer stated, "not in sight" to which the first officer responds, "not in sight missed approach."

According to the NTSB, the GPWS call outs were salient cues that should have caused the crew to check airplane's position and to act cautiously. From this moment there were three cues that should have initiated the crew to begin a missed approach procedure. The first is the GPWS call out 'five hundred', about 26 seconds before impact. The crew did not see the runway at that moment. If the airplane was correctly positioned along the approach course this would have implied that they passed already the MDA of 560 feet. If a missed approached had been initiated at the GPWS "minimums", 12 seconds before the first impact, the aircraft would likely have cleared the hill. Analysis also indicated that if an aggressive missed approach procedure was initiated 6 seconds before impact, that is the moment the plane was far through MDA and the runway was still not in sight, it is possible that the plane might have cleared the terrain.

4.3 Analysis of the Fixation Patterns

Analysis of the pilot's behavior on the basis critical variable SI focuses on the question, what do the pilots correct, although the result of their actions is undesired for the parties involved. When looking for correct actions, the almost correct descent profile is remarkable when the lateral position of the airplane is left out of consideration for the moment. At GUAM the distance measuring equipment (DME), the (UNZ VOR D0.0), is located 3.3 miles outside the airport, whereas normally the DME is located at the airport and rarely located outside the airport. Assuming that the crewmembers had a misconception about the location of the DME, they might have believed that the plane was 3.3 miles closer to the airport, which could explain the descent pattern shown in Figure 2 (NTSB, 2000).

Furthermore, the Air Line Pilots Association (ALPA) stated that although most nonprecision approaches are presented in a series of step-down altitudes, they are unacceptable for these category airplanes. These step-down altitudes are in fact directly contrary to the underlying concept of stabilized approaches because they require multiple power and pitch changes to be flown as charted [7]. This could explain why the captain prematurely reset the intermediate safety altitudes. Apparently the pilots concentrated on following a desired descent profile, a stabilized approach without the (NTSB) preferable but not required step-down fixes. By doing so, they lost touch with the actual desired result, a safe landing at the airport. When the descent profile is shifted 3.3 miles closer to the airport as is done in Figure 2 (represented by the dashed line), the approach would have been almost perfect

In case of unexpected behavior, the frame of reference of the subject has been changed such that the actual situation does not influence in negative sense the SI of the subject(s). Fixation is a form of behavior that shows this hypothesis clearly. From the NTSB report and the CVR transcript it becomes clear that in the last minute the salient warning signals of the GPWS indicating a serious problem in the actual situation have been neglected. When these signals had initiated a change of course of actions the crash could have been prevented. Though, since these GPWS warning signals represent a problem in accordance with the actual situation with the accompanying frame of reference, these signals do not affect the subject anymore since he changed his frame of reference to maintain his sense of influence.

The next question is how did the captain (and the crew probably) winded up in a fixation. The SI-model claims that someone's SI threatens to disappear when the expected result of his actions is uncertain. The uncertainty of the captain manifests itself clearly at 01.39:56 when he stated, "yes, yes it is working" at the question of the flight engineer if glideslope was working. Even so he reminded the crew that glideslope was inoperative in his approach briefing. There are two more clues indicating the captain's uncertainty about the nonprecision instrument approach. The first though is not a firm clue: the captain's cursory briefing of the instrument approach procedure which was likely due to a strong expectation that a visual approach was possible. The other firm clue was his question, "isn't glideslope working?" at 01.41:46.

The hypothesis of the captain's uncertainty about a nonprecision instrument approach is supported by the following statements of Air Line Pilot Association (ALPA) and a British Airways Boeing 777 captain who was a member of the Controlled Flight Into Terrain (CFIT) Awareness Task Force. They stated that nonprecision instrument approaches generally are much more complex than precision approaches and for many pilots they are less familiar. According to the ALPA, limited data indicate that airline transport crews conduct only about one to three nonprecision approaches per year and practice these approaches in a simulator "just as infrequently" [7].

The consequence of the analysis on the basis of the SI-model should be that the statement of the captain at 01.39:56 together with his approach briefing would immediately lead to the diagnosis a loss of the captain's SI in the actual situation and foreseen frame of reference, and consequently a change of frame of reference to maintain his SI. In this case the time to intervene and initiate a recovery from fixation would have been 2.5 minutes.

A second opportunity to diagnose loss of SI was the question of the captain at 01.41:46. In this case the time to intervene would have been much less (38 seconds) but still more than the time left with the 500 feet warning of the GPWS. Note that the newly proposed intervention techniques should reckon with the frame of reference of the fixated person and consequently should not represent a problem with the actual situation as the current alarm systems do.

5 DISCUSSION

Human factors research has shown that a common decision-making error is for humans to have a tendency to ignore evidence that does not support an initial decision; especially in high stress and workload situations. Human operators tend to seek (and therefore find) information that confirms the chosen hypothesis and to avoid information or tests whose outcome could disconfirm it, which produces inertia that favors the hypothesis initially formulated [13].

The aim of this study is to find ways to intervene in fixation such that the intervention results in a recovery from fixation. A prerequisite for such successful recovery is that the evolving event should show opportunities and time to intervene; both conditions are satisfied. The first sign of uncertainty of the captain was in the approach briefing. The evolution of the event took almost three minutes, from the first firm sign of the captain's uncertainty until the last GPWS warnings.

Analysis of the evolution of the event in hindsight showed that it is possible to point out several critical moments at which the crew could have decided to cancel the approach and to wait for a better glideslope status or to prepare properly for a nonprecision landing without the glideslope. It showed also that the alarm signals were not sufficient to change the plan of actions of the fixated crew in this case. The explanation for this "ignorance" of the alarm signals is that they represent only the state of the airplane and they do not reckon with the state of the captain (or the crew).

It is necessary to put a finger on the essence of fixation, before we can develop means to intervene in fixation in addition to the existing means that aim at error detection and error recovery. Yet in case of fixation the existing means for error detection and error recovery seem to be insufficient. We believe that fixation coincides with a change in context of the crew due to a threatening loss of sense of influence. If we can detect the context to which the fixated crew is acting only then we can find ways to intervene in fixation and recover from fixation.

To identify fixation without knowledge of the outcome requires a further validation of this model and a further assessment and sophistication of the methodology based on new cases of which the outcome is unknown beforehand. Nevertheless, the viewpoint presented in this paper based on autonomous systems could give new perspectives regarding unexpected behavior.

6 REFERENCES

[1] Endsley, M.R. (1995) Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors, 37: 32-64.

[2] Kanse, L., T. van der Schaaf (2001). Factors influencing recovery from failures. In Proceedings of CSAPC'01 (Neubiberg, Germany, September 2001), EACE Conference Series, 123-135

[3] Keyser V. de, D.D. Woods (1990). Fixation Errors: Failures to Revise Situation Assessment in Dynamic and Risky Systems. In: Colombo AG and Bustamante ASd, eds. Systems Reliability Assessment. Dordrecht: Kluwer Academic Publishers, 231-252.

[4] Mattheeuws, A., (1990). Het spel der vanzelfsprekendheden. Systeemtheoretisch Bulletin, 8: 255-287

[5] Maturana, H.R., F.J. Varela (1980) Autopoiesis and Cognition. Dordrecht, Holland: D. Reidel Publishing Company.

[6] Moray, N., I. Rotenberg (1989). Fault Management in Process Control: Eye Movement and Action. Ergonomics, 32: 1319-1342

[7] National Transportation Safety Board (2000) Controlled Flight Into Terrain Korean Air Flight 801, Boeing 747-300, Nimitz Hill Guam. Aircraft Accident Report NTSB/AAR-00/01. Washington DC

[8] National Transportation Safety Board (1994) A review of Flight-crew-Involved Major Accidents of U.S. Air Carriers, 1978 through 1990. Safety Study NTSB/SS-94/01. Washington DC.

[9] Rasmussen, J. (1983) Skills, Rules, and Knowledge: Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. IEEE Transactions on Systems, Man, and Cybernetics, SMC-13: 257-266.

[10] Reason, J. (1990) Human Error. New York: Cambridge University Press.

[11] Reason, J. (2000) Human error: models and management. BMJ, 320: 768-770.

[12] Van der vlugt, M., P.A. Wieringa (2003) De-fixating the Operator, in proceedings of IFAC'03 (Göteborg, Sweden, September 2003)

[13] Wickens, C.D. (1992). Introduction to Engineering Psychology and Human Performance. New York: Harper Collins.

[14] Woods, D.D., L.J. Johannesen, R.I. Cook, N.B. Sarter. (1994) Behind Human Error: Cognitive Systems, Computers, and Hindsight. Columbus: CSERIAC.

[15] Xiao, Y., C.F. Mackenzie (1995). Decision Making in Dynamic Environments: Fixation Errors and Their Causes. In Proceedings Human Factors and Ergonomics 39th Annual Meeting, 469-473.

CHLOE: A Technique for Analysing Collaborative Systems

Angela Miguel
Department of Computer Science
University of York
York, UK
angela.miguel@cs.york.ac.uk

Peter Wright
Department of Computer Science
University of York
York, UK
peter.wright@cs.york.ac.uk

ABSTRACT

CHLOE is a technique designed to help identify possible failures within collaborative work. It analyses both social collaborative (H-H) and technology-mediated collaborative (H-C-H) work. A question-based approach is taken by the method, in an attempt to reduce the requirement for analysts to have substantial knowledge of human factors or cognitive psychology. The CHLOE method consists of 4 stages: scenario description, task identification, error analysis (using questions based on a cognitive model of collaboration), and design suggestions. The method has been applied to a scenario of collaborative work by the designers of the method in order to assess its effectiveness. A small evaluation study has also been performed to assess the wider usability of the method.

Keywords

Collaborative Work, Collaborative Error, Human Error Analysis

1. INTRODUCTION

The importance of evaluating a design for possible human error is well recognised, and there are many different techniques available for performing this task which take several different approaches. However, despite the fact that most work takes place in groups or teams, most human error analysis techniques focus on errors that may happen during the interaction between a single individual and the system they are using. Collaborative work is susceptible to a different type of error. These are errors that emerge as a result of the distributed knowledge that this type of work involves, which places extra demands on participants. Collaborative errors may be caused by factors such as a lack of situation awareness or awareness of each other, misunderstandings between participants, conflicts, and failures of co-ordination. Therefore an alternative error analysis method is required. CHLOE has been developed not only to analyse collaborative work, but also to help suggest design improvements to support this type of work.

As Hollnagel [3] points out, the term error can have three different meanings. An error may be considered as the actual incorrect action that has been carried out, the visible consequences of that incorrect action, or the abstract cause behind the incorrect action and/or the visible consequences of this. The CHLOE error analysis questions tackle these cognitive causes (cognitive failures or errors) behind the error actions and consequences, and the reasons for these. Questioning the cognitive reasons for failures can help lead to design solutions because it focuses on *why* the observable failures occur, so improvements can be made to design on this basis.

In order to link the human error analysis section to redesign considerations, CHLOE takes a question-based approach to analysis. The questions link analysis and redesign because each one prompts the analyst to look for a particular aspect within the system being analysed that may cause collaborative problems, while at the same time providing information on the requirements of that aspect of the system to enable it to encourage successful collaboration.

Taking a question-based approach to error analysis potentially has some additional advantages. Questions should be easy to understand and simple to apply to the system. A question-based approach can help to guide the analyst to the correct areas or factors to examine for the possibility of error. THEA [7] has shown that this approach to error analysis enables non-'human factors' experts to carry out the analysis, assuming they have knowledge of the system.

2. THE CHLOE PROCESS

The CHLOE process consists of several stages (see Figure 1): Scenario Description, Goal Decomposition, and then Error Analysis. The Error Analysis questions are based on a model of collaboration. Finally, Design Issues are considered according to what has been discovered in the Error Analysis stage. Each of these stages is now described in more detail.

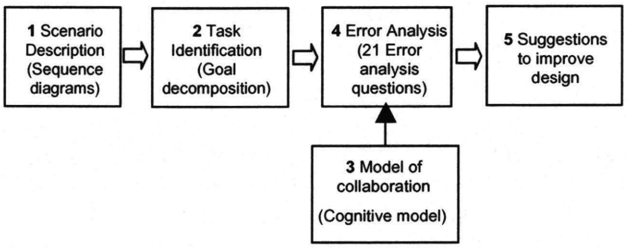


Figure 1: The CHLOE Process

2.1 Scenario Description

CHLOE uses scenarios to describe a system for error analysis purposes. There are several advantages to using scenarios, including the fact that they allow the system to be analysed in sections. This is often preferable to the alternative of trying to analyse a whole system at once, which will in many cases be impossible. Full models of systems also often show only how work *should* be done. Scenarios can be used to show many different possibilities of how work may be done. They can also show the dynamics of a system, which may become too complicated if trying to cope with an overall system view. Scenarios are also more flexible and can be used at several different stages in the life cycle of a system. They can therefore easily be used for both analysis and design.

Sequence diagrams are used to describe the scenarios in CHLOE (see Figure 2). They provide information on the sequence of actions, who communicates with whom and with which artefacts, and an indication of what information is passed between people and the technology they are using. The purpose of the sequence diagram is to provide background and contextual information about the scenario in full, which can then be used to help answer the error analysis questions to carry out the analysis. It is particularly important when analysing complex collaborative work to have a structured means of collecting the information required to answer the error analysis questions. Agents (human, computerised, and non-computerised) are shown along the top of a sequence diagram, with a timeline for each extending downwards. Arrows linking the timelines of agents show interactions between agents, including information passed between them. A dashed line indicates a reply to a previous interaction.

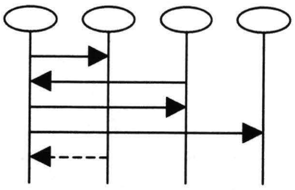


Figure 2: The structure of the sequence diagrams used for scenario description

2.2 Task Identification

In order to analyse a scenario, it must be structured in a way that supports the error analysis process. The sequence diagram provides the contextual information, but it contains too much detailed information to help structure the error analysis process. It would be unreasonable to carry out an analysis of each interaction on the sequence diagram individually. Analysing individual activities would also neglect the wider issues of how participants collaborate and how their actions/interactions combine to complete the task. In order to analyse the scenario at a higher level of abstraction a method is required to abstract out the tasks in a structured manner. A Hierarchical Goal Decomposition is applied to the scenario to achieve this. It also serves to structure the error analysis. The technique used is similar to Hierarchical Task Analysis (HTA) [4], but unlike HTA, no plans are required. This is because the method of carrying out the task is not being specified (this information is already in the scenario description); instead, the goals/sub-tasks that comprise the scenario are being identified.

2.3 A Model of Collaboration

Collaborative errors, like individual errors, can be regarded as stemming from failures in cognitive processing. The error analysis stage in the CHLOE method (see Section 2.4) is based on a cognitive model (see Figure 3). This model has been developed from a basic framework of collaboration by Dix [2] and shows various types of communication involved in collaborative work. As this model focuses on communication, failures in collaboration are framed as being failures in communication and understanding between participants. The model is composed of participants (P), an artefact of work (A), different possible types of communication, shared understanding and a simple cognitive loop [6]. This loop represents the cognition of each participant when interacting with other humans or machines.

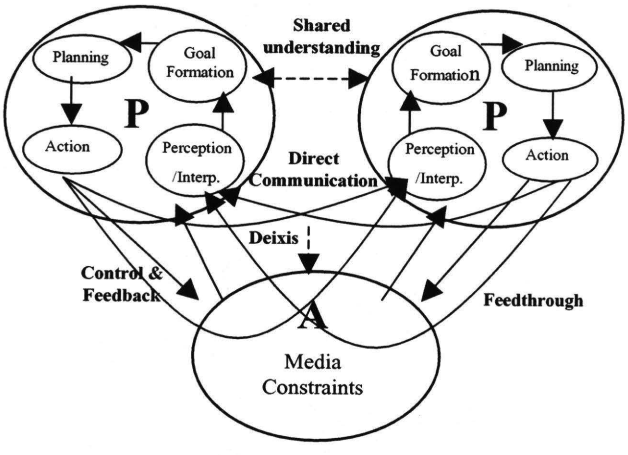


Figure 3: Model of Collaboration Used in CHLOE

The arcs in the diagram indicate how the participants and the artefacts of work are linked in each type of communication that may be involved in collaboration. They show how the actions[1] of one participant are perceived (directly or indirectly) and interpreted by the other participant, who then performs an action in reply. This is then perceived by the first, to form an ongoing cycle from action to perception between the participants[2]. Communication through the artefacts used, goal formation and planning are optional within this process.

The types of communication between participants and artefacts shown are:

- **Direct Communication** - The actions of one participant are perceived directly by the other. Therefore, after perceiving the actions of another he/she will interpret what they have perceived and go through whatever stages of the cognitive loop are necessary to carry out an action in reply, or the next action required (see Figure 4).
- **Deixis** – This describes the situation in communication where one person is speaking to the other(s) but is pointing or referring to something else, e.g. a common artefact (see Figure 4).

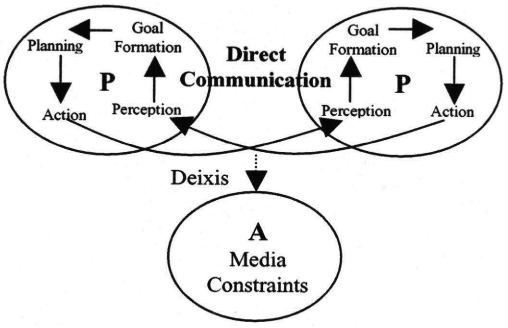


Figure 4: Direct Communication & Deixis

- **Control and Feedback** - One participant carries out an action on the shared artefact and then perceives the

---

[1] Talking is considered to be an action in this case, in addition to more physical actions.

[2] Everything inside the ovals, labelled P for participant, represents the internal workings of each participant

result of this action. This is not in itself a form of collaborative communication (see Figure 5).

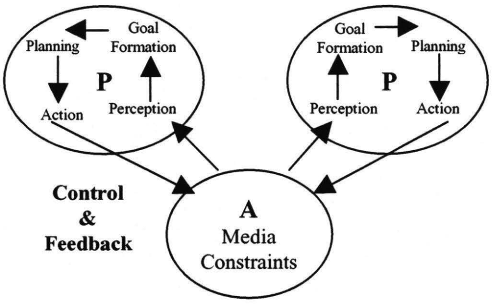


Figure 5: Control & Feedback

- **Feedthrough** - Communication takes place between participants through the artefacts of work. In this case, the actions of one participant on the shared artefact are perceived by the other participant, not directly, but *through* the artefact (see Figure 6).

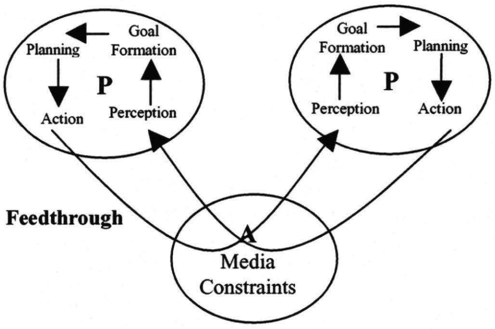


Figure 6: Feedthrough

- **Shared understanding** - All the different types of communication shown can help to create and support shared understanding, which in turn helps collaboration to work effectively

2.4 Error Analysis

2.4.1 Breakdowns in Collaboration

Collaborative failures are considered to be caused by breakdowns in cognition of the participants involved. The error analysis questions in CHLOE have been developed by applying guidewords to the collaborative model (see Figure 3) in order to examine where breakdowns may occur within the model, and the type of breakdowns that may occur. Basing the error analysis on failures within a cognitive model of collaboration allows the reasons behind the visible failures to be tackled.

The guidewords used were selected from the SUSI [1] modified HAZOP [5] technique, which was developed to analyse user-system interaction. In SUSI these guidewords were applied to the data flows, data stores and processes that make up a socio-technical system, to identify the ways in which that system may fail. CHLOE focuses on the people within the system and the communication between them, and examines how this may fail, leading to collaborative failure. CHLOE bases the error analysis on failures within a cognitive model of collaboration. Therefore the guidewords have been applied to the cognitive stages in the model of collaboration to identify cognitive failures in collaborative communication. These are the potential reasons behind failures of the data flows and processes that make up the collaborative system. By identifying the root causes of collaborative failures, design suggestions can be made to improve the system. The guidewords used describe 4 different types of failure (see Table 1).

Table 1: Cognitive Stages and their Failure Guidewords

Cognitive Stage	Failure Guidewords
Perception/Evaluation	(Corruption) Error Failure/None
Goals (Triggering & Initiation)	Conflict (Corruption) Error Failure/None
Planning (Coordination)	Interrupted/Incomplete (Corruption) Error Failure/None
Action	(Corruption) Error Failure/None Conflict Interrupted/Incomplete

Collaborative work (involving any of the types of communication shown in the model (see Figure 3)) may break down because of problems at any of the cognitive stages included in the model. Table 1 shows the four cognitive stages and the possible failures of them that were used. These failure guidewords were used in combination with the cognitive stages to create examples of how collaboration may fail. The analysis questions developed from these failures are therefore generally concerned with the following in collaborative work:

- failures of perception/evaluation – e.g. the collaborating participants may not all perceive or interpret the information in the same way, or may not be able to perceive what each other is doing
- failures of goal formation or triggering of tasks – e.g. each participant may not know what he/she is supposed to be doing and when, or their goals may conflict
- failures of planning and co-ordination – e.g. participants may understand incorrectly what they are supposed to do or the work may become disorganised and,
- failures of the actions themselves – e.g. one participant's incorrect actions have direct disastrous consequences for others' work, or no-one else can correct the incorrect actions once they have been done

If these types of failure shown in Table 1 are then considered in relation to the types of communication shown in the model of collaboration, more specific examples of failure can be created. Some examples of these types of failure are:

- Failure of Perception in Feedthrough – e.g. the second participant cannot see the result of the first's actions
- Failure of Action in Control and Feedback – e.g. one participant cannot carry out their actions because the system only allows one person to work on it at a time
- Error of Perception in Deixis – e.g. one participant refers to something, but the other participant misunderstands what is being referred to

- Conflict of Goals in Direct Communication – e.g. the participants communicating have conflicting goals.

Forty-eight possible collaborative failures were created from the combination of the twelve guidewords and cognitive stages, and the four types of communication in collaboration according to the model used. Questions were then developed around the possible reasons for these failures for each type of communication. From the original set of questions created, 21 questions were chosen to identify the main causes of failure for all types of communication. The question set was reduced from the original by removing repetition in the questions and then by selecting the questions that would cover the most important and most common types of collaborative failure in many different types of collaborative system.

The process of creating the error analysis questions has resulted in questions that can capture a range of potential causes for breakdown within a collaborative system. Different guidewords, or a greater number of guidewords could have been chosen to develop the questions, but the ones selected provide a good range of potential breakdowns for all cognitive stages. They tackle four basic types of breakdowns that may occur to prevent the cognitive stages from functioning correctly. Only a limited number of specific failures can be considered for each communication type because it would result in the requirement for more questions. The number of questions in the analysis must be kept to a minimum because the task-based approach to analysis involves the repetition of these questions for every task analysed.

2.4.2 The Error Analysis Questions

The 21 questions that form the error analysis section of CHLOE are split into four sections according to the cognitive stage they analyse. The collaborative aspects of the model have been captured in the error analysis questions, and so only the cognitive stages questioned as possible points of breakdown remain immediately visible as question sections. Six questions tackle perception/interpretation and evaluation in collaborative work. Problems in this area of collaborative work can cause a lack of situation awareness and misinterpretations of the system state or the actions of others. Five questions are concerned with goal formation. Problems in this area could result in collaborative failure because of actions not being carried out when they should be, participants having incorrect goals (leading to misunderstandings), or participants' goals conflicting. Five questions deal with planning and co-ordination problems. With many people working together it is important that everyone knows their role, when to carry out their tasks, and that work is co-ordinated so participants may work effectively without interfering with each other. Finally, five questions are about the actions that the participants must carry out as part of collaborative work. These consider conflict, access to resources, and ensuring that participants do not undo each other's actions. Example questions include: (Goals Q5) *'Are participant's goals or sub-goals likely to come into conflict (e.g. same resources required etc.)?'* and (Planning Q4) *'Is there a shared representation which is consistently visible and understood by all concerned, which can be referred to (e.g. pointing) when sharing information?'*

All 21 Error Analysis questions are applied to each task/activity as broken down in the Hierarchical Goal Decomposition for a thorough analysis. The information contained in the scenario description is used to help answer the Error Analysis questions.

To structure the answers and any resulting design suggestions, the tabular format shown as Table 2 is used.

Table 2: Tabular Format for Recording Error Analysis Results

Questions	Causes/Consequences	Design Issues
CHLOE Error Analysis Question	Issues raised by the analyst and the possible consequences of these	Suggestions about possible improvements /re-design ideas

2.5 Design Suggestions

The final stage of the CHLOE process is to think about possible design improvements to the collaborative system, which may help to reduce the likelihood or severity of errors/breakdowns in the areas susceptible to error, as highlighted by the error analysis. These design considerations can either be filled in after the error analysis has been completed for the whole scenario, or at the same time as each error analysis question is being answered for each task. As discussed in the previous section, a significant advantage of the question-based approach used by CHLOE is that design suggestions can grow from ideas provided through the questions and their corresponding answers, which highlight exactly what or how the current design of the system is lacking. Thus, each question and answer pair should help to indicate the requirements that a re-design needs to fulfil in order to better support collaborative work, and more specifically, what the difficulties with the current system are that need to be designed out. The information required to carry out this stage in the CHLOE method is therefore readily available, and only domain knowledge is required to produce design suggestions to improve the system.

3. CHLOE CASE STUDY

In order to test the effectiveness of CHLOE, it has been applied to a scenario of considerable size involving a lightning strike to an aircraft and the subsequent actions that must take place to stabilize the aircraft and land safely. This scenario involved several participants and artefacts including: a Pilot Flying, a Pilot Non-Flying, Air Traffic Control, a Tactical Communications Officer, and a Radar Operator. The information required to construct this scenario was collected from observations within a flight simulator. Two simulations of this scenario were observed.

From the data collected during observation of the simulations, scenario descriptions were created using sequence diagrams to detail the main interactions between the agents involved. These were then aggregated, in order to create a more generic version of how work takes place in the scenario. Goal Decomposition was then performed on the scenario to identify the tasks that must be achieved to attain the goal of the scenario, and which need to be analysed through the application of the Error Analysis questions. In this case the main goal was to stabilize the aircraft and return home safely. The full Goal Decomposition consisted of 3 levels and was broken down to 14 sub-tasks at the third level. An error analysis was performed on every one of these tasks to achieve a complete analysis of the scenario. However, this is not compulsory, and the analyst can choose to analyse only the important tasks if necessary.

From the application of CHLOE to the whole 'lightning strike' scenario, 128 potential difficulties for collaborative work were identified. Potential problems of all types were identified, but *Perception/Evaluation* contained the most problems. Table 3 details two perception problems highlighted from two different tasks in the scenario.

Table 3: Extract from the completed CHLOE Error Analysis

Questions	Causes/Consequences
Perception Q4 – Is it obvious to all that someone has altered the system/carried out an action, and if necessary who made these changes, so that this action is not repeated unnecessarily or undone by another?	It is not obvious that an action has been carried out unless both pilots know what the settings should be. A lack of situation awareness could result, and misunderstandings between the pilots will become more likely. One pilot may undo the others actions or try to redo them. It also means that no crosschecking will take place on the actions carried out.
Perception Q5 - Does the system interface/ the media/artefacts used enable the information from this action to remain available for checking and remembering purposes?	Information from ATC does not remain available for checking. This means there is only a brief chance to ensure that the pilot not talking to ATC also heard the message for checking purposes. This may easily be missed, and therefore reduces awareness of the situation and crosschecking ability.

The fact that the majority of the potential problems identified within this scenario were in the *Perception/Evaluation* section suggests that the greatest problem in this scenario is with clearly getting the information required. However, the *Planning* and *Actions* sections also had many potential difficulties. Perhaps inevitably, the more complex tasks involving the most collaborative effort are subject to the greatest number of potential difficulties. As would be expected however, the likelihood and severity of these failures varied considerably. A severity rating in some form is a useful means of highlighting which problems are most likely or have the greatest consequences. This can make the task of choosing which areas need re-design attention most urgently, slightly easier. A simple high, medium or low rating scale could be used as in THEA [7]. Alternatively a more detailed scale or matrix in which to mark the severity of the potential problem could be employed.

Redesign suggestions were also completed for the scenario, but due to a lack of detailed domain knowledge and technical expertise these were quite superficial suggestions, which only stated generally how the design needed to improve upon the current situation. CHLOE is aimed at engineers with little human factors knowledge, but who *will* have detailed technical knowledge of the system being analysed.

4. EVALUATION

A small evaluation of the CHLOE method has been performed in order to discover whether other analysts were able to use it successfully. A short Air Traffic Control scenario was used for this evaluation. This scenario focused on the path of one aircraft across a sector and the actions to be taken to ensure it did not come into conflict with other aircraft in the sector.

Seven participants were recruited. Two of these participants were very familiar with Human Factors and Human Error Analysis. Another two had limited awareness of some Human Error Analysis techniques, but did not have much knowledge of Human Factors. The final three had no knowledge of either of these topics. All of the subjects were either PhD students or Research Associates in the Department of Computer Science.

Firstly, the CHLOE method, its aims, and how to apply it were described briefly to the participants. Additional written material about the method was also provided. This included example sequence diagrams and goal decompositions in the description of the method and a list of the CHLOE error analysis questions along with explanations of each of these questions. The participants were also given a six-page description of Air Traffic Control to provide them with basic domain knowledge to allow them to answer the CHLOE analysis questions. This description included pictures of the environment in which the controllers sit, and diagrams of flight progress strips and strip boards to help them to visualise the domain. Also included in this material was a detailed description of the scenario that the participants were to analyse using the CHLOE method. They were allowed to read this material at their own pace and the second part of the evaluation took place the following day.

The participants were required to apply the full CHLOE method to the Air Traffic Control scenario. No time limit was applied. Afterwards they completed a questionnaire about the method and its application. Some of the participants were also interviewed later about the answers they had provided in the questionnaire in order to clear up ambiguities and collect additional information about their answers.

The questionnaire was designed to draw out any doubts or dissatisfaction with the method so that improvements can be made. It was split into five sections: modelling, error analysis questions, re-design issues, usability, and effectiveness. The first three sections questioned: the diagrams used to model the work for analysis, the ease of using the questions to perform the analysis, how easy it was to understand how and where the system needed to be re-designed, and how much support the questioning approach provided with this. The usability section was concerned with how easy it was to use the method and how much human factors knowledge the users felt they needed in order to apply it successfully. Finally, the effectiveness section questioned whether or not the participants felt the method had helped them to systematically identify potential collaborative problems.

In order to assess the results of this questionnaire, the participants were split into two groups. Those who had both knowledge of human factors and human error analysis were taken as one group (experts). All the others participants constituted the other group (novices).

4.1 Evaluation Results

The expert and novice groups often had quite different opinions about aspects of the CHLOE method and its success, but there was no major difference in the actual success of the application of the method by the two groups. Several important issues with the CHLOE method have been raised both through the views of the participants expressed in the questionnaires and interviews, and through the actual performance of these participants with the method.

4.1.1 Modelling Difficulties

Both groups thought that the sequence diagrams were easy to create. Within both groups there was variation in the quality and amount of detail in the diagrams produced. Choosing the agents was the most problematic area. The most severe mistake was made by a member of the novice group whose sequence diagram included only the humans in the scenario as agents. All actions were included, but not making artefacts as important as the paper strips into agents made it more difficult to capture the interaction in the diagram.

While most of the novice group thought that all the necessary information to capture collaboration was present in the diagrams, the experts disagreed. The experts' view is supported by evidence in the sequence diagrams actually created. Several of the analysts in both groups embellished their diagrams to try to display the additional information that they thought they needed to help them perform their analyses. Examples of this are numbering the sequence diagrams with the task numbers associated with each section of the goal decomposition, and adding overhearing information. The first of these suggests a desire for a closer relationship between the diagrams and between modelling and analysis. The experts also expressed this view that there was not a close enough relationship between the modelling of the scenario and the analysis process, and some analysts from both groups thought the diagrams were difficult to relate to one another. The second example demonstrates that the modelling does not show all the information that the analysts would like to perform their analysis.

The goal decompositions within both groups also showed problems. In each group there was one analyst who created a goal decomposition that was unnecessarily large and complicated for the scenario to be analysed. The goal decompositions of the rest of the analysts, including those who split the main goal into the same number of tasks, were all very different from each other. Therefore different tasks were analysed by most of the analysts.

4.1.2 Error Analysis Difficulties

The overall views expressed by the experts and the novices about the error analysis questions were very different. The experts were critical of the questions mainly because they felt that they did not direct the analyst as much as required. There was also nothing to motivate them to provide the details of the potential problems that were identified using the questions. Hence, it is possible to perform an analysis at a very superficial level using CHLOE. Such an analysis would provide no help in isolating the particular causes of problems in collaborative work or help with redesigns. The novices found the questions easy to understand, thought that they directed the user to the relevant issues, and also that the questions tackled issues affecting collaborative work by showing what the work required to be successful. Some potential problems with the questions were highlighted though. These concerned the potential overlap between some questions, the completeness of the questions, and the potential effort involved in applying the questions if the tasks or level of analysis has been chosen poorly.

Despite the different points of view concerning the questions, the error analyses of both groups displayed similar problems. All of the analysts provided simple Yes or No answers to many of the error analysis questions with no additional explanation. Other answers were often not detailed enough to provide anything more than a general indication that there was a problem in a certain area of the system. They were therefore not particularly helpful for showing what the specific problem was for collaboration within the system and feeding information into the redesign process. There was also evidence of potential misunderstanding and misapplication of the questions within the analyses by both groups. For example, providing inadequate reasons for the answer given and missing obvious problems. Some of the questions are clearly open to some interpretation because different analysts have used the same argument for different answers.

Although there were some problems with the error analyses performed, many potential problems were highlighted by the analysts in the novice group despite their lack of human factors knowledge. Many of the answers given are also particularly insightful given this lack of human factors knowledge.

4.1.3 Redesign Difficulties

Redesign suggestions were regularly not included where they should have been. However, where they were provided they were mostly of good quality. They tackled the issues raised by the question and the specific problem raised in the answer provided. The absence of many design suggestions is partly explained by the analysts' lack of domain knowledge. However, there are two potential difficulties with the method here that were highlighted by the analysts. These are that the method does not motivate the analyst to provide design suggestions as the questions simply point out that there is an issue, and the fact that some questions identify problems that did not necessarily lead obviously to design issues. Also, the suggestions made do not always adequately tackle the issue in question. The error analysis questions therefore do not prompt the analyst to consider the issues in enough depth.

The fact that the error analysis questions do not force the analysts to provide details of a potential problem and its causes and consequences was blamed for difficulties with the redesign suggestions. If an analysis is performed at too general a level and these details are not provided, there is nothing to feed into redesign considerations.

The novice group did express strongly that they thought the question and answer pair did help the analyst to see how the system should be redesigned though. There was not such a convincing response concerning whether the method showed where redesign was required though, and the need for prioritization of the design improvements was raised again.

The lack of a means to prioritise the analysis results was recognised by both groups as a weakness of the redesign section of the method.

4.1.4 Approaches to Performing the Analysis

Some interesting usability issues were raised through the evaluation of the CHLOE method with users other than its designer. Instead of fully analyzing one task at a time, the analysts took each question in turn and applied it to all the tasks. One analyst commented that this approach should lead to a more consistent application of the questions.

There was variation in the way that the analysts applied the method within both groups. Only one analyst in each group answered the questions for each task individually. One analyst in the novice group went to the opposite extreme of answering all the questions for all the tasks at once. However, most included elements of both, by answering questions for individual tasks sometimes, and for two, three or all tasks at others. Around half of the analysts thought that they had considered tasks specifically within the context of the scenario. Others sometimes considered the tasks or the system in general when answering the questions. One expert, who did not consider each specific task, explained this by saying that he found the scenario very restrictive. He wanted to consider difficulties in the system more generally for the whole analysis.

Of particular interest however, is the fact that one novice analyst often answered the error analysis questions with reference to problems with agents specifically, rather than tasks or even the scenario as a whole. For example, to answer some questions the focus was placed on problems with the layout of the flight strips, and these problems were identified for the system in general rather than for any particular task or tasks. This approach led to one of the most effective and detailed analyses produced by either evaluation group.

The experts and novices had similar views on what the greatest problems with the method were in terms of usability. These were mostly related to the creation of the goal decomposition. In particular choosing the right level of analysis and the optimum number of tasks. Interpreting the questions in the context of the task within the scenario was also highlighted as being difficult. Suggestions to improve usability included more guidance on creating good diagrams and linking them, tool support, and prioritization of redesign requirements. But the novice group agreed that the method did not take long to learn and was not difficult to use. It was also, and most importantly, thought that knowledge of Human Factors was not required to use the method successfully.

4.1.5 Effectiveness of Analysis

All but one of the analysts successfully found many potential problems within the scenario examined. However, the number of problems found by each of these analysts varied substantially. The lowest number identified was nineteen, and the highest was fifty-one. Much of this variation was caused by the way in which the tasks were split in the goal decomposition, and the amount of detail that the analyst chose to provide when answering the question. The problems identified were described at different levels of generality because of these differences.

Despite the different opinions on the effectiveness of the method, the two groups performed similarly. The experts' analyses were not any more successful than some of the novices, despite their human factors knowledge. Both groups also displayed the same difficulties. The most common problem was that the analysts sometimes provided answers that gave nothing more than a very general indication that there was a problem of a certain kind for a particular task. The details of the cause of a potential problem were not identified. The redesign section was therefore not always successful because the questions and their answers did not always prompt the analyst to provide a design suggestion.

The greatest problem concerning the effectiveness of the CHLOE method is the lack of consistency between the problems found by the analysts. Only a small proportion of analysts identified the same problems in the scenario for a given question. This was true both within and between the two groups. Each analyst put more or less focus on certain aspects of the system according to their goal decomposition. Therefore the different analysts saw different problems within the scenario. The analysis process was therefore not as thorough and systematic as it should have been.

Comparisons between the potential problems discovered by each analyst were difficult because of the different levels of generality at which the problems are described. This also led to very different results in terms of the numbers of problems found. All of this resulted from the different goal decompositions created by the analysts. While those who split the main goal into four tasks to analyse mostly found around 20-30 problems in the scenario, one analyst who split the main goal into six tasks to analyse discovered over fifty potential problems.

While the novices were happy that they had thoroughly and systematically analysed the scenario for potential collaborative error using CHLOE, the experts were not. They suggested that not enough structure had been imposed on the analysis process to provide a thorough and systematic analysis. This was blamed partly on the lack of a close link between the modelling and the analysis, which makes it more difficult for an analyst to recognise easily that something has been missed. The other part of the problem is that analysts create their own goal decompositions. These can vary greatly even for a small scenario. The focus on communication was also criticised as some of the more complex aspects of cognition that affect collaboration are neglected. The fact that the analysis was led by the tasks in the goal decomposition was criticised by the experts as it meant that the focus was not on collaboration to the extent that they thought it should be.

5. CONCLUSIONS

The CHLOE method has taken a question-based approach to the analysis of collaborative work. This approach was taken because it could potentially help lead to redesign suggestions, and reduce the need for human factors knowledge to enable the analysis. The application of CHLOE to scenarios by both the designers and other users has shown that the technique can help to highlight numerous potential difficulties within a collaborative system. It has also suggested that those without human factors training can perform an analysis as well as those who do. In particular, these novices have confirmed that the use of the question and answer pairs are helpful for the creation of redesign suggestions. However, the evaluation has also highlighted a number of problems with the method.

CHLOE has some problems that it is possible to correct quite easily. Examples of this are: the information missing from the diagrams to more effectively represent collaborative work, some of the error analysis questions not tackling specifically collaborative problems or helping redesign, and the redesign section needing a severity rating in some form. However, deeper problems with the approach taken by CHLOE have also been highlighted. These have led to a lack of consistency between different analysts using the method. The method does not force the analyst to be systematic or thorough because there is too much freedom allowed in its application. The fact that the analysts may analyse different sets of tasks, coupled with the lack of a strong link between modelling and analysis, allows aspects of the work being analysed to be easily missed.

Currently CHLOE also allows a superficial analysis to take place. The error analysis questions only direct the analyst generally as to what to look for, but not where. The analyst is also not forced to consider all the activities within a task because there is no external motivation to provide any details when answering the error analysis questions. This allows him/her to simply identify very general problems at the task level that do not get to the heart of the problem or help redesign. This problem is increased by the temptation for the analysts to group tasks together because it makes analysis faster and avoids repeating the questions so often. However, it increases the chance that the analysis focuses on certain areas and neglects others.

Most of the major problems with the CHLOE method are the result of design decisions taken to minimize the effort that analysis takes. Using a scenario and task-based approach means that many scenarios must be analysed in order to adequately analyse a system. Also, if a scenario is long, it will be split into many tasks in the goal decomposition. As the method has to be applied to each task, the amount of effort required for each round of questioning has to be minimized, or the overall effort required for analysis will become too great. This places a major constraint on the number of questions that the method can have. Therefore, the CHLOE questions have been developed from a few very general failure guidewords, and the questions themselves are quite general in nature. Even with these measures, the low number of questions possible results in the questions trying to ask too much at once.

To get more from an analysis, the questions need to direct the user more precisely in order to support the identification of more specific problems that can provide more precise details for the redesign process. More questions are required in the method to be able to do this and still provide some coverage of the types of error that may affect the system. Therefore, a different approach to error analysis from the one that CHLOE takes may be required to provide a more thorough and systematic analysis of collaborative work.

Despite these problems, the CHLOE method demonstrates the potential benefits of using a question-based approach to the analysis of collaborative work. It considers issues that are important (and particular) to collaborative work, which would be missed if using other error analysis/evaluation approaches. For example HAZOP-like methods [5] that use Guidewords miss the combinations of factors that may lead to collaborative error. Collaborative work is different from individual work and is therefore at risk of failures caused by different problems. CHLOE attempts to deal with these.

Further evaluation of the method with more subjects and using true experts in human error analysis and human factors are required to conclude whether the CHLOE method can be used effectively with corrections made to the design within the current approach, or whether a new approach that supports analysis using a greater number of more specific questions is required.

6. REFERENCES

[1] Chudleigh, M.F. & Clare, J.N. (1993) The Benefits of SUSI: Safety Analysis of User System Interaction. In J. Gorski (Ed.) SAFECOMP'93. Proceedings of the 12th International Conference on Computer Safety, Reliability & Security, Poznan-Kiekrz, Poland. Springer Verlag. pp 123-132.

[2] Dix, A., Finlay, J., Abowd, G. & Beale, R. (1998) Human-Computer Interaction 2nd Ed. Prentice Hall Europe, Harlow, Essex.

[3] Hollnagel, E. (1998) Cognitive Reliability and Error Analysis Method (CREAM). Elsevier Science Ltd.

[4] Kirwan, B. & Ainsworth, L.K. (Eds) (1992) A Guide to Task Analysis. Taylor & Francis Ltd, London.

[5] Kletz, A. (1999) Hazop and Hazan: Identifying and Assessing Process Industry Hazards.4th Ed. Rugby, UK.

[6] Norman, D. A. (1988) The Psychology of Everyday Things. Basic Books

[7] Pocock, S., Harrison, M.D., Wright, P.C. and Johnson, P.D. (2001). THEA: a technique for human error assessment early in design. In M. Hirose (Ed.) IFIP TC 13th International Conference on Human-Computer Interaction. IOS Press. Ohmsha. pp 247-254.

Interaction management in collision avoidance at sea

Marcel Andro
Université Bretagne Sud (UBS)
LESAM, Centre de Recherche
Rue de Saint-Maudé
F-56325 LORIENT Cedex
Tel : (33) 0297874521
marcel.andro@univ-ubs.fr

Christine Chauvin
UBS - LESAM
CNRS, IRCCyN, PsyCoTec
B.P. 92101
F-44321 NANTES CEDEX 03
Tel : (33) 0297874521
christine.chauvin@univ-ubs.fr

Gilbert Le Bouar
Université Bretagne Sud (UBS)
LESAM, Centre de Recherche
Rue de Saint-Maudé
F-56325 LORIENT Cedex
Tel : (33) 0297874521
gilbert.le-bouar@univ-ubs.fr

ABSTRACT

Collision is one of the main dangers that threaten a boat. To manage the risk of collision, officers of the watch on board merchant vessels have to apply the Collision Regulations. Previous studies have shown that these regulations give rise to different interpretations. The present paper aims at gaining a better knowledge of the strategies used by watch officers to manage interactions with others in collision avoidance situations. It relies on fieldwork carried out in the Dover Strait. It picks out the features of the actual behaviour of watch officers on board cargo ships and on board ferries. It shows that they follow informal rules which are shared among a certain category of vessels and which may present several deviations from the formal rules. In addition it shows that watch officers on board ferries have a general strategy which aims at mastering the situation. This strategy is efficient to prevent accident or incident, but is not always applied by other kinds of vessels which also cross the Dover Strait. The discussion deals with the advantages and limits of a new support system for avoiding near-misses occurring in a busy waterway.

Keywords
Interaction management, social rules, cognitive strategies, decision making, ship handling, collision avoidance.

1. INTRODUCTION

One officer of the watch (OOW) working alone does the navigation on board merchant ships, in open water. He is responsible for navigation and bridge management activities but also for collision avoidance. In fact, conflict detection and resolution are assumed by each OOW and not by a traffic control centre. This feature distinguishes maritime traffic operation from air traffic control and means interactions between OOW can be compared with interactions between car drivers. Furthermore, in both cases, the two participants involved do not communicate or communicate very little. In both cases, conflict resolution is regulated. In fact, on board ships, each OOW has to take the Collision Regulations [1] into account. Collision Regulations define different kinds of interaction situations (crossing, overtaking and head-on situations) and different status of vessels (the 'give-way' vessel shall keep out of the way and the 'stand-on' vessel shall keep her course and speed). They give recommendations about the direction of manoeuvre (to starboard rather than to port). They provide also recommendations concerning time and amplitude of the course or speed alteration. Finally there are traffic lanes for ship handling as well as for car driving. In fact, traffic separation schemes (TSS) separate opposing streams of traffic in busy waterways, such as the Dover Strait. A vessel using a traffic separation scheme shall 'proceed in the appropriate traffic lane in the general direction traffic flow for that lane' but shall also take avoiding action when she sees another power-driven vessel crossing on her starboard side.

Studies dealing with collision avoidance show that different interpretations of the Regulations generate uncertainty concerning the actions of vessels. Uncertainty concerns:

i. the evaluation of the risk made by OOW [2]. If they do not consider that a risk does exist, they will not alter the vessel course;

ii. the distance at which the 'give-way' vessel will alter her course [3];

iii. the direction of the course alteration: port or starboard [4].

We have shown, in a previous work [5], that :

- a fairly high number of mariners achieve a sufficient miss distance by altering course to port;
- some patterns of action are misunderstood and lead the stand-on vessel to perform an emergency manoeuvre;
- as a consequence, OOW try to anticipate both the action and reaction of the other.

This study aims at identifying and analysing the strategies used by OOW to manage their interactions with others in collision avoidance situations. It relies on the theoretical frameworks of 'social interaction' and 'social rules', which leads to consideration of the informal rules governing the behaviour as well as the formal ones. This study was carried out from ecological data recorded in the Dover Strait. In this area interaction situations concern, mainly, cargo ships operating in the traffic separation scheme and ferries crossing the lanes.

The ultimate purpose of this work is to examine to what extent the combination of formal and informal rules is efficient to prevent accident and incident and to discuss the possible advantages and limits of a new support system.

2. THE THEORETICAL FRAMEWORK OF 'SOCIAL INTERACTION' AND 'SOCIAL RULES'

Social interaction patterns have been studied in driver behaviour. In this field, a social interaction is defined [6] as 'a condition in which behaviour of Driver A is determined by his organismic features as well as the environment of which the behaviour of Driver B is an integral part. Driver B's behaviour in turn is in part determined by the behaviour of Driver A, whose subsequent act is a determinant of B's next action, and so forth'.

In collision avoidance at sea, the action of an OOW is determined by the behaviour of a target ship, because his action has to be co-ordinated with this behaviour. Zhao & al. [7] describe the co-ordination of the manoeuvres with each other, in the following table (Cf. Table 1). They consider three possible actions (turn to starboard, turn to port, stand on) and eight combinations of actions.

Table 1. The co-ordination of the manoeuvres with each other

Manoeuvre	Turn to starboard	Turn to port	Stand-on
Turn to starboard	Co-ordinated	Uncoordinated	Co-ordinated
Turn to port	Uncoordinated	Co-ordinated	Uncoordinated
Stand-on	Co-ordinated	Uncoordinated	X

Four combinations of actions are uncoordinated and are unsafe, since they could cause a collision. Four co-ordinated actions are safe; in this case, the action of one ship may, furthermore, facilitate action of the other one.

This work does not take the vessels' status into account: in practice, either the give-way or the stand-on vessel may take action first. Therefore, eight co-ordinated actions are possible :

- Give-way vessel takes action first and turns to port; the stand-on vessel turns to port too or keeps her course and speed.
- Give-way vessel takes action first and turns to starboard; the stand-on vessel may turn to starboard too or keep her course and speed.
- Stand-on vessel takes action first and turns to port; the give-way vessel turns to port too or keeps her course and speed.
- Stand-on vessel takes action first and turns to starboard; the give-way vessel turns to starboard too or keeps her course and speed.

These different combinations of actions relate to different kinds of strategies, which are more or less close to the formal rules (i.e. the collision regulations).

In our work, we address three questions:

- Are different kinds of ships (namely cargo ships and ferries) likely to choose different kinds of strategies?
- To what extent do these strategies differ from the formal rules?
- What are the determinants of those strategies?

3. METHOD

Quantitative data concerning collision avoidance was collected at the Gris-Nez Vessel Traffic System. This centre watches over the traffic in the Dover Strait. Traffic in this area is heavy; it consists of cargo ships operating in the Traffic Separation Scheme and of ferries crossing between England and France, so that four interaction situations are possible (Cf. Figure 1). Traffic in this area was observed for one month, and 63 interaction situations between cargo ships and ferries crossing between Dover and Calais were recorded.

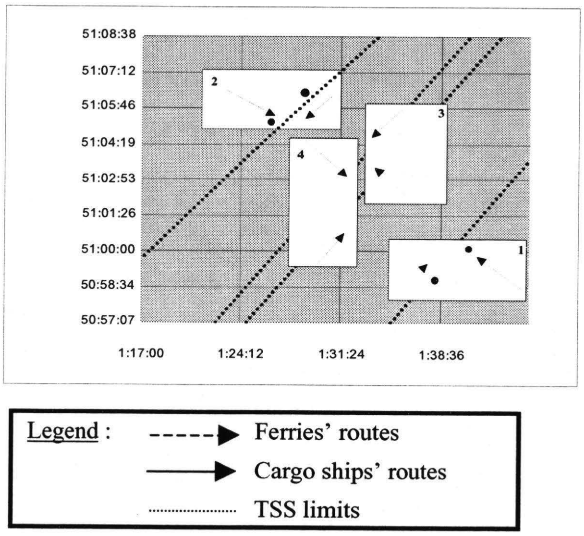


Figure 1 –

The different kinds of interaction situations analysed

For each of these situations, priority was defined (the ferry being either the stand-on vessel or the give-way vessel) and features of the context were described in terms of visibility, wind force and direction, strength and direction of the current; every 36 seconds for about 15 minutes the following values were noted:

- speed and course of each vessel,
- distance between vessels,
- DCPA (the Distance at Closest Point of Approach) and TCPA (Time to Closest Point of Approach).

Quantitative data were analysed using several statistical methods, namely: tests of correlation between the main variables, linear progression, analysis in principal components and discriminating analysis.

4. RESULTS

Statistical analysis allows [8]:

- the informal rules applied by ferries crossing the Dover Strait and by cargo ships operating in the Traffic Separation Scheme to be distinguished and,
- those rules to be compared with the formal rules (namely, the Collision Avoidance Rules).

The formal rules describe the behaviour of the 'give-way vessel' but take also into account the behaviour of the 'stand-on' vessel.

4.1 The behaviour of the 'give-way' vessel

It is written in the Collision Avoidance Rules that :

- *When two power-driven vessels are crossing so as to involve risk of collision, the vessel which has the other one on her own starboard side shall keep out of the way and shall, if the circumstances of the case admit, avoid crossing ahead of the other vessel* (Rule 15).
- *Every vessel which is directed to keep out of the way of another vessel shall, so far as possible, take early and substantial action to keep well clear* (Rule 16).

When the cargo ship is the give-way vessel, we note that she does not always perform a manoeuvre. Decision tree techniques show that the decision to perform an action depends on the speed of the ship.

- In fact, more than half of the slow give-way cargo ships (speed less than 10-11 knots) do not perform a manoeuvre.
- The behaviour of the faster give-way cargo ships overlaps the formal rules. They alter their course to the right, with an amplitude of around 26°, at a distance of around 2.6 nautical miles of the target, to achieve a miss distance of 0.7 mile.

Concerning the give-way ferries observed in a situation of interaction with cargo ships, we note that:

- Ferries alter their course to the right (64.5%) or to the left (23.5%), with an amplitude of 18°, at a distance of around 3.5 nautical miles to cross astern of the cargo ship at a distance of 0.7 nautical miles or ahead at a distance of 1 nautical mile.

Alteration of course to port is inconsistent with the formal rules, since it leads to crossing ahead of the stand-on vessel. However, it is not random behaviour. In fact, decision tree techniques show that it is connected with the vessel speed (ferries' speed is often higher than cargo ships' speed) and with a positive bow centre range (the ferry will cross ahead of the cargo ship if no manoeuvre is carried out). It obeys an economic strategy, since alteration of course to port is less important than an alteration of course to starboard.

4.2 Behaviour of the 'stand-on' vessel

The 'stand-on' vessel *may take action to avoid a collision by her manoeuvre alone, as soon as it becomes apparent to her that the vessel required to keep out of the way is not taking appropriate action in accordance with these rules* (Rule 17).

We observed that 13 of the 29 stand-on ferries altered their course. For seven of these cases, the action was carried out at a short distance (less than 2.6 nautical miles, the average distance at which the cargo ships habitually manoeuvre) and consists generally in altering course to starboard, even if this action generates a decrease in the DCPA and forces, therefore, the 'give-way' vessel to take co-ordinated action.

In the other cases, the manoeuvre is carried out very early and stops the other from taking action.

These actions are two possible interpretations of rule 17 [9]. In fact, rule 17 defines four stages relating to the permitted or required action for each vessel (Cf. Figure 2):

1. at long range, both vessels are free to take any action;
2. when a risk of collision first begins to apply, the give-way vessel is required to take proper action to achieve a safe passing distance and the stand-on vessel must keep her course and speed;
3. when it becomes apparent that the give-way vessel is not taking appropriate action, the stand-on vessel is required to give the whistle signal and is permitted to take action to avoid collision by her manoeuvre alone;
4. when collision cannot be avoided by the give-way vessel alone, the stand-on vessel is required to take such action.

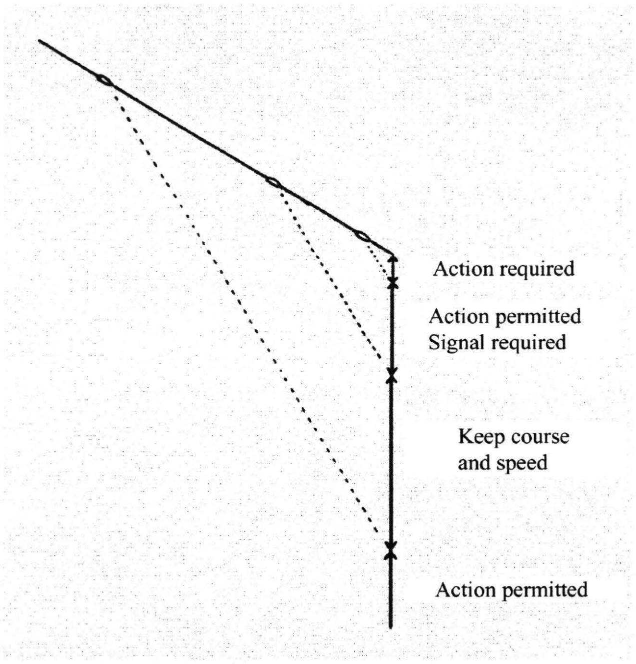


Figure 2 - The four stages relating to the permitted or required action for the 'stand-on' vessel

The interpretation of the rule relates to the distances at which the various stages begin. In the observed situations, the outer limits of the second stage seem to be of the order of 3 miles and the outer limits of the third stage seem to be of the order of 2 miles.

As a conclusion, it is possible to say that the observed behaviour of ships may disagree with rule 15, either because the give-way vessel cargo ship does not take action, or because the give-way ferry alters her course to port. They consist also in different interpretations of rule 16, because actions taken by ferries are more anticipatory than those by cargo ships. Behaviour of 'stand-on' ferries consists also in different applications of rule 17, each of them aiming at mastering the situation.

These patterns of behaviour are not random; they depend on the context (speed ratio between ships, positive or negative bow centre range) and they follow informal rules which are shared among certain groups of ships (ferries, fast cargo ships, slow cargo ships). They represent two general strategies: an individual one aiming at reducing course alterations and a more general one aiming at mastering the interaction situations.

However, these patterns of behaviour are not completely predictable, since two officers operating on the same kind of ships and in the same context may prefer to follow either the informal or the formal rules.

5. DISCUSSION

The results raise two questions. First of all, it is important to know if the combination of formal and informal rules is efficient to prevent collisions and incidents in crossing situations. If this is not the case, it seems interesting to use these results in order to assess a new support system, which is in the process of being implemented on board merchant vessels.

5.1 Risk management

Risk of collision depends on the areas and on the traffic density; collisions are much more likely to happen in coastal waters and in areas of restricted navigation [10]. Statistics published by the MAIB[1] [11] (concerning accidents involving merchant UK-flagged vessels or merchant foreign-flagged vessels in UK waters) show that collisions represent 11% of the accidents recorded between 1994 and 2001, the same as fire or explosion and more than grounding (8%). Nevertheless, in a given area such as the Dover Strait, collisions are too rare to constitute a data-base for an in-depth analysis. In order to answer the first question - concerning the efficiency of the rules actually followed by OOW – it seems, therefore, necessary to examine the incidents observed in this area. This examination will deal with: *i)* the efficiency of the interpretation of rule 17 made by OOW on board ferries, *ii)* the risk related to 'give-way' cargo ships that do not take action, *iii)* the incidents caused by ferries when OOW choose to alter course to port.

5.1.1 Efficiency of the interpretation of rule 17 on board ferries

Belcher [12] points out the great number of near miss encounters occurring in the Dover Strait. He analysed all ship movements within the south west bound traffic lane and all those crossing the traffic separation scheme in a 24-hour period and found 175 near miss encounters[2] (for 255 vessels involved in the observed movements) and 41 very close near misses[3]. But these very close near misses particularly concern overtaking situations and very few crossing situations (only 6%). The author explains that this result is due to the early action taken by the OOW on the ferry for vessels crossing from their own port side and confirm the efficiency of the current interpretation of rule 17.

In most cases, OOW on board ferries try to master the interaction by manoeuvring very early, when they are on board the 'give-way' vessel but also when they are on the 'stand-on' vessel. This general strategy is efficient to avoid accidents as well as incidents, whatever the behaviour of the cargo ships. It is widely shared among the ferries operating in the Dover Strait.

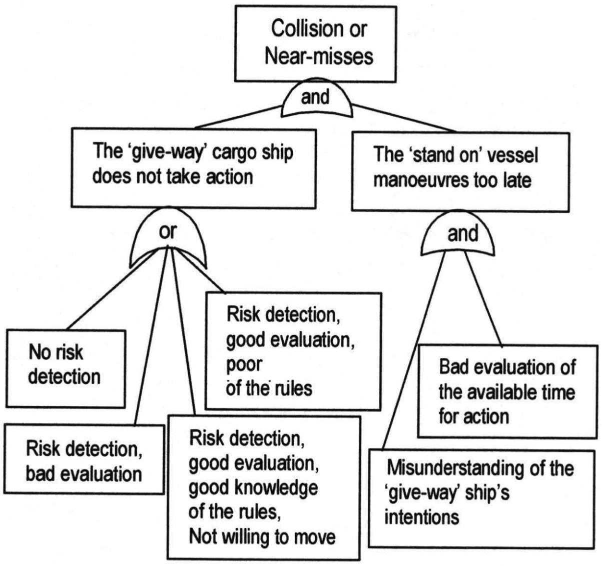
5.1.2 Risk related to the 'give-way' cargo ships which do not comply with rule 15

Most of the crossing situations involve a cargo ship and a ferry and are managed in this manner. However, crossing situations may be more complicated when they involve more than two ships. They may be also unusual, when they involve several cargo ships (one cargo ship operating in the TSS and one or two cargo ships crossing the Dover Strait). In such cases, incidents may occur due to lack of knowledge of the area's habits and, consequently, to lack of stored predefined schemata of actions and reactions. An incident or an accident may occur *i)* when the give-way vessel does not obey formal rule (namely rule 15) and does not take effective avoiding action and *ii)* when the stand-on vessel alters her course or reduces her speed and applies rule 17 far too late. In this case, the behaviour of the two OOW is uncoordinated. Most of the near misses observed in the Dover Strait and reported by the MAIB these last years are results of such situations. It is the case, for example, of the following incident [13].

The 6,391gt reefer vessel, Saratau, was proceeding in the south-west bound lane of the Dover Strait TSS on a course of 227°. Another reefer vessel, the 4,574gt Polestar, was in the opposite lane and heading north-east, but bound for the pilot station off Dover. To achieve this she made her heading 350° to cross the TSS. It was not an uncommon situation. Saratau first detected Polestar at a distance of 6 miles, and determined that a risk of collision existed. As the stand-on vessel in accordance with Rule 17 she maintained her course and speed. She was watching Polestar carefully and expected her to take avoiding action. By the time the distance between the vessels had reduced to approximately 1 mile, the bridge team on board Saratau had become very concerned that the other vessel appeared to be doing nothing to given way. She tried, first, to attract the other vessel's attention by using sound signals in accordance with Rule 34(d), and then by VHF radio, channel 16. As the distance between the vessels continued to close, Saratau altered course to port. Polestar, the give-way vessel, eventually reduced speed and then stopped her engines. The vessels passed each other at a distance of 1 cable. Polestar passed ahead of Saratau.

Causes of such situations are manifold (Cf. Figure 3). The first cause is non compliance with the formal rule. It concerns the give-way vessel using the Traffic Separation Scheme. It may be due to *i)* a poor look-out, *ii)* a lack of appreciation of the risk, *iii)* poor knowledge of rule 10 applying to Traffic Separation Schemes (OOW on board the cargo ship may believe that the crossing rule does not apply in narrow channels), *iv)* a reluctance to take an action that may disturb the traffic flow pattern. The second cause relates to late action of the stand-on vessel, which is due itself to a lack of appreciation of the available time for action or/and to the difficulty to foresee the action of the give-way vessel or to know her intentions.

Figure 3 – Failure tree



5.1.3 Risk related to ferries which alter course to port

Another kind of incident reported by the MAIB is related to ferries altering their course to port. In that case, the risk is due

---

[1] Marine Accident Investigation Branch of U.K.

[2] A near miss encounter is determined when two ships pass within 8 cables of each other (0.8 nautical mile).

[3] A very close near miss encounter is determined when two ships pass within 3 cables (0.3 nautical mile) or less of each other, leaving no room for error.

to bad evaluation of the speed ratio between the 'give-way' ferry and the 'stand-on' cargo ship. It may make the achievement of a sufficient miss distance difficult and therefore forces the 'stand-on' cargo ship to 'help' the ferry by altering also her course to port.

Several means could be used to improve the present system and avoid such near-misses.

To avoid risk related to 'give-way' ferries which alter course to port and cannot achieve a sufficient miss distance, a new decision assistance tool on board ships would be useful. In fact, only a computer tool with adequate calculation power, is capable of processing fuzzy and multiple data such as wind force, sea state, targets' speed and DCPA of several targets in order to assess or to suggest avoidance routes.

To avoid risk related to 'give-way' cargo vessels which do not take action, two major ways of improvement would be *i)* to implement a shore system to manage the traffic in busy waterways, *ii)* to improve interactions between OOW through training or with a new communication system. Such a system is now in process of being implemented on board merchant ships, thanks to the Automatic Identification System (AIS).

5.2 Toward a new aiding tool: the AIS

Interaction situations at sea are situations where cognition could be distributed among several actors, in order to reduce uncertainty concerning intentions and actions of targets vessels. In fact, they are not. Our study shows that OOW on board ferries prefer to master the situation by avoiding the interaction or by constraining actions or reactions of the targets.

Distributed cognition would require communication between actors and the knowledge of their intention. Nowadays, OOW are reluctant to make contact by VHF radio, due to the difficulty of distinguishing an approaching vessel from other vessels in the vicinity and the added problem of communicating with a crew of a different nationality.

The Automatic Identification System (AIS) is now required for all ships of 300 gross tonnage and upwards engaged on international voyages and cargo ships of 500 gt and upwards not engaged on international voyages and passenger ships irrespective of size. Certain vessels are already fitted with AIS (ships engaged on international voyages and ships constructed on or after 1 July 2002) and others are to be fitted (not later than 1 July 2003 for passenger ships). AIS broadcasts the ship's identity, position and other data locally at regular intervals. Among other features, the aim of AIS is to make verbal communication easier, thanks to identification of targets.

The question is to know if this feature could avoid incidents and accidents.

AIS may improve the knowledge of other's intentions. Nevertheless, verbal contacts are likely to introduce a certain complexity into the communication process, which may be time consuming and may delay actions and reactions [14]. Nielsen and Petersen [15] point out that spoken communication may not be the ideal medium for all types of information exchange in collision avoidance, due to issues of lack of persistency, format differences between instruments and 'speak able' information, labour intensiveness etc. To avoid misunderstanding, a solution could be to exchange written and predefined messages rather than verbal ones. To avoid time consuming negotiation, another solution could be to rely on the traffic control centre which could collect and process the intentions of those involved in order to make proper decisions.

In addition, AIS cannot improve look-out and risk assessment on board the give-way vessel unless it is used to send alarms to the target. It cannot have any effect on lack of knowledge. It cannot improve the appreciation of the available time for action on board the stand-on ship.

6. CONCLUSION

This study pinpoints the interpretations of the Collision Regulations, which govern collision avoidance in a busy waterway. It shows some deviations from the rules, particularly in the case of 'give-way' cargo ships which do not take action. It shows also that actions and reactions of ferries are generally convenient; carried out very early, they allow the interaction to be mastered and to avoid any incident.

Nevertheless incidents may occur, particularly with other kinds of vessels crossing the Dover Strait less often than ferries and which do not have suitable routines of behaviour. For such cases, the new support system which is being implemented on large cargo ships does not seem to be able to cover all the problem's aspects. It should make the verbal communications easier. It could contribute to a real improvement in sharing information among those concerned. On the other hand, it could complicate communications, it could also be time consuming and – moreover – will not be capable of dealing with certain causes of the incidents observed in crossing situations (causes which be may be related to knowledge of the regulations and to the look-out for example). It seems, therefore, that such a system should be seen and designed as part of a whole system taking into account not only the superficial aspects of a given problem but also its roots.

7. REFERENCES

[1] HMSO (1972). *International Regulations for Preventing Collisions at Sea*. HMSO.

[2] Hara, K. (1993). A safe way of collision avoidance manoeuver based on manoeuvring standard using fuzzy reasoning model. Paper presented at the *MARSIM'93 - International Conference on Marine Simulation and Ship Manoeuvrability*. St. John's, Newfoundland, Canada, Sept.

[3] Habberley, J.S., & Taylor, D.H. (1989). Simulated collision avoidance manoeuvres: a parametric study. *The Journal of Navigation, 42-2*, 248-254.

[4] Hinsch, W. (1996). Traffic rules to coordinate collision avoidance manoeuvres at sea. In J. Zhao, P.A. Wilson, Z. Hu, & F. Wang (Eds.), Proceedings of *International Conference on Preventing Collision at Sea – Collision'96 (pp. 166-172)*. Dalian: Publishing House of Dalian Maritime University.

[5] Chauvin, C. (1999). From analysis of watch officer' activity to ergonomic assessment of Automatic Radar Plotting Aid. In J.M. Hoc, P. Millot, E. Hollnagel, & P.C. Cacciabue (Eds.), Proceedings of the *Seventh European Conference on Cognitive Science Approaches to Process Control (pp. 29-34)*. Valenciennes : Presses Universitaires de Valenciennes.

[6] Wilde, G.-J.S. (1976). Social interaction patterns in driver behavior: An introductory review. *Human Factors, 18-5*, 477-492.

[7] Zhao, J., Price, W.-G., Wilson, P.-A., Tan, M. (1995). The uncertainty and uncoordination of mariners' behaviour in collision avoidance at sea. *The Journal of Navigation*, 48/3, 425-435.

[8] Andro, M., Chauvin, C., Le Bouar, G. (2002*). Les interactions entre navires*. Etude réalisée pour la Direction Scientifique et Technique du Ministère des Transports, de l'Equipement et du Logement, du Tourisme et de la Mer, 56 p.

[9] Cockroft, A.N. & Lameijer, J.N.F. (1996). *A guide to the collision avoidance rules* – 5th ed. Oxford: Butterworth-Heinemann Ltd.

[10] Pilley, R.J. (1996). Collisions: Some facts and figures. In J. Zhao, P.A. Wilson, Z. Hu, & F. Wang (Eds.), Proceedings of *International Conference on Preventing Collision at Sea – Collision'96 (pp. 19-25)*. Dalian: Publishing House of Dalian Maritime University

[11] MAIB (2001). Annual Report. http://www.maib.dft.gov.uk/ar2001/index.htm

[12] Belcher, P. (2003). A day in the life of the Dover Strait. *Safety at sea international. 57/408*, pp. 15-16.

[13] MAIB (2002). Safety digest, 2/2002. http://www.maib.dft.gov.uk/sd/0202/10.htm

[14] Chauvin, C., Saad, F. (2000). Communication and interaction in dynamic control tasks. Paper presented at the *International Conference on Traffic and Transport Psychology ICTTP*. Bern, Switzerland, September 4-7, 2000.

[15] Nielsen, M., Petersen, J. (2001). Collision avoidance at sea – Practice and Problems. In M. Lind (Ed.), *Proceedings of the 20th European Annual Conference on Human Decision Making and Manual Control (p. 81-90)*. Kongens Lyngby, Denmark, june 25-27.

Healthcare Staff Attitudes towards Management, Job, Teamwork and Leadership in Japanese Hospitals

Kenji Itoh
Tokyo Institute of Technology
Tokyo, Japan
ken@ie.me.titech.ac.jp

Toshiko Abe
Tokyo Medical and Dental University
Tokyo, Japan
abet.ns@tmd.ac.jp

Henning Boje Andersen
Risø National Laboratory
Roskilde, Denmark
henning.b.andersen@risoe.dk

ABSTRACT

The present paper reports the results of a questionnaire-based survey of safety culture in hospitals including healthcare staff's attitudes towards and perceptions of hospital management, work goals, leadership and teamwork. Approximately 600 responses were collected from physicians, nurses and pharmacists working in five Japanese hospitals. The questionnaire was adapted from Helmreich's "Operating Team Resource Management Survey" and contained, in addition, questions about respondents' reporting of their own errors and information to patients who have suffered adverse events. This paper describes results of the survey that relate to healthcare staff attitudes towards safety-related issues including comparisons between departments and wards as well as work positions. In addition, we compare the attitudes of healthcare staff with those of ship officers that have been elicited using a similar type of questionnaire. Based on the survey results, we discuss professional culture of Japanese healthcare systems that are closely relating to patient safety.

Keywords

Safety culture, Patient safety, Medical staff attitudes, Adverse events, and Questionnaire-based survey

1. INTRODUCTION

It is widely recognised that human error is the predominant cause of accidents not only in human-machine systems involved in, mainly high-tech, transport and industrial domains such as aviation, railway, ship handling and nuclear power production but also in healthcare systems and in particular in hospitals [12]. In recent years, there has been a much-heightened focus on the impact of organisational factors on safety [13]. Similarly, it has been pointed out that the dominant type of contributing causes of major accidents involve the organisations that themselves shape the safety culture or climate within which their employees operate [5, 14]. The concept of safety culture has been defined in various ways by researchers. A frequently cited definition was provided by the UK Advisory Committee on the Safety of Nuclear Installations which defined safety culture as "the product of individual and group values, attitudes, perceptions, competencies and patterns of behaviour that determine the commitment to and the style and proficiency of an organisation's health and safety management" [1]. In other words, the concept is coupled not only to management's commitment to safety, its communication style and the overt rules for reporting errors but also to employees' motivation, morale, attitude to management and their perception of errors and performance shaping factors [2].

At the same time, in recent years a number of projects have sought to uncover the safety culture of individual organisation in the above types of high-tech industries. In these projects, for example, operators' safety culture related attitudes have been found to correlate with incident/accident rates in railway operations and to be important indices of safety performance [9, 10, 11]. In the healthcare domain, a recent study found that a number of aspects relating to safety culture – such as acknowledgement of human error and power distance – were correlated with the rate of incident reporting of individual work units [8], and it was suggested that they might in turn impact on patient safety.

In the present study, we performed a questionnaire-based survey to identify characteristics of safety culture in Japanese hospitals. In our former article [8], we described findings about healthcare staff attitudes towards incident and error reporting including their actions vis-à-vis the individual patient who has been injured by medical error. The findings were based on the analysis of responses to questions relating to two fictitious adverse events that are a part of the questionnaire applied in the present study. In this paper, however, detailed results of other parts in the survey are reported, particularly focusing on healthcare staff perceptions of and attitudes towards hospital management, job, leadership and teamwork. As part of the safety cultural structure in healthcare systems, we identify differences and similarities in staff attitudes between departments/wards, positions and organisations as well as between physicians, nurses and pharmacists. In addition, the questionnaire responses from physicians and nurses are compared with those of ship officers that have been collected in our previous studies using a similar type of questionnaire [3, 9]. Based on the integrated results of the questionnaire survey, we discuss some current issues of safety culture in Japanese hospitals as well as factors that jeopardise patient safety.

2. QUESTIONNAIRE

The questionnaire applied in this study comprises five parts and has an additional demographic section where respondents fill in their department or ward specialty, position, experience and age group. Four of the five parts of the questionnaire have been adapted from Helmreich's "Operating Team Resource Management Survey" [6]. The Helmreich questionnaire has several derivatives focusing on specific domains and allows us to compare the results with ones derived from other domains, e.g., maritime operations and aviation [3, 6, 9]. One of the greatest advantages of using the adapted questionnaire is the opportunity it provides for comparing professional culture across domains. We have transformed terms and statements from the original "Operating Team Resource Management Questionnaire" to fit the working situation of physicians, nurses and pharmacists working not only in the operating room but also in other types of departments and wards, keeping the same meaning and intention for each question. Finally, the questionnaire has been translated into Japanese.

Part I of the questionnaire contains 57 question items about perceptions of hospital management as well as general questions about factors or attitudes that may impact on safety performance. Respondents are asked to rate each item on a five-point Likert scale between 1 and 5 (from ‘strongly disagree’ to ‘strongly agree’). The question items can be classified into distinct groups in terms of organisational and human aspects that form hospital safety culture. In the present study, with reference to the original classification by Helmreich and Merritt [6], we arranged all the items into nine categories of distinct “safety culture aspects”: (1) power distance, (2) communication, (3) teamwork, (4) recognition of own performance degradation under high stress or workload, (5) stress management for team members, (6) morale and motivation, (7) satisfaction with management, (8) recognition of human error potential, and (9) awareness of own competence. Each category includes several items. For example, the category, power distance comprises twelve items among which the following examples illustrate the format and style of the questions: “The senior person should take over and make all the decisions in life-threatening emergencies”; “senior staff deserve extra benefits and privileges”; and “physicians who encourage suggestions from team members are weak leaders.”

The second part of the questionnaire was developed for Danish survey of physicians’ and nurses’ attitudes [4], in which respondents are asked about their behaviour and actions in terms of reporting or talking with their leaders and colleagues their own errors as well as their information to patients who have become victims of such errors. Respondents’ reactions are elicited as responses to two fictitious adverse events. The respondent is asked to study each case vignette and subsequently to rate his or her likelihood of engaging in various actions described in the questionnaire on a five point Likert scale.

Part III asks respondents about their perception and preference of leadership styles, offering descriptions of four different styles of leadership varying from an autocratic to a democratic type. For example, a sample description of the most autocratic style is: “A leader usually makes decisions promptly and communicates them to subordinates clearly and firmly. He or she expects them to carry out the decisions loyally without raising difficulties”. Respondents are then asked two questions: (1) which style you most prefer to work under, and (2) which style you find yourself actually working under most often in your organisation.

In Part IV, 15 questions about work goals are included involving issues such as security of employment, changing work routines with new, unfamiliar tasks, and working with people who cooperate well with one another. Like the other parts of the questionnaire, respondents are asked to rate each item on a five-point Likert scale (from ‘of very little or no importance’ to ‘of utmost importance’), considering his/her ideal job. In the last part, respondents describe their personal perception of the quality of teamwork and cooperation with different professional members – working in several specific departments such as internal medicine, surgery and anaesthesiology – at various positions.

The questionnaire was distributed to physicians, nurses and pharmacists working in five hospitals located in different areas in Japan. The survey was made between December of 2001 and January of 2002. A total of 66, 486 and 43 responses were obtained from physicians, nurses and pharmacists, respectively. The mean response rate was 91% across the three professional groups. Among physicians, 33 respondents were heads of department, 22 consultants or physicians after residents, and 9 residents. In the nurse group, responses were collected from 32 matrons and 97 deputy leaders while 354 were from ordinary nurses. In the pharmacist group, samples came from two leaders, 11 deputy leaders and 30 from ordinary staff.

3. STAFF ATTITUDES IN HOSPITAL

3.1 Hospital Safety Culture

Using the questionnaire responses to Part I, percentage agreement and disagreement for each safety culture aspect mentioned in the last section are shown in Table 1 across the three professional groups. The percentage [dis]agreement is defined as the following rate: the nominator represents 5 and 4 responses, i.e., “strongly agree” and “slightly agree” [the 1 and 2 responses, i.e., “strongly disagree” and “slightly disagree”]; and the denominator represents the total number of responses for the specific items of each aspect. Before calculation of these indices, items that represent negative meaning in terms of the aspect had their figure reversed, i.e., 5 and 4 responses were reversed to 1 and 2, and vice versa. Finally, tests of significance (Kruskal-Wallis) were performed for each safety culture aspect as well as for each question item to identify significant differences between the professional groups.

Table 1 Percentage (dis)agreement of safety culture aspects

Safety culture aspects		Physician	Nurse	Pharma.	Total	X_0^2
I. Power distance	% agree.:	30.4%	21.8%	27.6%	23.2%	0.88
	%disagree.:	59.7%	60.4%	59.2%	60.3%	
II. Communication		88.1%	85.9%	89.4%	86.4%	14.75**
		4.9%	3.8%	2.9%	3.9%	
III. Teamwork		57.6%	65.0%	55.2%	63.5%	16.17**
		26.0%	15.7%	24.8%	17.5%	
IV. Own performance under high stress		49.2%	41.0%	42.6%	42.0%	3.92
		38.1%	35.7%	32.9%	35.8%	
V. Stress management for team member		69.5%	69.4%	66.8%	69.2%	5.12
		19.8%	15.8%	21.6%	16.6%	
VI. Morale & motivation		72.9%	65.7%	65.9%	66.5%	14.75**
		16.0%	15.1%	18.5%	15.4%	
VII. Satisfaction with management		45.5%	51.3%	51.7%	50.7%	10.40**
		39.6%	28.8%	31.7%	30.1%	
VIII. Recognition of human error		60.6%	60.7%	55.4%	60.3%	2.32
		26.3%	21.3%	28.6%	22.4%	
IX. Awareness of own competence		58.2%	44.8%	40.2%	46.0%	17.52**
		27.1%	24.8%	30.9%	25.5%	

**: p<0.01, *: p<0.05

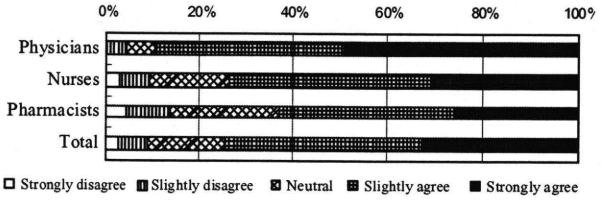


Figure 1 Responses of the item "I like my job"

The overall trend of results from the five hospitals surveyed in this study is that the healthcare staff indicates a relatively high morale and motivation as well as relatively positive perception of communication within their organisations. One of the most typical items representing motivation is “I like my job”, and responses from the three groups to this item are shown in Figure 1. As can be seen from this figure, physicians in particular have a high level of motivation, compared to nurses and pharmacists. Regarding the awareness of healthcare staff of own competence, responses to this aspect vary among three professional groups, and, similar to the results about motivation, physicians show a greater degree of awareness of their own competence compared to nurses and pharmacists. The respondents also perceive teamwork within their work group at a reasonably high level. In particular, nurses’ perception of teamwork was the most positive and approximately two thirds of nurses agreed that they have good teamwork in their respective hospitals. Compared with these safety culture aspects, satisfaction with management is not high, and the physician’s satisfaction was significantly lower than that of the two other professional groups.

One of the most interesting safety culture aspects is *power distance*, which refers to the psychological distance between leaders or superiors and subordinate members. A small power distance reflects, for example, that leaders and their subordinates have open communication initiated not only from leaders but also, more critically, from juniors. The results of the survey suggest that there is a relatively small power distance in Japanese hospitals; in addition, no significant difference in perception of this aspect was uncovered between physicians, nurses and pharmacists. For example, 91% of physicians and 90% of nurses disagreed (?2 =2.56; p>0.05) with the item “team member should not question the decision or actions of senior staff except when they threaten the safety of the medical activity”. A similar pattern was obtained for the item “senior staff should encourage questions from junior staff during their activities if appropriate”, but for this item physicians’ agreement was significantly higher than that of the nurses (?2 =13.72; p<0.01), as shown in Figure 2.

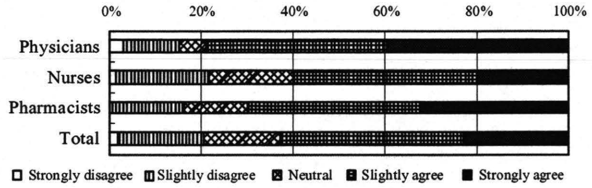


Figure 2 Responses of the item "Senior staff should encourage questions from junior members"

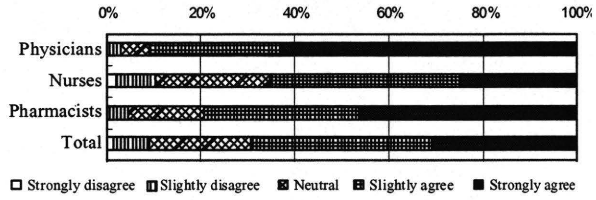


Figure 3 Responses of the item "Human error is inevitable"

A large part of the healthcare staff surveyed showed positive attitudes to and a realistic recognition of human error. As a representative question of this aspect, Figure 3 depicts responses to the item “human error is inevitable”. As can be seen from this figure, most respondents agreed with this item (91% and 65% of agreement for physicians and nurses, respectively; ?2 =41.65, *p*<0.01). They disagreed with the statement that “errors are a sign of incompetence” (80% and 71% disagreement; ?2 =1.92, *p*>0.05). However, for the item regarding error reporting, “I am encouraged by my leaders and co-workers to report any incidents that I may observe” a largely positive response was observed, but there was a quite large and significant difference in responses to this item between the three professional groups (?2 =79.13, *p*<0.01). More than 85% of nurses agreed with this question while the percentage agreement of physicians was less than 45%.

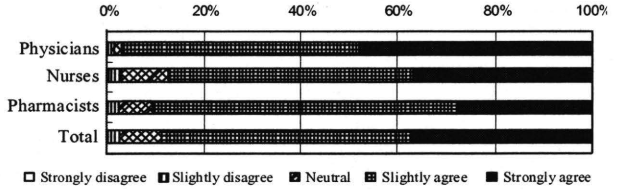


Figure 4 Responses of the item "Team members should monitor each other for signs of stress or fatigue"

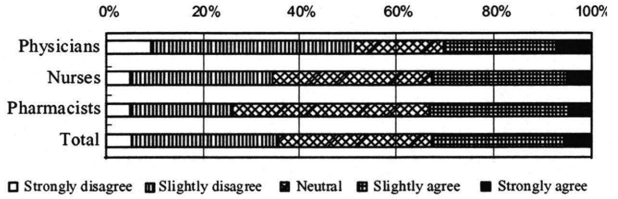


Figure 5 Responses of the item "I am more likely to make errors or mistakes in tense or hostile situations"

Regarding attitudes to stress management for team members, most of healthcare staff recognised the need for monitoring colleagues’ levels of stress and workload. For example, more than 90% of respondents agree with the item, “team members should be monitored for signs of stress and fatigue during task” (see Figure 4). In contrast, respondents did not exhibit any great awareness of the effects of stress on their own performance. More than half of the doctors and one third of the nurses disagreed with the item, “I am more likely to make errors or mistakes in tense or hostile situations”, as indicated in Figure 5. Similarly, only 5% of doctors agreed that their performance is reduced in a stressed or fatigued situation (89% disagreement). Percentage disagreement of this item was slightly lower at 78% for nurses.

3.2 Differences between Departments/Wards

Analysis of similarities and differences within the physician group between departments – or rather, specialties, i.e., internal medicine (N=19), surgery (N=22) and others (N=25) – showed no significant differences (with one exception), although the lack of significant differences may be due to the relatively small response samples (type 2 error). The only aspect that turned out to show a significant difference between specialties was awareness of own competence. Thus, the agreement of surgeons was higher by more than 10 percentage points than those of the other two specialty groups.

For nurses, percentage agreements and disagreements of each ward group are shown in Table 2 for all the safety culture aspects, being classified into eight groups: internal medicine (N=129), surgery (138), ICU (intensive care unit; N=39), outpatient (N=55), paediatrics (N=12), mixed ward (N=52), and operating room (N=30). Unlike the department-based analysis of physicians mentioned above, there were significant differences between the nurse’s ward groups in several safety culture aspects: communication, stress management for team member, morale and motivation, and recognition of human error. Among the eight groups, two stood out as remarkable types in terms of responses to these aspects. One type comprises nurses working in the operating room and paediatrics.

Table 2 Ward- and position-based comparisons of nurses in safety culture aspects

Safety culture aspects		Ward-based								Position-based			
		Internal medicine	Surgery	ICU	Out-patient	Paediatrix	Mixed ward	OR	f Ȭ	Ordinary	Chief	Matron	f Ȭ
I. Power distance	% agree.: %disagree.:	20.0% 61.8%	22.8% 59.1%	24.6% 57.1%	23.7% 61.0%	24.3% 64.6%	20.2% 59.9%	20.7% 62.8%	13.30	22.7% 58.7%	19.5% 64.7%	17.9% 68.0%	16.92**
II. Communication		87.4% 1.7%	83.9% 5.9%	83.3% 2.1%	88.1% 4.5%	88.1% 3.4%	84.6% 3.9%	90.7% 3.3%	17.50*	84.1% 4.5%	90.3% 1.7%	94.2% 1.3%	15.66*
III. Teamwork		66.0% 14.2%	64.4% 16.3%	60.9% 13.9%	66.2% 19.9%	65.2% 21.7%	67.3% 14.1%	68.1% 15.1%	4.61	66.4% 14.5%	62.2% 17.8%	59.2% 22.4%	5.35
IV. Own performance under high stress		40.4% 35.6%	41.2% 36.8%	43.0% 31.1%	36.0% 42.3%	47.7% 34.6%	41.7% 29.8%	50.4% 32.6%	21.30	42.6% 33.6%	37.4% 38.8%	36.5% 47.4%	16.84**
V. Stress management for team member		71.5% 14.3%	68.8% 17.4%	71.1% 10.5%	71.2% 15.0%	69.5% 15.3%	68.0% 15.8%	65.8% 19.5%	3.68	67.4% 16.5%	73.9% 13.4%	77.8% 13.3%	11.26**
VI. Morale & motivation		70.3% 12.6%	61.7% 17.8%	62.2% 14.5%	76.9% 10.6%	63.2% 26.3%	59.6% 13.6%	61.5% 17.6%	35.40**	61.0% 17.3%	76.4% 9.8%	85.3% 6.4%	74.96**
VII. Satisfaction with management		55.0% 26.5%	49.6% 27.7%	47.4% 32.6%	56.8% 32.1%	52.5% 33.9%	50.2% 26.7%	48.3% 34.2%	10.27	48.1% 30.1%	58.1% 26.8%	67.3% 16.4%	31.96**
VIII. Recognition of human error		64.9% 19.5%	58.7% 21.9%	53.3% 23.7%	61.1% 23.1%	60.9% 26.1%	60.9% 17.8%	67.5% 20.8%	15.42*	59.7% 20.9%	63.5% 22.3%	64.2% 22.8%	6.37*
IX. Awareness of own competence		48.9% 23.0%	42.8% 25.5%	45.6% 20.4%	48.0% 26.3%	42.3% 32.4%	39.9% 20.3%	41.8% 30.5%	12.17	41.7% 26.7%	49.6% 21.0%	63.6% 14.4%	35.01**

**: $p<0.01$, *: $p<0.05$

Compared to the other ward groups, these ward groups of nurses expressed greater agreement with the importance of communication and they showed a higher level of realistic acknowledgment of their own performance limitations under stress conditions as well as a more realistic acknowledgement of human error, and, finally, a relatively lower level of morale and motivation. Nurses working in internal medicine and with outpatients compose the other ward type. In contrast to the operating room and paediatrics nurses, they had the highest morale and motivation and expressed greater agreements with the items about stress management for team members, but a lower level of appreciation of their own performance limits under stress condition. The latter results might in part reflect differences in tasks and work conditions.

3.3 Differences between Positions

Similar to the results of the department-based analysis, there were few significant differences across positions, i.e., residents (N=9), consultants (N=22) and leaders (N=33), in physicians' responses about safety culture aspects – but again, this result might well be due to the relatively small sample. Only morale and motivation (?2 =14.45, $p<0.01$), and awareness of own competence (?2 =11.72, $p<0.01$) were found to show significant differences between leaders, consultants and residents. Physicians in a leading position, i.e., heads of department, exhibited the highest morale and motivation and showed the greatest awareness of their own competence. No difference in morale and motivation was observed between consultants and residents, but consultants' awareness of own competence was slightly stronger than that of residents.

In contrast, as shown in Table 2, there were significant differences within the nurse group in responses to all the aspects between their positions, i.e., ordinary (N=354), chief (N=97) and matron (N=32). As an overall trend, nurses at a higher position showed higher morale and motivation and they exhibited more positive or realistic attitudes to management, error recognition and other organisational issues, but they also showed greater power distance. However, opposite patterns of perceptions of own performance under stress condition were indicated between the position groups, that is, ordinary nurses had the most realistic acknowledgment of their own performance limitations under stress conditions.

3.4 Differences between Hospitals

Using response data obtained from nurses – since only a small number of physician's responses were collected from three out of the five hospitals (N=91, 113, 88, 100, and 94) surveyed in this study – we performed hospital-based comparisons of the safety culture aspects. Significant differences were identified in most aspects between the five hospitals: power distance (?2 =13.28, $p<0.05$), communication (?2 =34.22, $p<0.01$), recognition of own performance degradation under high stress (?2 =33.88, $p<0.01$), morale and motivation (?2 =39.28, $p<0.01$), satisfaction with management (?2 =45.06, $p<0.01$) and recognition of human error potential (?2 =15.65, $p<0.01$). These differences may suggest that each hospital has a different style and procedures concerning risk management, error reporting, manuals and checklists, safety training and rules, etc., which shape its own local safety culture.

In our previous study, applying a similar type of questionnaire to railway operators [11], we have elicited responses to question items that can differentiate low-incident and high-incident work units – we call these items "risk-identifying items" – based on integrated results of questionnaire responses and accident and incident statistics, both of which were collected from the same railway operation company. Most of these question items fall into two safety culture aspects: on the one hand, morale and motivation, and on the other, recognition of own performance limitations under stress situation. Results of applying these question items to hospital-based responses of the nurse group are shown in Table 3 in terms of percentage agreement and disagreement as well as chi-square values calculated by the Kruskal-Wallis test. As can be seen in this table, there are significant differences between the five hospitals surveyed in this study for all the risk-identifiable questions except for the item, "I like my job" for which a significance level was at less than 10%. Again, the relatively small sample size may possibly make it more likely that a type-two error is made if we conclude that there is in fact no difference.

Table 3 Hospital-based percentage (dis)agreement of items that differentiated high-/low-incident railway organisations

Items		Hospitals A	B	C	D	E	f Ô
Even fatigued, I perform effectively	%agree.: %disagree.:	78.0% 8.8%	75.0% 14.3%	67.0% 15.9%	62.6% 27.3%	74.5% 4.3%	13.92**
I do my best work when I am alone		93.4% 2.2%	93.6% 0.9%	86.9% 4.8%	82.1% 4.2%	94.6% 0.0%	14.80**
My decision-making ability is as good in emergencies as in routine situations.		28.7% 28.7%	38.9% 24.8%	26.7% 40.7%	14.0% 54.0%	29.3% 21.7%	32.27**
Regular debriefing is an important part of maintaining effective coordination		72.5% 5.5%	80.9% 4.5%	90.8% 4.6%	91.9% 2.0%	91.3% 2.2%	15.18**
I am more likely to make errors in tense situations.		15.6% 46.7%	34.8% 39.3%	35.6% 26.4%	49.0% 21.4%	27.7% 37.2%	28.85**
My performance is not adversely affected with an inexperienced team member		42.2% 27.8%	29.5% 32.1%	32.6% 31.4%	27.6% 48.0%	14.0% 44.1%	17.54**
I am proud to work for this hospital		37.5% 33.0%	73.2% 8.0%	54.5% 14.8%	59.6% 14.1%	61.7% 16.0%	42.48**
A truly professional can leave personal problems during medical activity		62.2% 21.1%	68.1% 19.5%	47.7% 29.1%	49.0% 25.5%	62.8% 16.0%	11.51*
I like my job		74.4% 12.2%	79.6% 7.1%	79.3% 8.0%	62.6% 16.2%	71.0% 4.3%	8.24

Table 4 Leadership style in healthcare

(a) Style most preferred

	•©Autocratic 1	2	3	Democratic•¨ 4
Physicians	9.2%	49.2%	24.6%	16.9%
Nurses	5.5%	53.4%	10.1%	31.0%
Pharmacists	4.9%	61.0%	19.5%	14.6%
Ship officers	10.6%	58.6%	23.8%	6.9%

(b) Style most often found

	•©Autocratic 1	2	3	Democratic•¨ 4
Physicians	39.3%	27.9%	14.8%	18.0%
Nurses	33.0%	28.1%	18.6%	20.4%
Pharmacists	13.3%	43.3%	30.0%	13.3%
Ship officers	45.9%	30.4%	17.5%	6.2%

As an overall trend, nurses working in Hospitals B and E showed relatively high morale and motivation. For example, their percentage agreements of the items, "I am proud to work for this hospital", "I do my best work when people leave me alone" and "I like my job" are higher than nurses in the other hospitals. However, their recognition of effects of stress, fatigue and workload was relatively less realistic than nurses in Hospitals C and D, whose level of morale and motivation was lower than the other hospitals. Similar to the circumstances under which we collected data in the railway survey [11], at the time of the present survey, the healthcare staff involved had not received any training about effects of stress, fatigue, workload and other psychological factors on task performance and quality. In such a situation, these items relating to recognition of one's own performance degradation under high stress might often project the staff's morale. Thus, in the railway study it was found that percentage agreements on these stress-related items were higher – more realistic recognition of stress effects – for high-incident work units, i.e., branches within a company, whose rate of accidents/incidents was higher [11]. Considering the situation in the Japanese hospitals surveyed in this study, there seems to be the same relationship between responses to recognition of stress effects and the level of morale. In a future, when an appropriate training on these issues is provided to the healthcare staff, this relationship may change, i.e., disappear or change its order of correlation.

These results seem to indicate each hospital has its own safety culture, and therefore, if it is legitimate to generalise the relationship between safety culture response and incident risk found in our previous studies of railway, the hospitals may well have different levels of risk of medical adverse events (We are not suggesting, of course, that the level of risk of incident is determined solely by safety culture; only that safety culture is a co-determinant of risk of incidents).

4. LEADERSHIP, TEAMWORK AND WORK GOALS

4.1 Preferred vs. Actual Styles of Leadership

Respondents' perceptions of leadership issues are summarised in Table 4. This table includes responses from Japanese ship officers about the same leadership questions – the survey of ship officers will be briefly mentioned in the next section – for the purpose of professional safety culture comparisons. Desirable leadership may be affected not only by professional culture but also by national culture. Helmreich and Schaefer [7] reported that the "consultative" style – "A leader usually consults with subordinates before reaching decisions. He/she listens to their advice, considers it, and then announces decision. He/she expects all to work loyally to implement it whether or not it was in accordance with the advice they gave." – was supported by more than half of operating room staff in a German hospital. In contrast, in our survey, the style most preferred by the three professional groups in Japanese healthcare as well as by the Japanese ship officers was the "mildly autocratic" style that was described in this way: "A leader usually makes decisions promptly, but, before going ahead, tries to explain them fully to subordinates. He or she gives them the reason for the decisions and answers whatever question they may have". Still, approximately 30% of the nurse group preferred the "democratic" style followed by the "autocratic" style. This may well project a difference of professional culture between physicians and nurses.

In actual workplaces of healthcare, more autocratic styles are frequently found than what the staff desired both in Japan and in Germany – the autocratic style for more than 40% [7]. In Japanese hospitals, as can be seen in Table 4, physicians and nurses found the autocratic style most frequently followed by the mildly autocratic style. The pharmacist's perception of leadership is different from these two groups. They find that the mildly autocratic style is the most frequent, and, in addition, 30% of this group report that the consultative style is what they most often observe. It is noteworthy that both physicians and pharmacists prefer a style of leadership that is *less* "democratic"

Table 5 Perceived Teamwork with healthcare personnel

	Doctors			Nurses		
	Physicians	Surgeons	f Ô	Intern.med.	Surgery	f Ô
Internal medicine						
Leader	38.9%	42.1%	1.02	26.3%	•|	9.06
Consultants	55.6%	63.2%	2.15	27.1%	•|	10.44
Residents	41.2%	55.6%	3.82	30.8%	•|	14.82*
Nurses	44.4%	22.2%	5.16	63.7%	31.5%	34.39**
Surgery						
Leader	33.3%	80.0%	9.01*	23.4%	36.8%	14.11*
Consultants	50.0%	89.5%	10.06**	17.5%	33.0%	15.18*
Residents	28.6%	75.0%	6.36*	•|	34.1%	10.50
Nurses	6.7%	73.7%	13.93**	30.8%	51.5%	17.19*
Anaesthesiology						
Leader	21.4%	95.0%	20.12**	•|	•|	•|
Consultants	7.7%	81.3%	17.55**	•|	•|	•|
Nurses	•|	•|	0.58	•|	•|	•|

Figures: % agreement of good teamwork. (rate of "very good" + "good") or "consultative" than the ones they most often work under, though few of them prefer the outright autocratic style.

Different patterns of preference and perception were observed between specialties of the physician group. Surgeons preferred a more democratic leadership style, i.e., the consultative style (48%) while physicians supported the mildly autocratic style (53%). However, in their actual workplace, the surgeons (50%) found the autocratic style much more frequently than the physician group (33%). There is an almost identical pattern as the one mentioned above across the nurses' ward group.

4.2 Perceived Teamwork in Hospital

In Table 5 is shown respondents' perceptions of "very good" and "good" teamwork within their own groups and with other groups expressed by the physicians and nurses in internal medicine, surgery and anaesthesiology, respectively. We have calculated the percentage agreements of teamwork for each professional staff group for which more than 50% of response were collected from each department/ward group. In this table, a dashed mark ("–") indicates less than 50% responses from each professional group (lack of item response means having no teamwork experience of a respondent with a specific professional group). It is a common pattern of teamwork perception that respondents in each professional group and specialty/ward has the most positive attitudes to teamwork within their own group. For example, the nurse group working in the surgical ward has the most positive teamwork perception of their own group compared with other groups, and so does the internal medicine nurse group. Surgeons' perception is much more positive than the other groups on average. In addition, nurses' perception of teamwork with physicians is very low, even their teamwork with the physician who have the same specialty. On the other hand, physicians do not have the same relatively negative perception of nurses.

Besides these common patterns, there are several specific characteristics of teamwork perception of each group. For example, physicians' perceptions of other specialty groups are much more negative than their perception of their own group, particularly with respective to anaesthesiologists as well as to nurses in the surgical ward. Surgeons' teamwork perception of internal medicine physicians is also quite negative compared to that within their own group. However, surgeon's perception of physicians in another specialty, i.e., anaesthesiologists, is very positive and is similar to their perception of their own group. Finally, the nurse group in the internal medicine ward also perceives teamwork with physicians with whom they frequently cooperate as rather negative, and in fact just as negative as their teamwork with surgeons with whom they seldom work together.

These results suggest that teamwork perception is greatly affected by how much cooperation or collaboration opportunities a given group has another. There are, however, two exceptions, as just indicated: teamwork with surgeons is perceived as being relatively more positive by the other groups, even by those who work less frequently with surgeons. In contrast, teamwork with internal medicine physicians is perceived as being relatively poor even by the groups who frequently work with them.

4.3 Healthcare Staff's Work Goals

Responses of questions on work goals are summarised in Table 6 in terms of rank between fifteen items of the questionnaire as well as percentage agreement for each item. For the purpose of comparing with another professional culture, this table includes the responses from the Japanese pilots obtained by Helmreich and Merritt [6] that applied their early version of questionnaire – two old items were replaced with three new items in their later version, which we applied in the present study. As can be seen in this table, there is little difference in work goals among the three healthcare professional groups. The most important work goal for the Japanese healthcare staff is the interpersonal relationship with collaborators within a hospital such as "working with people who cooperate well with one another" and "maintaining good interpersonal relationship with all other medical personnel".

Next to the most treasured work goals, interpersonal relationship, there is a slight difference with respect to the second-most valued work goal between physicians, nurses and pharmacists, though very minor between the latter two groups. More than 80% of the physician group emphasised work factors such as freedom to adopting their own work procedure, and enough time to consider more than one solution. On the other hand, the physicians' concern with work itself or work content was not high: priority ranks for the work-related items, "job or career that will bring them prestige and recognition from others", "job about which they know everything with no surprise" and "changing work routine with new, unfamiliar tasks" were in the lowest ranks. In contrast, nurses and pharmacists put more stress on personal issues than physicians. More than 80% of these groups put emphasis on security of employment and sufficient time left for their personal or family life.

Table 6 Work goals of each professional group

Work values	Physicians		Nurses		Pharmacists		Japanese pilots[6]
	Rank	%agree.	Rank	%agree.	Rank	%agree.	
Good interpersonal relationships	2	89%	2	93%	2	88%	6
Opportunity for higher-level jobs	10	65%	12	48%	12	61%	8
Security of employment	6	79%	3	88%	3	88%	2
Environment where group's achievement are valued	12	54%	10	61%	11	62%	•|
Live in a desirable area	8	72%	7	80%	9	61%	3
Routine with new, tasks	13	58%	13	41%	13	47%	13
Time to consider solutions	2	83%	5	90%	4	81%	•|
Warm relationship with superior	5	80%	6	84%	6	77%	11
Freedom to adopt own approach	2	89%	9	74%	7	77%	•|
Opportunity for high earnings	9	70%	8	76%	9	58%	7
Challenging tasks to do	7	73%	11	62%	8	70%	5
Know everything about the job	15	17%	14	20%	15	10%	10
Sufficient time for personal life	10	62%	4	88%	5	80%	1
Work with people cooperating	1	97%	1	96%	1	95%	4
Job or career bringing prestige	14	23%	15	21%	14	19%	•|
Find the truth, the one solution	•|		•|		•|		9
Observe strict time limits	•|		•|		•|		12

There are no large differences between positions and in particular between positions of the nurse group. That is, nurses at any level of positions put the highest emphasis on work with people who cooperate well with one another, as well as maintaining good interpersonal relationship with all other medical personnel. For the physician group, there is little difference between leaders and consultants, but the work goal of the residents is slightly different from these two higher position groups. The resident group also put the highest emphasis on work with people who cooperate well with one another (100% of agreement). However, in addition, they also put emphasis on work factors such as freedom to adopt their own approach, time to consider more than one solution, and challenging tasks from which they get a personal sense of accomplishment. No difference was observed between departments or wards both for the physician and the nurse group.

In contrast to the healthcare staff, according to the result by Helmreich and Merritt [6], Japanese pilots attach greater importance to personal issues such as sufficient time left for their personal or family life, security of employment, and life in an area desirable to their family. In addition, pilots also put stress on interpersonal relationships as well as getting challenging tasks giving a personal sense of accomplishment. These differences between healthcare staff and the pilot may in part result from the working style or tasks and it may possibly also be related to the selection and recruitment processes of the professions. In hospitals, healthcare staff members work in teams most of the time, while pilots must frequently work for long periods away from their family, often for many days in a row during long-distance flights.

Table 7 Comparisons with ship officers in percentage (dis)agreement for safety culture aspects

Safety culture aspects		Healthcare staff			Ship officers
		Physician	Nurse	Pharma.	
I. Power distance	% agree.:	5.7%	8.6%	4.8%	8.3%
	%disagree.:	89.3%	79.2%	87.5%	81.4%
	X_0^2:	11.72**	10.58**		
II. Communication		86.4%	85.4%	85.7%	98.8%
		6.1%	3.8%	7.1%	0.7%
		21.83**	163.69**		
IV. Own performance under high stress		48.6%	42.5%	44.1%	38.3%
		38.4%	34.7%	31.5%	43.1%
		16.92**	53.73**		
V. Stress management for team member		71.6%	67.4%	68.4%	91.5%
		18.7%	17.7%	19.9%	3.0%
		48.16**	294.18**		
VI. Morale & motivation		80.5%	73.9%	71.7%	82.3%
		11.3%	9.5%	12.6%	7.8%
		0.002	61.04**		
VIII. Recognition of human error		36.2%	53.3%	39.3%	50.8%
		47.7%	33.6%	45.2%	36.2%
		17.11**	0.63		

Bottom row : Chi square:between doctors/nurses and ship officers (**: $p<0.01$, *: $p<0.05$)

5. COMPARISON WITH SHIP OFFICERS

For the purpose of comparison with ship officers' attitudes, we used response samples (a subset of questions that overlapped in the two questionnaires) from similar surveys in the maritime domain using an earlier, derivative version of the questionnaire of the present study, the SMAQ (Ship Management Attitudes Questionnaire) [3, 9]. With the SMAQ, we collected 444 samples from Japanese ship officers working in two Japanese ship companies. Comparisons between healthcare staff and ship officers are shown in Table 7 in terms of percentage agreement and disagreement of safety culture aspects. With regard to characteristics of professional culture, a comparison between the two samples show significant differences between physicians/ nurses and ship officers in all the aspects except for morale and motivation (where physicians and officers are alike) and for recognition of human error (where nurses and officers are alike). Both the physician/nurse and the ship officer groups had high motivation and morale. Similar to the hospital work environment, there was not a large power distance on the ship bridge. For the other safety culture aspects, ship officers assign a greater importance to communication during task performance and to stress management for team members than do both physicians and nurses. Moreover, ship officers' attitudes to human error are more realistic than that of physicians and, in terms of percentage agreement, identical to the attitudes of nurses. In contrast, physicians and nurses had slightly more realistic perceptions of the effects of stress on their own performance. Integrating these professional comparisons, we find that the safety culture among ship officers seems to be characterised by a somewhat greater safety awareness than that of hospital staff.

6. CONCLUSION

This paper reported the results of a questionnaire-based survey about safety culture related attitudes among hospital staff. We aimed at identifying safety-related perceptions and attitudes among healthcare staff in relation to patient safety. To elicit characteristics of hospital safety culture, we compared the questionnaire results with the data obtained in our former studies of the maritime domain [3, 9]. Finally, in addition to the healthcare staff attitudes towards managerial issues, we surveyed their perceptions and views of teamwork quality, leadership styles and work goals.

We analysed the healthcare staff's attitudes to safety culture related issues, classifying 57 question items into nine aspects: (1) power distance, (2) communication, (3) teamwork, (4) recognition of own performance degradation under high stress, (5) stress management for team members, (6) morale and motivation, (7) satisfaction with management, (8) recognition of human error potential, and (9) awareness of own competence. As general characteristics of safety culture in Japanese hospitals, we identified a moderate power distance between superiors and subordinate members within these professional groups, appropriate recognition of importance of communication, relatively high morale and motivation, and reasonable attitudes to stress management for team members. However, their satisfaction with management was rather low. In addition, healthcare staff is reluctant to acknowledge degradation of their own performance under high stress, workload, fatigue and other performance shaping and psychological factors. Recognition of human fallibility was reasonably high, but not realistic enough to be fully recognised as a substantial risk factor. Physicians' attitudes to and perceptions of safety culture related issues as well as other issues treated in this paper, cf. statements summarised below, seemed to be homogeneous across positions and specialties – although this conclusion can be made only tentatively, since there is a risk that we might make a "type 2 error" due to the small number of physician samples from individual specialties. On the other hand, differences were observed in most safety culture aspects between positions and wards for the nurse group as mentioned in Sections 3.2 and 3.3.

Respondents tended to judge the quality of teamwork more highly in relation to groups with whom they cooperate more frequently. In addition, surgeons' perceptions of teamwork were more positive than those of physicians, both within their own group and with other specialties. A similar pattern was identified between nurse groups working in surgery and internal medicine. Regarding perceptions of the leadership issue, most of Japanese hospital staff preferred slightly less democratic style than what European doctors and nurses do, i.e., a mildly autocratic style in which a leader usually makes decisions promptly, but tries to explain them fully to team members before going ahead. However, the leadership style that respondents found themselves working under most frequently is the autocratic style – e.g., a leader who usually makes decisions promptly and communicates them to subordinates clearly and firmly a perception identical to results from a German hospital. Finally, the Japanese hospital staff put the greatest emphasis on good interpersonal relationships within their work environment as their work goals. Doctors also stressed the value of working ways that allow discretion of work procedures and sufficient time given for examining several options to a problem. For nurses and pharmacists personal issues such as security of employment and sufficient time left for their personal or family lives are slightly more important for physicians.

In our former study [8], we derived some hypotheses concerning correlations between the risk of having incidents and some of the safety culture aspects, e.g., recognition of human fallibility, and power distance, based on a comparison between actual reporting statistics of incidents and the questionnaire responses obtained from a single hospital. For example, a particular professional group who has a relatively larger power distance and unrealistic recognition of human errors will be liable to produce a greater number of incidents. In a future study, we will examine the hypotheses based on combined results of a greater number of samples to a questionnaire like the one applied in the present paper and analysis of incident statistics obtained from multiple hospitals.

Finally, based on our efforts to examine the statistical correlations between actual incident rates and the perceptions and attitudes of healthcare staff, we would suggest that a questionnaire-based method may be a useful tool to estimate the present level of safety culture in relation to patient safety in a specific organisation or work unit in the medical domain. This is particularly of importance whenever incident reporting is incomplete or when reporting criteria are heterogeneous. Equally, while incident reporting is a retrospective index of safety levels, survey data may be used prospectively and, in combination with a proactive regime, to identify points at which a specific local safety culture may need to be strengthened.

ACKNOWLEDGMENTS

We are heavily indebted to Robert L. Helmreich, the University of Texas at Austin who provided not only the original “Flight Management Attitudes Questionnaire” and “Operating Team Resource Management Questionnaire” but who also inspired the adaptation of this survey instrument for patient safety. We would like to acknowledge Naomi Kitazawa, Tokyo Medical and Dental University for their cooperation in this project. We are also grateful to the Danish patient safety survey project group for permission to use their cases and question items. Special thanks are due to the hospital heads as well as the risk managers in the hospitals surveyed in this project for allowing us to analyse their incident reports.

REFERENCES

[1] ACSNI. *Advisory committee on the safety of nuclear installations: Human Factors Study Group Third Report: Organising for safety*. HSE Books, Sheffield, 1993.

[2] Andersen, H.B. Assessing safety culture. Technical Report R-1459, Risø National Laboratory, Roskilde, Denmark, 2002.

[3] Andersen, H.B., Garay, G. and Itoh, K. Survey data on mariners: Attitudes to safety Issues. Technical Report I-1388, Systems Analysis Department, Risø National Laboratory, Roskilde, Denmark, 1999.

[4] Andersen, H.B., Madsen, M.D., Hermann, N., Schiøler, T. and Østergaard, D. Reporting adverse events in hospitals: A survey of the views of doctors and nurses on reporting practices and models of reporting. *Proceedings of the Workshop on the Investigation and Reporting of Incidents and Accidents*. 127-136, Glasgow, UK, July 2002.

[5] Hee, D.D., Pickrell, B.D., Bea, R.G., Roberts, K.H. and Williamson, R.B. Safety management assessment system (SMAS): A process for identification and evaluating human and organization factors in marine system operations with field test results. *Reliability Engineering and System Safety*, 65: 125-140, 1999.

[6] Helmreich, R.L. and Merritt, A.C. *Culture at work in aviation and medicine: National, organizational and professional influences*. Ashgate, Aldershot, UK, 1998.

[7] Helmreich, R.L. and Schaefer, H.-G. Team performance in the operating room. In M.S. Bogner (ed.), *Human error in medicine*. 225-253, Lawrence Erlbaum Associates, Hillsdale, NJ, 1994.

[8] Itoh, K., Abe, T. and Andersen, H.B. A survey of safety culture in hospitals including staff attitudes about incident reporting. *Proceedings of the Workshop on the Investigation and Reporting of Incidents and Accidents*. 144-153, Glasgow, UK, July 2002.

[9] Itoh, K. and Andersen, H.B. Motivation and morale of night train drivers correlated with accident rates. *Proceedings of the International Conference on Computer-Aided Ergonomics and Safety*. Barcelona, Spain, May 1999 (CD ROM).

[10] Itoh, K., Andersen, H.B, Seki, M. and Hoshino, H. Safety culture of track maintenance organisations and its correlation with accident/incident statistics. *Proceedings of the 20th European Annual Conference on Human Decision Making and Manual Control*. 139-148, Copenhagen, Denmark, June 2001.

[11] Itoh, K., Andersen, H.B., Tanaka, H. and Seki, M. Attitudinal factors of night train operators and their correlation with accident/incident statistics. *Proceedings of the 19th European Annual Conference on Human Decision Making and Manual Control*. 87-96, Ispra, Italy, June 2000.

[12] Kohn, L.T., Corrigan, J.M. and M.S. Donaldson (eds.). *To err is human: Building a safer health system*. National Academy Press, Washington DC, 1999.

[13] Reason, J. Managing the management risk: New approaches to organisational safety. In B. Wilpert and T. Qvale (eds.), *Reliability and safety in hazardous work systems*. Lawrence Erlbaum Associates, Hove, 1993.

[14] Reason, J. *Managing the risk of organizational accidents*. Ashgate, Aldershot, UK, 1997.

Individual Contribution to Safe Collective Management in Dynamic Situations. The Case of a Medical Emergency Centre (SAMU).

Jacques Marc
Laboratoire Cognition & Activités Finalisées
Université Paris8-CNRS, 2, rue de la liberté, 93526 Saint-Denis Cédex2
jmarc@idf.ext.jussieu.fr

Janine Rogalski
Laboratoire Cognition & Activités Finalisées
Université Paris8-CNRS, 2, rue de la liberté, 93526 Saint-Denis Cédex2
rogalskij@univ-paris8.fr

ABSTRACT

This paper presents an empirical study about the individual contribution to the safety of collective management in a medical emergency centre (SAMU). The study aimed at extending the model of ecological safety developed by Amalberti (1996) for individual activity to the case of collective management. More precisely, the study focuses on how individual contribute to safety of collective action in the case of risky dynamic environment management. On the basis of long-term observation of work activity, a "pseudo-simulation" was designed for confronting actors in an emergency centre (physicians and "on-call operators") with a series of "errors" in the functioning of "their" centre. They were asked to comment what they observed during short periods of simulated activity and to assess safety and situation mastery by the various actors observed. Results are coherent with the existence of error management in collective action as being a matter of ecological safety: error management was depending on potential error consequences: Subjects focused mainly on errors with may have operational consequences, on the global quality of safety management in the centre, and ascribed errors more frequently to actors to whom they assessed a lower level of situation mastery.

Keywords

Error management, ecological safety, simulation, risky dynamic environments, emergency centre.

1. INTRODUCTION

This paper presents an empirical study about the individual contribution to the safety of collective management in a medical emergency centre (SAMU). Safety is considered from the point of view of the patients concerned by the assistance calls. More generally, in rescue or emergency centres, the objective is to limit as far as possible the damage issued from events threatening human integrity. Safe management in such types of situations is assessed by how the system fills out its "duty of means". Studies on how operators contribute to the un(safety) of systems involving risk were initially focused on human errors, with the aim of decreasing their occurrence.

Research has then shown that error is an intrinsic component of human activity: that led to the notion of error-tolerant systems. In the line of a deep analysis of error management in risky dynamic situations, Amalberti developed a model of "ecological safety": It postulates that operators behave under a cognitive compromise between the "external risk" of negative issues of their making errors and the "internal risk" of their loosing the cognitive control of the situation, due to cognitive overload involving error management (the operator may "ignore" errors even if s/he identifies them).

The study aimed at extending this model of ecological safety assessed for individual activity to the case of collective management. More precisely, a "pseudo-simulation" was designed for confronting actors in an emergency centre (physicians and "on-call operators") with a series of "errors" in the functioning of "their" centre. The design of the pseudo-simulation and of the "experimental" task individual prescribed to the centre operators will be described in some details. Three types of quantitative data will be presented: how do individuals acting as observers of collective activity manage errors depending on their nature; to whom were they ascribing errors; how high or low do they assess situation mastery by the various actors they observed.

Results are coherent with the existence of error management in collective action as being a matter of ecological safety: error management was depending on potential error consequences: Subjects focused mainly on errors with operational consequences, on the global quality of safety management in the centre, and ascribed errors more frequently to actors they assessed a lower level of situation mastery.

2. ECOLOGICAL SAFETY

Human error has been widely analysed from the point of view of individual activity (Reason, 1990), with a more recent emphasis on risk management within an organisational point of view (Reason, 1997; Rasmussen, 1997). The model of ecological safety (Amalberti, 1996; 1997) was developed within this context.

2.1 Ecological safety for individual activity

Amalberti (1996) proposes a model of ecological safety for individual activity, which explain results from various studies on errors in risky dynamic environment management, showing that even expert activity is not free of error and that not all perceived error is recovered.

This model postulates that operators regulate their production to reach the main objective of their tasks, while not engaging too much cognitive resources for too a long time. This regulation does not lead to the best cognitive activities in situation representation and in decision making, but it allows operators to maintain a "sufficiently good" level of overall performance, and to reach their main objective, with cognitive constraints at a level as low as possible leaving them resources enough to adapt to possible unexpected situation evolution. This regulation has thus three roots, which may be contradictory: 1) the need to reach the main objective, 2) the need to do ensure safety, and 3) the need to preserve their own capacities (both mental and physiological) as "reserves" in case of unexpected further work load and for maintaining a sufficiently good overall performance over time.

For Amalberti, the human operator is the only system who can adapt in anticipation to the dynamic needs of situations. There are consequences from the point of view of the representation of the situation (situation assessment) and from the point of view of error management. Operator elaborates dynamically a representation of the situation, rich enough to permit decision making and simple enough to be easily managed within the scope of the situational constraints (for information taking and for acting).

Operator is thus constantly doing a cognitive compromise between his level of situation awareness, his knowledge of possible situation evolution and disturbances, and his knowledge about his own competencies to face future situation states — that is his meta-knowledge (Valot, Grau & amalberti, 1993). For this he used to simplify his representation and choose the less demanding way to behave, leaving the resources free to manage situation evolution. The cognitive compromise can be seen at another level: When simplifying his representation, operator makes a compromise between an *external risk* (or "objective" risk) of negative issues through him making errors and an *internal risk* (or "subjective" risk) of him loosing the cognitive control of the situation.

In these situations, errors play a crucial role. They are both cues of system break and operator's performance cues: their number and type inform the operator on the quality of his cognitive compromise and on the resistance of the system he is controlling. If they become too numerous or threatening to lead to serious consequences, they inform the operator he is loosing his mastery of the situation. But if they are few or with minor consequences (with regards to the main goal), they can be accepted. Moreover, the operator may use them to act "at the limits" of the system in order to regulate his internal risk. So, the operator can have a feeling of situation control with the presence of errors, and being effectively within a safe envelop of action. More, in dynamic situation, it could be dangerous to try to recover every error, even those that have no serious consequences, because this activity might mobilise resources to the detriment of situation management.

Such a model was developed for and supported by studies on individual error management. A further issue is to what extend it is valuable for analysing collective activity.

2.2 Issues of error management and ecological safety in collective activity

The issue of collective action in dynamic risky situations has been strongly focusing on collective performance (Brannick, Salas & Prince, 1997), and training (Rogalski, 1994), particularly for team resource management. However, linking human error assessment and human reliability assessment is not straightforward, even in individual activity (Kirwan, 1992; 2001). It is particularly the case for collective management of dynamic situations with a lot a variability, both impeding the design of procedures which could be used as references and requiring complex co-operation processes under time pressure (Antolin-Glenn & Rogalski, 2002).

Collective safety can be analysed at three levels:

- organisational level (the group as a sub-system functioning within a larger system): this is the case in Rasmussen's analysis (1997);
- collective level (the group as a self-active system of interacting actors),
- individual level (the group action results from individual activities: for each actor, the other ones are components of the world to be managed).

We used the two last approaches of error management in collective activity, and will detail the one that was used in the case of a medical emergency centre through a "pseudo-simulation". The first approach considers the team (whatever its organisation) as an entity and transposed methods of individual activity analysis to the activity of such a *virtual operator* (Rogalski, 1991). Errors production, detection, recover, are then analysed without consideration of who is the individual actor involved; such an approach is strongly task oriented. It is possible to analyse error management as for individual activity (and directly test the validity of the model of ecological safety).

The second approach focuses on what is developing within the team, and aims at identifying cognitive activities oriented towards both the individual activities and the articulation of individual activities in the common task. Operative communication is a classical resource for such analyses. Besides, analysing the conditions of operative communication in complex collective activity allowed showing that what Rognin (1996) called "*pluri-addressing*" (pluri-addressed communication) was an important point for collective reliability.

From the last point of view, an individual can have four functions in collective safety through error management:

- producing / managing his/her own errors as an individual actor
- managing other operators' errors, as observer and co-actor
- producing positive or negative interferences (Loiselet & Hoc, 1998) as regards to error management by other actors (including production and recovery)
- managing the global process of co-ordination of individual activities in a team activity.

Depending on the situation to be studied, these methods are more or less usable. The characteristics of the co-operation within a medical emergency centre led to use a method focusing on one of the function of individual activity with regards to collective action: managing errors as an observer of collective activity. Such a method has been used for comparing novices and experts with regards to individual error detection (Doireau, Wioland & Amalberti, 1997). We will discuss the benefits and limits of such an approach from the point of view of activity analysis and from the point of view of training purposes.

3. THE EMPIRICAL STUDY

The study presented here took place in a medical emergency centre (Marc, 2002), where several operators act collectively under strong constraints due to workload, responsibility and temporal pressure. From an operational point of view, such a centre is organised into two systems: 1) a regulation centre receives calls for assistance and has to identify their emergency for deciding which type of means has to be sent; 2) an operation system, consisting in intervention teams (ambulances) who have to transport the patient to hospital emergencies.

3.1 The situation of reference: Collective activity in a medical emergency centre

The regulation centre (where the study was done) provides a case of complex collective work: numerous operators; distributed co-operation: interaction with a diversity of "callers" (Navarro & Marchand, 1994), decision about medical emergency, orders to operation teams; hierarchical differences: physicians and phone —on-call— operators; computer-based shared tools (file for each call); temporal pressure; information ambiguity; responsibility for human life; phases of high workload. The task is strongly prescribed in terms of administrative procedures to be followed, and of task allocation between the emergency centre members.

The medical emergency regulation centres (in France) are organised with a common basis: physicians are responsible to interact with callers in order to diagnose how urgent is the medical problem motivating their distress call, while call-takers (in French, *Permanenciers Auxiliaires de la Régulation Médicale*: PARM) are in charge of the first interaction, and can play some filter role, when it is clear that there is not a urgent medical problem -for instance drunk people calls, or jokers. Call-takers are also giving missions to the medical ambulances (*Service Mobile d'Urgence et de Réanimation*: SMUR), a radio operator being responsible of a specific communication network (this task may be devoted to a call-taker). More generally, they are in charge of external interactions, to the exception of the medical ones.

The basic composition in SAMU 75 is two call-takers (one of them is in charge of the radio), and one physician. Depending on the workload (expected emergency calls), there may be more call-takers and physicians in the regulation room (until 6 call-takers and 3 physicians, one being specialised in child care). It is up to physicians to ensure teamwork quality: in this respect, they play a role similar to those of captains in civil crews (Jentsch et al., 1999).

Activity is organised around cases (or "stories" or "affairs"), which constitute units of action: a call triggers a case management. Each case is (has to be) linked to a record in a file, which is open by a call-taker, will be also used by the physician and constitute both a tool for information sharing and a common object of action (filling it as required is an administrative and legal obligation).

In normal situations, a case management involves a lot of "micro-procedures" concerning patient, file, and internal communication. So, the first actions to be undertaken are the following: the call-taker takes the caller, opens a file, locates the call, performs a first sorting and emergency assessment, "informs" the file, informs the physician, closes the file while commuting the call to the physician —with the file number; the physician takes the call, opens the file, assesses medical emergency and if possible diagnosis, fills in the medical record fields, chooses a mean of action, write down this information, informs the PARM about his/her decision, closes the file, and if necessary relays the call to the PARM. Action continues until the patient has been taken in charge, and the case file closed. Any departure from the requested organisation of such micro-procedures is considered as an 'error': not all of them may result in threatening the patient safety, but all may have administrative, financial or legal consequences. This was the basis of the data we analysed.

The complexity of the co-operation situation, the time pressure and the variability of the cases processed in the regulation centre make it very difficult to record co-ordinated data about the activity of the team members in real time: the method for analysing the pertinence of the model of ecological safety was based on an observation of collective work, a record and analysis of errors and their management from the point of view of an individual (a PARM), and the use of a "pseudo-simulation" for a systematic study of error management from SAMU operators acting as observers (their position was similar to those used in Doireau, Wioland and Amalberti (1997) for individual activity. We will here focus on this last situation.

3.2. From observations to pseudo-simulation situation

After several months of observation by the first author, the activity of a PARM was systematically followed during two months. It enables to record and analyses a wide set of errors management occurrences (Marc et Amalberti, 2002). The major results are in line with the ecological safety model: operators do not detect or notify all errors; only some of them are managed at short or mean term. This results concern the errors made by the observed operator as well as errors made by others (identified by the observer, and which could have be noticed by the operator observed). An important component of error management is group oriented.

From this set of data, a "pseudo simulation" was designed. It was based on a series of static views of the regulation centre, with three actors involved: two phone operators and a physician. These views were linked by a continuous audio tape (operative communication with the callers, the operation teams and within the regulation centre, including noise currently heared in the regulation room). The centre activity was seen and hear from a PARM point of view, including the screen with the current state of the open file. The "pseudo-simulation" lasted about 10 minutes, with stops after periods about 2 minutes.

Actors of the regulation centre were put in the position of observers of the centre activity as shown in the pseudo-simulation. They were ask to comment each short session between two stops. They were finally asked to assess how the various actors were mastering the situation, and to evaluate the level of risk.

In order to avoid errors to be detected on the basis of their further consequences, each error that was made in a session has its end on another session. Physicians and PARMs (13 of each group, almost all the regulation centre personal) were individually asked to observe the simulated situation, and to resume it during simulation breaks.

The term of "pseudo-simulation" was used in order to emphasise the fact that the operators (PARMs and Physicians of the regulation centre) were not in a position of actors —at the difference with ordinary simulations— but were in a position of observers. They could comment, evaluate, propose actions but were no able to act. With regards to the transformation made from the real situation in the regulation centre, there was a "decoupling" (Samurçay & Rogalski, 1998) of the functions actors may perform in collective work: the function of observing other actors' actions was conserved, but the function of being ownself an actor was prevented.

The evaluation of the model of ecological safety for collective work is done through the following points:

- not all errors are categorised as such (even if the corresponding activity is detected and memorised), and that there is a hierarchy between types of safety events, in favor of those linked to the main operational goal: patients safety

- errors ascribed to actors will be related with the evaluation of mastery of the situation; more precisely, if the model is relevant, actors considered as less mastering the situation will be more often considered as errors producers.

3.3 Categories of safety events and actors initiating error management

In order to evaluate the relevance of the model of ecological safety for the case of collective work, we categorised safety events introduced in the pseudo-simulation with regards of their nature and of their "producer".

The "pseudo-simulation" was designed from eight (observed) stories; and organised into 49 episodes, involving 76 safety events. An episode is corresponding to a static view of the movie and the audio information associated. It reports the group collaboration on one or more event.

For example, in episod 12, PARM2 receives information about the means asked for a cardiac problem PARM1 was managing, and she interacts with these means. PARM2 has not completed the file of a previous case, that the physician was begining to treat.

This episod concerns two cases (or stories). It involves three safety events: a negative interference (of PARM2 with PARM1's activity); and two errors in administrative management: PARM1 tramsmits a file for evaluation before patient localisation, the Physician accepts an incomplete file (no defence against the previous error).

Safety events were categorised into seven categories, as concerning their nature:

- positive defences (coded Defences Pos.): protection against error, anticipation for avoiding conditions for error, error recovery

- negative defences (coded Defences Neg.): a possible protection is not performed, anticipation is inadequate, possible error recovery is not performed

- error in medical regulation (coded Medical Reg.): An actor makes an error that has an implication on medical regulation: bad selection of means

- error in interpretation (coded Interpretation): bad representation of a situation at the informational level

- error in administrative management (coded Administrative): bad management of a file by an actor

- negative interference (coded Interference): An actor interferes with another actor's activity, (without being asked for, and without proposing it).

- other safety events (coded Others): actions involving safety and not clearly categorised into the previous categories (such as discussion about cases addresses).

Elsewhere, safety events can be categorised depending on the actor(s) who was their initiator: PARM1, PARM2 or Physician. (Two safety events — a medical regulation and an interference— were involving PARM2 in interaction with another actor).

Table 1 presents the distribution of safety events in the pseudo-simulation with regards to these categories.

Table 1. Distribution of safety events (numbers) depending of their nature and their initiator (P1: PARM1; P2: PARM2; PHY: physician; T: interaction of two of them)

		Actors involvement in the safety events			
Type and number of the safety events		**P1**	**P2**	**PHY**	**T**
Defences Pos.	15	7	5	3	
Defences Neg.	6		6		
Medical Reg.	13	2	3	7	1
Interferences	15	2	10	2	1
Administrative	12	4	7	1	
Interpretation	9	2	4	3	
Others	6	1	3	2	
Total	76	18	38	18	2

3.4 Method for data analysis

Two groups of actors in the regulation centre —13 PARMs and 13 physicians— were participating to the experiment (allmost the whole centre, to the exception of two experts who evaluate a first prototype and contribute to the final choices, and a few persons due to operational duties). They all consider the simulation to be quite a realistic view of what could happen in a regulation centre. Some were even trying to "take the hand" on the file open on the simulation screen in front of them.

Several analyses of the recorded comments and of the evaluations of situation mastery were performed (Marc, 2002). They were done in three steps. At the first step, quality of operational memorywas attested: in effect, the place of safety events management can only be assessed on the basis of a sufficiently good operational memory of cases they were concerning and episodes in which they were occurring. In a second step, we identiy references to safety events in comments of the observed simulation. In the last step, we analyse evaluations of situation mastery and level of risk (on scales form 0 : no mastery at all, to 7: perfect mastery of the situation).

4. RESULTS

4.1 Operational memory

As frequently observed in real operational settings, operational memory was quite high.

To the exception of one case that did not imply neither safety nor the main goal of the regulation centre, PARMs and Physician remembered all cases of the pseudo-simulation. Episods were variously reported, but a majority (more than 55%) of the 49 episods were reported by more than three quarters of the subjects. Only 6 episodes were reported at a low level in one of the groups. The episodes highly reported were reported by the two groups of subjects (PARMs and Physicians) at a high or medium level, to the exception of two episods, which were quite differently memorised by physicians and PARMs: these were more aware of the episode involving the actor of the professional group they were belonging to.

The quality of operational memory has to be evaluated in taking into account the fact that subjects were not asked to memorise cases, episodes or events but to comment what they observed in the successive sessions of the simulation.

4.2 Actors referred to in the comments

In order to identify some possible effects of individual position (in the situation of reference, that is normal activity in the regulation centre) in operational memorisation, we analysed the distribution of references to the various actors in the pseudo-simulation (without taking into account decontextualised verbalisations, which were not concerning the *hic et nunc* observed situation in the simulation)

References to individual actors and their actions in the various episodes was dominant, with a similar weight for physicians and PARMs (more than 60% of the verbalisations concerning the observed situation). References to the actions of the regulation centre as an entity was more frequent by physicians than by PARMs (20% *vs* 13,7%): the later were more often refering to groups of identified actors.

The physician in the pseudo-simulation was almost the first reference (29,3% for physicians, 24,9 % for PARMs) although the simulated situation was seen from PARM2's point of view, who performed a lot more of observable actions (and errors) than the physician (he was refered to in about 20% of verbalisations concerning the situation).

Such differences can only be interpreted as resulting of the position of the simulated actors with regards to the situation of reference: the importance of the refernces to the physician in the subjects' verbalisations may be understood by the central position the physician occupies in the decision hierarchy and responsibility. They also attest the ecological validity of the pseudo-simulation, with actors considered as in the real regulation centre life.

4.3 Comments about safety events

The first element with regards to the ecological safety model is that almost 50% of the verbalisations issued from the demand of comments were effectively concerning safety: error detection, identication of defences, anticipation of consequences or errors, proposition of solutions. Error detection represents about half of these verbalisations, physicians being slightly more sensible than PARMs (52.6% vs 42.6%).

A second cue is the hierarchy of type of safety events in safety concerns. The other half being error identification.

We consider the distinction between detection and identification to be an important point in evaluating the relevance or the model of ecological safety. In error detection, the subject just notices that there is something wrong, it is like the error suspicion for Alwood (1984); in error identification, the subject makes a diagnosis about the error, even sometimes a prognosis about how it could be recover.

As observed in natural settings in so-called "naturalistic decision making" (Klein et al., 1993), errors identifications often called for (propositions of) solutions by the subjects commenting the simulated situation. This is a sign that there was a strong implication of operational representations about safety which were "triggering" these proposals: the solution is already active and just "waits" for sufficient cues of external or internal risk to be launched.

The fact that, on the one hand, safety events were detected in actors' actions and not simply remained unnoticed, but, on the other hand, were not for all that identified as being errors, is in favour of the hypothesis of a differential error management, depending on the type of safety events.

Table 2 presents a view about the differences between the types of safety events, depending on how many subjects refered to them.

Table 2. Distribution of the level of reference to the various types of safety events (Low: less than 9 subjects refering to such an event; High: more than 17 subjects; Average: between 9 and 17)

Type (and number) of safety events		Level of reference	
	Low	Average	High
Regulation (11)	3	7	5
Defences Pos. (15)	5	7	3
Defences Neg. (6)	2	3	1
Administrative (12)	5	6	1
Interpretation (9)	3	5	1
Interference (15)	11	3	1
Others 6	2	4	-

From Table 2, it can be seen that safety events were in fact considered from the point of view of their potential consequences: regulation errors were the first focus, then positive defences (recovery or avoidance through anticipation); administrative errors, and problems of interpretation (without direct regulation consequences) were not highly considered. Interference problems, concerning the collective activity in itself were the lowest focus of interest, as if the main focus was on "what happen to the critical objects" (action decision about patients) issued from the centre as a whole, and if activity itself was in the last place, until it does not lead to critical safety events.

Beyond these data, some observations are of interest: subjects were checking, until the end of the error life, if it was producing critical consequence. Some subjects, at the end of the experimentation, were speaking about errors done at the beginning of the simulation. Such a supervision of errors along time was observed in analysing individual error management.

4.4 Ascribing safety events

When commenting about a session, or at the end of the simulation, subjects could ascribe safety events (and particularly errors) to a specific actor, an interaction between actors, the group as a whole. They could also speak about errors as "facts" and not actions, without refering to any actor.

There was no important differences between physicians' and PARMs' verbalisations, except that physicians were relatively more focused on the regulation as an entity, and PARMs on physician / PARMs interactions.

Table 3 presents the distribution of global verbalisations, depending on their "contextualisation": about the stories of the pseudo-simulation, or general verbalisations, without references to any story of the simulation script.

Table 3. Distribution of safety events ascribed to actors or groups of actors in per cents of verbalisations on safety events (references to the simulation stories, or general references)

Responsible actor	**Types of verbalisations**	
	Stories	General
Physician	26,8	26,7
PARMs	27,4	5,1
PARM2	15,6	3,8
PARM1	11,8	1,3
Centre as an entity	10,3	27,7
Interactions	28,7	18,7
physician /PARM	11,8	10,7
between PARMs	7,9	8
No actor ascribed	12,8	17,7
Actors outside the regulation centre	3	5
Total number of verbalisations	533	159

In ascribing safety events observed in the pseudo-simulation, subjects were stronly oriented towards group activity: Entity + Interactions. Morevoer, it appeared an overestimation of the physician's role: almost 40% of the safety events mentionned were ascribed to the physician's actions, compared with the effective 24% in the simulation script (see Table 1). The decalage between the actor's initiative in safety events in the script and in the verbalisations about the simulated situation is quite interesting.

Even when taking into account the relative importance of safety events (Table 2), the importance of the physician as the first actor in safety events remains remarkable. It looks like if he was perceived as the first responsible of the quality of the regulation centre life, as regards to safety.

The point is reinforced in the decontextualised verbalisations (not speaking about the stories of the simulation script). In fact, these verbalisations were stronlgly oriented towards the overall group activity (centre as an entity + no actor ascribed + interactions = 64,2 % of the general verbalisations). The importance of ascribing safety events to the physician (26,7%) is all the more highlighted.

4.4 Evaluation of situation mastery

In their comments, subjects evaluated the actors' activities in the "pseudo simulation" as being "normal" ones. As a result, they globally evaluated situation mastery at the mean point (3,4, on the 0 to 7 scale), lower for the group of physicians than PARMs (mean evaluation: 3,1 *vs* 3,6). However, as shown in Table 4, there were significant differences between their evaluations of the situation mastery for the various actors in the simulation.

Table 4. Evaluation of level of mastery (on a 0-7 scale) by the various actors of the "pseudo simulation" and the centre as an entity, depending on the group of subjects (PARMs or Physicians)

Actors in the script	**Subjects**		
	PARM	PHYS	**Mean**
Physician	3,3	2,1	2,7
Entity	3,5	3,1	3,3
PARM 2	3,5	3,5	3,5
PARM 1	4,0	3,9	3,9

The order in the differencial evaluations was the same for the two groups of subjects: PARMs and Physicians. The physician was evaluated as having the lowest mastery of the situation. This was particularly strong for the groups of physicians who were highly critical toward their colleague in the script of the simulation.

The increasing order of evaluation of the situation mastery: physician, centre as an entity, PARM2, PARM1, was observed for every subject, with only two exceptions (concerning the relative place of the entity, and/or of the two PARMs).

5. DISCUSSION AND CONCLUSION

The results support the hypothesis that the model of ecological safety based on the notion of cognitive compromise was pertinent for collective action. The "pseudo-simulation" defined an observer position for the regulation centre members. Such a position was the results of "uncoupling" the different positions of the members of the regulation centre (observer, actor, decision-maker).

Data are converging toward the conclusion that safety events are considered depending on how their consequences main affect the actors' main goal (here: patients safety): regulation issues are most often referred to, then positive defences, while administrative errors and interference problems were the least often referred to (the order did not reflect the effective place in the simulation script). It does not result from insufficient observation and/or memorisation of what was happening in the observed centre, as operational memory was quite high. This result is coherent with observations done about individual action.

Another perhaps more interesting result does concern the collective dimension itself. Subjects ascribed safety events in a distorted way with regards to the objective script: when commenting about the stories of the script, subjects strongly overestimated the physician's involvement in safety events. At the contrary, less safety events were ascribed to the two PARMs, and there was less differences between them than in theirs producing errors in the script (PARM 2 was producing more errors). Moreover, while interferences as such were not highly noticed, interactions between actors were considered as producers of safety events.

The differences between PARMS can be explained by the type of safety events they produced (PARM 2 was making a lot of errors in administrative management, which were underestimated as types of errors). It is not the case for the overestimation of the physician's involvement.

Evaluation of the mastery of the situation is converging with ascribing safety errors to actors. The difference between PARMs was more close of theirs producing errors: PARM2 was considered as mastering situation at a lower level than PARM1. This could indicate that errors (even when seen as minor ones) are cues of the mastery of the situation, from the point of view of action observers in collective work.

The physician was considered the less in control of the situation, in quite a coherent way with the evaluation of his involvement in safety errors. This was particularly underlined by the physicians' evaluations.

This can be considered as a specific point for a model of ecological safety in the case of collective work: centering safety events management on their consequences for the main operational goal and on the responbility of these consequences leads to focus on the quality of the activity of actors in charge of the collective work. In the same line, results concerning the centre as an entity, producer of safety events as well as mastering the situation more or less well, emphasise the collective dimension in a model of ecological safety, from the point of view of the virtual operator constituted by the group as such.

The strong implication of subjects in observing the simulated collective activity and in commenting and evaluating it, leads to propose that such a pseudo-simulation could be used both as a basis of reflection of actors about their practice as a group, and as a basis for training newcomers, whatever their position: physicians and call-takers (PARMs).

6. REFERENCES

Allwood, C.M. (1984). Error detection processes in statistical problem, *Cognitive Science*, 8, 413-437

Amalberti, R. (1996). *La conduite des systèmes à risques*. Paris: PUF.

Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems. *Safety Science*, *37*, 109-126.

Antolin-Glenn, P., & Rogalski, J. (2002). Expertise in distributed cooperation: a Command Post in operational management. In S. Bagnara, S. Pozzi, A. Rizzo & P. Wright (Eds.), *ECCE'11. Cognition, culture and design* (pp. 143-150). Catania IT: CNR. Istituto di Scienze e technologie della cognizione.

Brannick, M.T., Salas, E., & Prince, C. (Eds.) (1997). *Team performance assessment and measurement. Theory, methods and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Doireau, P., Wioland, L., & Amalberti, R. (1997). La détection des erreurs par des opérateurs extérieurs à l'action. *Le Travail Humain, 60*(2), 131-153.

Hoc, J.-M. (1998). How can we evaluate the quality of human-machine cooperation ? In F. Darses & R. Zaraté (Eds.), *COOP'98 Third International Conference on the Design of Cooperative Systems* (pp. 121-130). Le Chesnay, FR: INRIA.

Jentsch, F., Barnet, J., Bowers, C. A., & Salas, E. (1999). Who is flying the plane anyway? What mishaps tell us about crew member role assignement and air crew situation awareness. *Human Factors, 41*(1), 1-14.

Kirwan, B. (1992). Human error identification in human reliability assessment. *Applied Ergonomics, 23*(5), 299-317.

Kirwan (2001). Assessing Human Error in air traffic management systems design : methodological issues. *Le Travail humain*, 64(3), 269-289.

Klein, G., Oranasu, J, Calderwood, R., & Zsambock, C. (Eds.) (1993). *Decision making in action: models and methods*. NJ: Ablex

Leplat, J., & Rasmussen, J. (1987). Analysis of human errors in industrial incidents and accidents for improvement of work safety. In J. Rasmussen, K. Duncan & J. Leplat (Eds.), *New Technology and Human Error* (pp. 157-168). New York: Wiley & Sons.

Loiselet, A., & Hoc, J.-M. (1999). Assessment of a method to study cognitive cooperation. In J.-M. Hoc, P. Millot, & E. Hollnagel (Eds.), *Proceedings of CSAPC'99* (pp. 61-66). Valenciennes: P.U.V.

Marc, J. (2002). *Contribution individuelle au fonctionnement " sûr " du collectif: Protections cognitives contre l'erreur individuelle et collective (le cas du SAMU)*. Thèse de doctorat de psychologie, Université Paris-VIII.

Marc, J., & Amalberti, R. (2002). Contribution individuelle au fonctionnement " sûr " du collectif : l'exemple de la régulation du SAMU. *Le Travail Humain, 65*, 217-242.

Navarro, C., & Marchand, P. (1994). Analyse de l'échange verbal en situation de dialogue fonctionnel : étude de cas. *Le Travail Humain, 57*(4), 313-330.

Rasmussen, J. (1997). Risk management in a dynamic sociaty, a modelling problem. *Safety Science, 27*(2-3), 183-214.

Rogalski, J. (1991). Distributed decision making in emergency management : using a method as a framework for analysing cooperative work and as a decision aid. In J. Rasmussen, B. Brehmer & J. Leplat (Eds.), *Distributed decision making. Cognitive models for cooperative work*, (pp. 303-318). Chichester: Wiley and Sons.

Rogalski, J. (1994). Formation aux activités collectives. *Le Travail Humain*, 54(4), 367-386.

Rognin, L. (1996). *Activités collectives et sûreté de fonctionnement*. Thèse de doctorat, Paul Sabatier Toulouse.

Samurçay, R., & Rogalski, J. (1998). Exploitation didactique des situations de simulation. *Le Travail Humain, 61*(4), 333-359.

Valot, C., Grau, J.-Y., & Amalberti, R., (1993). L'opérateur ne gère pas que le processus : il doit aussi se gérer lui-même. Représentation de ses propres compétences. In A. Weill-Fassina, P. Rabardel & D. Dubois (Eds.), *Représentation pour l'action*, (pp. 271-293). Toulouse: Octares

Remembering and Forgetting Effects in Analyzing Search Engine List Results

Francesca Rizzo
University of Siena
Via dei Termini 6
53100 Siena
francy@media.unisi.it

Oronzo Parlangeli
University of Siena
Via dei Termini 6
53100 Siena
parlangeli@unisi.it

Sebastiano Bagnara
Politecnico of Milano
Via Durando 38/A
20131 Milan
sebastiano.bagnara@polimi.it

ABSTRACT

The diffusion of technology to support information seeking and browsing has conducted to an exponential increase of the overall available information. However, though many studies have been conducted to model human search behavior supported by information systems (Marchionini, 1995, Navarro-Prieto, Scaife & Rogers, 1998; Nielsen, 1997; Shneiderman, 1998), a scant attention has been paid to the issue of the amount of cognitive resources requested to a person when she/he is using a search engine to solve a particular information problem.

This work aims at investigating how people manage demands on cognitive resources when accomplishing an information seeking task on a list of results produced by a search engine. More specifically, the two experiments reported here, in which subject had to perform a query using a simulated search engine, were conducted (1) to analyze the relation between the complexity of the information problem people have to solve using a search engine and the amount of information they can access; (2) to understand the role of forgetting in managing the information elaboration process.

The pattern of results obtained in the two experiments suggests the relevance of the reiterative activity inside the working memory, since different loads on working memory determine corresponding level of forgetting. Under the conditions considered here the hypothesis of the occurrence of deliberate forgetting in order to facilitate information elaboration and learning does not appear to be theoretically necessary.

Keywords

Information processing, forgetting, expertise, search engine.

1. INTRODUCTION

Many studies conducted to analyze human searching behavior (Marchionini, 1995, Navarro-Prieto, Scaife and Rogers, 1998; Nielsen, 1997; Shneiderman, 1998) showed that seeking activities, far from being stable and well structured, appear to be highly uncertain and dependent on different variables as the nature of the problem, the context in which the problem occurs, the nature and the accessibility of tools to conduct the search etc.. In addition, the huge amount of information provided by many information searching tools may be considered as affecting searching behavior as well. They not only increase information accessibility, but make people connectivity possible every time and everywhere (while driving, working, playing, shopping etc..) and change the same searching activity. In this new scenario, searching is not only realised through cognitive strategies such as information narrowing, evaluation, selection and so on. People, on the contrary, must clarify and understand specific states of the world. In fact, it seems reasonable to describe information seeking as a sort of meaning making activity that people exhibit when dealing with unknown and not easily delineable circumstances. As a consequence, when searching people are more likely to focus only on information items they judge relevant to satisfy their needs of knowledge. From this point of view, the information seeking process consists more in a balance between information acquisition and loss than in an infinite browsing of the information space by a person. In the process of searching, people consistently loose information they judge not useful to solve their information problem. Though this loss may assume different forms, generally people define this as "forgetting" and they associate a negative feeling to it. Forgetting is considered something to be avoided, a negative side effect. The hypothesis of forgetting as a negative loss of information is directly referred to the theory that considers knowledge stored in memory as traces (Rock and Ceraso, 1964; Paul, 1967; Neisser, 1967; Hintzman, Curran and Oppy, 1992). According to this hypothesis, human memory is able to produce copies/script of experienced events and to preserve them. These recordings, or traces, may deteriorate eventually. Therefore if the original event does not occur again to reinforce its traces, its memory will be automatically lost.

Another explanation of forgetting is the theory of interference. The interference hypothesis focuses on a competition mechanism that might occur among pieces of knowledge sharing some cognitive resources (Underwood, 1957; Bower and Mann, 1992). In this light the cause of forgetting is not related to natural cues decay in time, but to interference that intervene among similar memory items. Another hypothesis considers forgetting as a phenomenon that can be produced as the result of a failure occurring during information retrieval processes (Neisser, 1967; Baddeley, 1986). In these cases the retrieval cues to access stored information in memory are totally or partially unavailable. By and large, notwithstanding the different theoretical hypotheses on forgetting, a general negative appreciation related to this phenomenon can be noted.

Contrary to this evaluation, it seems that people need to forget as well as they need to remember. The ability to acquire new information is necessarily related to the lost of no longer useful old one. More tuned with this positive consideration of forgetting a recently and highly promising explanatory hypothesis indicates, for example, the possibility that forgetting is a consequence of inhibitory mechanisms activated to solve the problem of retrieval interference (Anderson, Bjork and Bjork, 1994; 2000).

The experimental paradigm generally adopted to evalutate this theoretical hypothesis is known as deliberate forgetting (Johnson, 1994; Golding and Long, 1998; Epstein, 1972; Bjork, 1972; Zacks et alii, 1996; Marks and Dulaney, 2001). It focuses on the role of forgetting as an effective cognitive mechanism in managing information that is irrelevant in order to perform a task. A series of experimental studies has been conducted on this issue (Marks and Dulaney, 2001, Oram and Mac Leod, 2001, Altmann and Gray, 2000), and the investigation on deliberate forgetting is, today, a consolidated experimental working area in the field of the cognitive psychology. The possibility to consider forgetting phenomenon as a positive cognitive mechanism in the information elaboration processes suggested us to conduct the two experiments reported in the following.

2. EXPERIMENT 1

Some studies on expertise (Chase and Simon, 1973; Chi, Glaser and Farr, 1988; Ericsson and Smith, 1991) demonstrated that experts differ from novices in the way they process information. In particular, experts are able to organise the new information into significant units and, this, in turn, can imply that experts do better than novices in performing information recalling tasks. Basing on this consideration, we decided to investigate whether a difference exists between experts and novices in analysing the list of results of a search engine and, things being so, if it also depends on the complexity of the information problem people are required to manage.

We started by considering that items presented in the list of results of a search engine might impose different demands on working memory. This depending on the nature of information problem to be solved (the number of items the subjects have to remember) and on the level of participants' expertise. In particular, we hypothesised that a converging problem will require a low cognitive effort to be performed since subjects engaged in executing it have to report, in a final recalling task, less information items from the list of results. This should produce both a higher level of elaboration by the working memory for relevant items and a high necessity to forget items evaluated as irrelevant. On the contrary, when subjects have to accomplish a divergent information problem, we expected that the problem would require more consistent cognitive elaboration to be performed: subjects engaged in solving it have to report more items, thus making consistent the activity of working memory and making the value of forgetting less relevant. In the latter case, since much must be remembered less has to be forgotten.

The experiment 1 was thus designed in order to identify whether some significant difference, in terms of remembering and forgetting effects, may be registered between two different groups of people, experts vs non experts dealing with either a converging or a diverging information searching problem.

2.1 Method

The experiment 1 was conducted using a new paradigm. Subjects were asked to perform a search activity using the interface of a search engine as Google. Two tasks were assigned to two different groups of subjects to represent two different circumstances. The first one represented a situation in which subjects needed to solve a converging information searching problem, since they had to remember just one item out of the list of results produced by the search. The second one described an unfocused situation in which participants needed to remember more items. Two scenarios which asked for performing a query were designed to implicitly contain an instruction to forget part of the materials subjects had been exposed to during the results analysis phase. In fact, in the first scenario the subject was required to find the best result that solved his/her information searching problem and this implicitly implied to forget all the others. In the second scenario the subject was required to find and remember all the results that satisfied his/her information needs, thus implicitly requiring not to forget any items but very irrelevant ones. Finally, people were acknowledged that after 10 minutes from the beginning of the experiment they had to report to the experimenter the information items they judged relevant for the information problem.

2.1.1 Subjects

Participants were selected among the undergraduate students of the University of Siena. 88 subjects were recruited and allotted to two different groups. The first one was composed by 44 students who had attended the marketing class and had passed the exam with grade A during the preceding academic year. The second one was composed by 44 students who had not yet attended a marketing class. These groups were then rearranged in order to have two experimental groups both with half experts and half non experts subjects. The two experimental groups differed in relation to the task they had to perform.

2.1.2 Materials

A simulation of Google search engine was purposely implemented for the study. Using this simpulation, each subject, after having typed in some keywords to perform the search, was exposed to the same list of result. Results were 35 items structured as they had actually been produced by Google. 5 items had been formerly evaluated by two experts (Professor of Marketing) as completely relevant to the task, 25 as partially relevant, and 5 as completely irrelevant.

2.1.3 Tasks

Subjects were given one of two possible scenarios. In the first one subjects had to pretend to be students who had to elaborate a dissertation thesis on brand management. They were required by their tutor to add to the references of their thesis the best web site concerning David Aaker. Thus they had to perform the query and, after having analysed all the results they had to report the most complete and informative site they had found.

Everything was the same in the second scenario, but the request from the tutor. In this case subjects were required to report all the sites they judged as relevant with no limits about their number.

After subjects had received one of the this two possible scenarios they were required to execute the following tasks:

- to conduct only one search on Google and to analyse all the obtained results in 10 minutes;
- to describe to the experimenter the result/results he/she judged to be relevant in solving the problem at hand. Subjects were not allowed to use pencil and paper during the execution of the task.

2.1.4 Procedure

Each of the subjects was introduced in a room and was seated in front of a computer. Then an experimenter provided a sheet of paper on which was written one of the two scenarios. Supplementary instructions were provided when necessary. Participants were free to type in whatever keywords they wanted, but they were allowed to perform the query just once. The list of results was the same for each of the four experimental conditions (expert dealing with the converging problem; expert dealing with the diverging problem; novice dealing with the converging problem; novice dealing with the diverging problem). Subjects had 10 minutes to analyse the list of results then they left the workstation and sat in front of the experimenter who, contrary to their expectations, asked them to recall all the results they could remember. In fact, they expected having to refer only the result/results they had judged to be significant to the information problem. Student could recall items by title, abstract, URL, or one of the possible combinations of these cues.

2.2 Results

Data have been analysed through a two factors ANOVA. The first one, "expertise", relates to the two levels of knowledge, experts or novices, on the problem domain. The second one, "task", indicates the two levels "converging" vs. "diverging" related to the information search activity. The first significant result is relative to the effect of expertise (F(1.84)=17.556, p<.0001) on the final recalling task. These figures indicate that experts (4.273) remember more items than novices (2.886), (see figure 1).

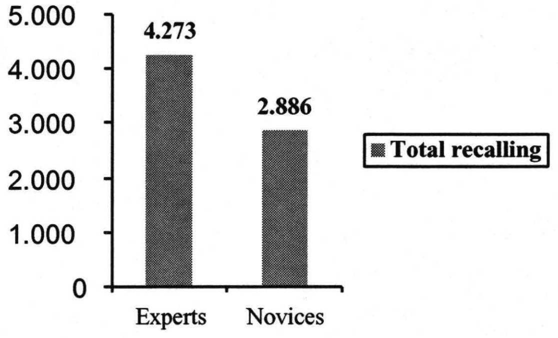


Figure 1. "Experts" vs. "Novices" total recalling means

This is consistent with previous findings about the experts ability in using working memory resources (Chase and Simon, 1973; Ericsson and Smith, 1991). The experts' ability to organise the relevant items in larger information structures allows them to maintain relevant items in working memory stores more efficiently than novices.

Also the "task" factor has a significant effect on the total recalling (F(1,847)=5.780, p<.0184). However this result (figure 2) is contrary to the experimenters' expectations, since it indicates that when just one item must be remembered subjects are able to report more items than when they are requested to remember as many items as they can.

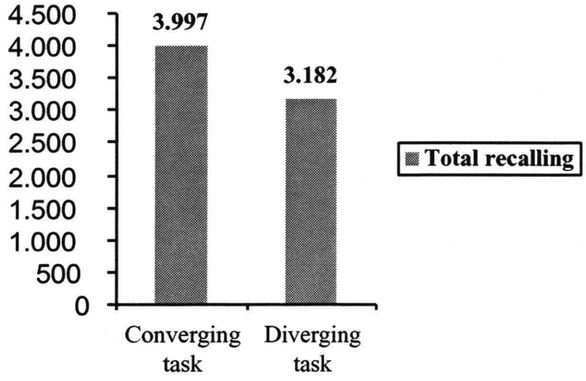


Figure 2. "Converging" vs "Diverging" total recalling means

A tentative explanation for this result may be the following. The nature of the converging information problem requires to maintain in the working memory only one useful result, i.e. the best. From a cognitive perspective this implies that the cognitive resources involved to maintain the selected items in the working memory are totally dedicated to just one item in turn. As a consequence, items judged relevant remain in the working memory with a high level of activation and, if requested, they can be referred during the final recalling task, even if they had been formerly dismissed. In this context the effects of the implicit instruction to forget does not appear as very effective.

The nature of the diverging information problem requires people to store in memory, all the information judged useful, and this, in turn, means that a high amount of cognitive resources is requested. In addition, the demand of the cognitive resources needed to carry on the comparison and evaluation activities increases as the information elaboration process goes on. This results in a competition between the cognitive resources needed to maintain information and those needed to elaborate it. The solution of this competition seems to produce a shift of part of the cognitive resources from the maintenance to the elaboration process. As a consequence, subjects engaged in solving the diverging problem recall less items than subjects engaged in solving the converging one. The forgetting phenomenon seems to be something that occurs not as an effect of a "to forget" implicit instruction but as a consequence of the nature of the task to be performed.

Another significant result obtained in this experiment is related to the effects of the interaction between task and expertise on the final recalling (F(1,84)=4.534, p<.0362), (see figure 3).

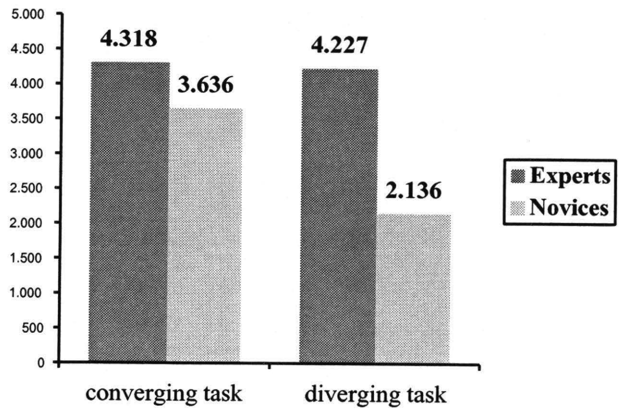


Figure 3. Recalling means for each of the experimental conditions.

More precisely, the difference between the means of recalling for the two groups of novices is significant: t(42)= 3.262, p<.0022. However the same difference is not significant for the two groups of experts. This result supports the interpretation by which the difference in the performances between experts and novices for the final recalling task does not depend on the implicit instruction "to forget", but on the different problem complexity. The explanation suggesting that people, in dealing with a complex information problem, simply shift part of the cognitive resources needed to maintain the items in memory to the processes of information elaboration, thus allowing the manifestation of some forgetting mechanism seems to be confirmed. In addition, the forgetting phenomenon seems to play a role only when people are not able to exhibit any other cognitive strategy to circumvent the limits of the working memory, as experts do. This explains why the groups of experts did not exhibit any significant difference in their performances. Finally this explanation is also supported by data about the difference between the recalling means of the experts and the novices engaged to solve the diverging problem t(42)= 4.890 p<.0001. The superiority of the experts in evaluating and comparing information items allows them to control the competition for cognitive resources in the diverging problem solving activity. Thanks to this competence experts can exhibit a better performance than novices in recalling items elaborated during the complex information problem solving activity.

3. EXPERIMENT 2

The experiment 2 has been conducted in order to verify if the results obtained in the experiment 1 were due to the experiment temporal constraints (people were informed that they had 10 minutes to find the useful results from the list presented on the search engine interface).

Subject in the converging condition had to spend some time in the matching activity necessary to compare the best item in the working memory with the one currently analysed. And this time could be considered as increasing in the diverging condition since it is reasonable to suppose that many items are in the working memory. In this condition people could not have enough time to look as many items as in the converging condition. And this could explain the difference in the results obtained comparing the means of recalling.

3.1 The hypothesis

Experiment 1 and experiment 2 differ in the final task people are required to perform: a recalling task and a recognition task respectively. This experiment was designed on the basis of the assumption that people performance on recognition tasks does not heavily depend on the depth of information elaboration, but on the exposition to the information *per se*. The hypothesis tested in experiment 2 has been elaborated considering the difference between recalling and recognition tests. If the results of the recognition task will be proportional to the results of the recalling task of the first experiment, then the recalling and forgetting effects registered in experiment 1 will probably be due to the temporal constraints imposed by the design of the experiments. On the contrary, if the second experiment will not show any significant difference in the level of recognition among the groups, then it is possible to maintain that the results of experiment 1 depend on a different level of elaboration people dedicated on each of the elements of the list.

3.2 Method

In experiment 2 the same method already adopted in experiment 1 was adopted. What changes from the first experiment to the second one is the final task subjects were asked to perform: a recalling and a recognition task respectively.

3.2.1 Subjects

80 subjects Participated in the experiment. They were selected and allotted in two groups as in experiment 1.

3.2.2 Materials

The same interface and list of results already adopted in experiment 1 were used in this experiment.

3.2.3 Tasks

The searching tasks subjects had to perform were the same as in experiment1.

3.2.4 The procedure

The procedure was the same already adopted in experiment 1. However, in this case at the end of the experiment students were asked to indicate which of the items presented on a printed list were also in the list of results they had obtained using the search engine. The printed list had 70 items, half were the old ones ad half were completely new for the subject.

3.3 Results

The 2X2 ANOVA performed on the data did not produce any significant result (see table 1).

Table 1. Recognition means for each of the four experimental conditions

	Total recognition means
Experts, converging problem	11.150
Experts, diverging problem	12.800
Novices, converging problem	11.400
Novices, diverging problem	12.750

This result confirm a general finding in cognitive psychology studies about the relative ease of recognition tasks when compared with recalling tasks. In addition, it is evident that no differences occur between experts and novices when they are engaged in a recognition task. This evidence falsifies the hypothesis that the recalling task is influenced by the experimental temporal constraints. Data about the recognition task do not have the same distribution of the data about the recalling task and this implies that the differences found in experiment 1 are not ascribable to time constraints. We explained these results referring to the processes of comparison and elaboration. In fact, in order to perform the task, it appears necessary to separate the information items in two groups, i.e. the group of elements judged relevant for the problem solution and the group of elements judged not useful and then rejected. This interpretation clarifies why people recognise more items than what they recall. The different level of elaboration received by any of the items of the list is adequate to support people performance on the recognition task, but becomes a critical feature when people need to retrieve some information that was not adequately processed by the working memory.

This explanation seems adequate also to understand why in experiment 1 we registered different performances in relation to the two factors considered, i.e. "expertise" and "task". All the results of the list presented by the search engine underwent some degree of cognitive processing, though the elements judged pertinent received a higher level of activation than the elements judged not pertinent. The higher level of activation can be thought as a more consistent reiteration activity on the selected items. As a consequence, the elements that are reiterated more consistently are not subject to a decay phenomenon. Also the items judged as not relevant are associated with a certain level of activation due to the processes of interpretation and evaluation to which their are subjected. However, since these items are either scantly or not at all reiterated, their level of activation is below the threshold necessary to recall them. In this condition, it does no seem necessary to consider deliberate forgetting as necessary to explain the results obtained here. Differences in the two tasks considered and in relation to the varying knowledge between experts and novices in experiment 1, and the lack of effects in experiments 2 are all results explainable considering different levels of elaboration inside the working memory.

4. DISCUSSION

First of all it seems possible to claim that the implicit instruction to forget, embedded within the two scenarios, did not produce any significant effect on the recalling performance when people are required to solve the converging problem. This in contrast with our expectations that people would have been able to report a very little number of information items in this circumstances. Actually, it does not appear necessary to forget items that have been evaluated as pertinent and subjected to a certain degree of reiterative activity in favour to a more promising candidate. Subjects only reiterate the best item they have encountered until the last one and, when a better candidate is found, they simply do not reiterate the former any longer.

Also the forgetting phenomenon observed for the diverging research condition, (the one in which the scenario required the participants to individualise all the information items they evaluated positively), seems to depend on the way in which cognitive resources are used. In this case there is a higher demand of cognitive processing. The problem requires to be managed through the reiteration of many items and through a consistent series of evaluations, comparisons, and dismissals. Under these conditions, forgetting seems to occur as a consequence of an excessive workload, as a side effect due working memory limits. This effect could be expressed saying that the more you have to learn the less you retain. In addition, the occurrence of the forgetting phenomenon appears more evident when the participants were not able to exhibit any efficient cognitive strategies to manage both the information overload and the concurrent need of cognitive resources to elaborate the newest information. This is what presumably happens for the group of novices involved in performing the diverging information problem in experiment 1.

Finally, the results obtained in the experiment 2 suggested that the remembering and forgetting effects found in the experiment 1 did not depend on the experimental temporal constraints. Furthermore, the remembering and forgetting effects registered in the experiment 1 seems to depend on the different degree of cognitive elaboration spent by the participants on each of the encountered information items. The different degree of cognitive elaboration, where elaboration can be considered essentially as a reiterative process, seems to cause the forgetting effects not by definitively erasing the items judged to be not relevant, but segregating the relevant items from the not relevant ones. Things evaluated as appreciable are considered and kept alive, the remainder is left to its destiny.

5. REFERENCES

[1] Altmann, E. M. & Gray, W. D. (2002). Forgetting to remember. *Psychological Science*, 13, 1, 27-33.

[2] Anderson, M. C., & Bjork, R. A. (1994). Mechanisms of inhibition in long term memory: a new taxonomy. In D. Dagenbach & T. H. Carr (Eds), *Inhibitory processes in attention, memory and language*. San Diego, Academic Press.

[3] Anderson, M. C., Bjork, E. L. & Bjork, R. A. (2000). Retrieval induced forgetting: Evidence for a recall specific mechanism. *Psychonomic Bulletin & Review*. Vol 7, No 3, 522-530.

[4] Baddeley, A.D. (1986). *Working memory*. Oxford, Oxford University Press.

[5] Bjork, R. A. (1972). Theoretical Implication of directed forgetting. In A. W. Melton & E. Martin (Es) *Coding process in human memory,* 217-235. Washington, DC: Wiston.

[6] Bower, G.H. & Mann, T. (1992), Improving recall by recoding interfering material at the time of retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 1310-1320.

[7] Chase, W. G. & Simon H. A. (1973). Perception in Chess. *Cognitive Psychology*, 4, 55-81.

[8] Chi, M. T. H., Glaser, R. & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbum.

[9] Epstein, W. (1972). Mechanisms of directed forgetting. In G. H. Bower (Ed.), *The Psychology of learning and motivation.* New York, Academic Press.

[10] Ericsson, K. A. & Smith, J. (1991). Prospects and limits of the empirical study of expertise: An introduction. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: prospects and limits*. Cambridge, England: Cambridge University Press.

[11] Goding, J. M., & Long, D. L. (1998). There's more to intentional forgetting than directed forgetting: an integrative review. In Golding J.M. & MacLeod C. M. (Eds.), *Intentional Forgetting*. London, Erlbaum Associates.

[12] Hintzman, D.L., Curran, T., E Oppy, B. (1992), Effects of similarity and repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 667-680.

[13] Johnson, H. M. (1994). Processes of successful intentional forgetting. *Psychological Bulletin*, 116, 274-292

[14] Marchionini G. (1995). *Information seeking in electronic environments*, Cambridge University Press, Cambridge, Eng..

[15] Marks W. & Dulaney C. (2001). Encoding Processes and Attentional Inhibition in Directed Forgetting. *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol 27, No. 6, 1464-1473.

[16] Navarro-Prieto R. N., Scaife M. Rogers Y. (1999). Cognitive strategies in web searching.Http://www.zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/index.html

[17] Nielsen, J. (1997). Search and you may find. Http:://www.useit.com/alertbox/9707b.html

[18] Oram, J. & Macleod, T. (2001). Remembering to forget: Modeling Inhibitory Mechanisms in Human Memory. In *Proceedings of the cognitive Science Society*, 23, 738-743.

[19] Paul, I. H. (1967). The concept of schema in memory theory. In R. R. Holt (Ed.), *Motives and thought: Psychoanalytic essays in honor of David Rapaport*. New York: International University Press.

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

[21] Rock, I. & Ceraso, J. (1964), A cognitive theory of associative learning. In C. Scheerer (Ed.), *Cognition*. New York: Harper & Row.

[22] Shneiderman B., Byrd D., Croft W. B. (1998). Sorting out of searching. *Communication of ACM*, 41, 4, 95-98.

[23] Shneiderman, B. (1997). Design information aboudant Websites: issues and recommandations. Http://www.cs.umd.edu/hcil/members/bshneiderman/ijhcs/main.html.

[24] Underwood, B. J. (1957). Interference and forgetting. *Psychologycal review*, 64, 49-60.

[25] Zacks, R. T., Radvansky G. & Hasher, L. (1996). Studies of directed forgetting in older adults. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22, 143-156.

Session 3 Human Analytical Tools

Causation patterns of accidents versus near misses: a critical review of the literature, and an empirical test in the railway domain.

Linda Wright
Formerly: CIRAS
University of Strathclyde
Glasgow, UK
lindabwright@yahoo.com

Tjerk van der Schaaf
Safety Management Group
Eindhoven University of Technology
Eindhoven, The Netherlands
t.w.v.d.schaaf@tm.tue.nl

ABSTRACT

An essential assumption for the usefulness of basing accident prevention measures on minor incidents is the common cause hypothesis: that causal pathways of near misses are similar to those of actual accidents (such as injuries and damages). The idea of a common cause hypothesis was originally proposed by Heinrich [1] in his seminal book "Industrial Accident Prevention". In this paper it is argued that the hypothesis of similarity of causes for major and minor accidents has become confounded with the interdependence of the ratio relationship between severity and frequency. This confounded view of the hypothesis has led to invalid tests of the hypothesis and erroneous conclusions. The evidence from various studies is examined and it is concluded that the hypothesis has not been properly understood or tested. Consequently, such a proper test was carried out using data from the UK railways which were analysed using the CIRAS (Confidential Incident Reporting and Analysis System) 21 cause taxonomy. The results provide qualified support for the common cause hypothesis with only 3 out of the 21 types of causes having significantly different proportions for the three consequence levels investigated: 'injury & fatality', 'damage' and 'near miss'.

Keywords: common cause hypothesis; iceberg model; ratio triangle; accident causation; near miss reporting

1. INTRODUCTION

The validity (or refuting) of the common cause hypothesis has major implications for accident prevention and analysis. If the different levels of severity really do have completely different patterns of causes, then industry has been concentrating on levels of severity (near misses, small failures) which may have little impact on the frequency of accidents which cause the greatest injuries. On the other hand, if common causal pathways can be demonstrated then a concerted effort is required to collect appropriate data (i.e. via voluntary near miss reporting schemes) and to ensure that causal analysis techniques become more widespread.

When the common cause hypothesis (the similarity of causes of major and minor accidents) is discussed there is inevitably discussion of the ratio data studies performed by Heinrich [1], Bird [2] and Skiba [3]. This very starting point is central to the way thinking about the common cause hypothesis has become focussed. Heinrich's original triangle was not intended to convince the reader of the commonality of causes between different accident outcomes, but to illustrate the fact that prevention need not wait until an accident occurred, and that prevention should not only be aimed at the most severe consequences but also to events at the lower triangle levels. In this endeavour Heinrich was successful. The ratio triangles or icebergs are used profusely in industry today. However, Heinrich did not base the common cause hypothesis upon the ratio relationship between major accidents, minor accidents and no injury accidents, although the proposed ratio relationship seemed (to him) to substantiate the idea of a common causal pathway. Today the common cause hypothesis has come to imply a ratio relationship of consequences (and not of causes). How then, did this confusion arise, and where did the common cause hypothesis spring from?

Heinrich [1] presented three inter-related concepts which demonstrated the benefits of addressing near misses (no-injury accidents) as well as actual accidents. These were the ratio triangle (showing that for every major injury there are 29 minor injuries and 300 no-injury accidents); the hidden costs of accidents (which demonstrated that hidden costs of incidents were 4 times as great as the actual costs – this was later called the 'iceberg model' by Bird [2]); and the common cause hypothesis (the *causes* of no-injury accidents were the same as the causes of major and minor injuries). The terminology has now become inextricably linked and when one talks about the common cause hypothesis or iceberg models they are often taken to be synonymous. They are however, very different aspects of the relationship between near misses and more serious incidents.

Ratio models suggest that a certain number of less serious incidents occur for every major accident or fatality. The ratio model has been tested at different times for different domains, in different countries. However, in each case a different ratio has been obtained. The ratio relationship has not been re-tested over time using a similar data set i.e. the ratio relationship has not been demonstrated to be *stable* over time. A ratio that is not stable is therefore not predicitive of the number of near misses likely to occur at a later date, and may simply be descriptive of the situation at a given time. Explicit in the ratio model is the hypothesis that the causes of frequency are the *same* as the causes of severity, i.e. reduce the number of incidents at one level of severity and the other levels should also reduce in proportion. As research into this particular topic has shown, reducing the number of incidents at one level of severity does not lead to a similar reduction at a different level of severity (Petersen [4], [5]). However, this is to be expected as the causes of severity (i.e. outcome/ consequence) can be entirely different from the causes of frequency.

This can perhaps best be illustrated by an example. A cyclist rides the same route every day and encounters a junction where he rarely meets vehicles. This being the case he does not actively check for traffic. On one journey in a hundred a car passes the junction. The causes of frequency of an accident at this junction between the cyclist and a motor vehicle are related to how often they will encounter each other. On one occasion, the cyclist and a car collide at the junction. The severity of the accident is then related to whether the cyclist wore a helmet or not, how fast the car and bicycle were travelling, the type of vehicle (e.g. a mini or a four-wheel drive with bull bars). Thus we can see that the causes of frequency and severity are not the same. However, the causes of the accident itself (no matter the outcome or the frequency) are failing to stop and check for oncoming traffic. Hopefully this example has illustrated the fact that the ratio model and the common cause hypothesis are not the same and that testing the ratio model has no implications for the common cause hypotehsis.

2. THE CONFOUNDING OF THE RATIO MODEL VS THE COMMON CAUSE MODEL

How then did these separate models become so interdependent? Heinrich [1] discusses the relevance of the ratio triangle model as providing evidence for similarity of cause of frequency and severity (i.e. reduction of the number of events at the bottom of the triangle should lead to a reduction in the number of events at the top of the triangle), and also suggests that this leads to the assumption that the causes of these events are the same. Although, Heinrich made this assumption as a (to him) logical leap from the ratio triangle, the evidence for the common cause hypothesis itself emerged from data analysis rather than simply from a deduction based on the ratio model. Unfortunately Heinrich provides little evidence of how the data were collected and analysed, although he provides a detailed description of the causal taxonomy used. The Heinrich taxonomy, although innovative at the time, is now seen as somewhat limited, assigning only one cause. Despite the fact that Heinrich's figures are convincing he did not further discuss the common cause hypothesis in subsequent editions, and finally in the final edition co-authored with Petersen and Roos [6], the common cause hypothesis is rejected. The rebuttal of the common cause hypothesis by Petersen [4], [5] and by Heinrich et al [6] which was based on the ratio relationship not holding when the number of serious injuries are reduced via prevention efforts, has further confounded the interdependence of the two models. The ratio model as evidence for a common causal pathway has become embedded in the literature and is not questioned. This is further evidenced by the way the terms 'iceberg model' (i.e. ratio triangle) and 'common cause hypothesis' are now generally used to mean the same thing. The ratio of the number of accidents and incidents occurring says nothing about the underlying causal factors – not to mention the barriers in place (e.g. Svenson [7]) or the error recovery processes that took place (e.g. Van der Schaaf & Kanse [8]) – that it is difficult to understand how the ratio relationship has become so entwined with causation.

3. LITERATURE REVIEW

Table 1 below summarises the limited literature in the area and describes the type of data used to test the common cause hypothesis, whether the authors held a view of the common cause hypothesis that rested upon the ratio model or not, and summarises the conclusions reached.

Table 1: Review of literature testing the common cause hypothesis

reference	**type of data used**	**confounded view of the iceberg model ?**	**conclusions**
Saloniemi & Oksanen [9]	Frequency data for major and minor accidents in manufacturing and construction industries.	Yes. Confuses ratio of minor to major incidents as being the same as causal mechanisms of major and minor incidents.	As ratios not in agreement with original iceberg theory, concluded that different causal mechanisms present between major and minor accidents.
Lozada-Larsen & Laughery [10]	Frequency data Comparison of the type of activity taking place prior to accident occurring e.g. during manufacture.	Yes. Basic misunderstanding of what constitutes causality. Confusion over activity being performed prior to incident and causes of incident.	Supports similar cause hypothesis, as similar tasks were undertaken in the various categories prior to incidents occurring.
Tinline & Wright[11]	Frequency data comparing the occurrence of lost time accidents and loss of containment.	Yes. Confuses ratio data for causal data.	As ratios not in agreement with original iceberg theory, concluded that different causal mechanisms present between major and minor
Shannon & Manning	Number of accident events assigned to non-lost time	Yes. Basic misunderstanding of what constitutes causality. Confusion over	As differences observed between the number of accident events (struck by; cut by etc) assigned to

[12]	and lost time accidents as assigned by victim. Data are descriptive of the type of event which resulted in the accident e.g. struck by; cut by; punctured by.	what caused the injury rather than the accident	the consequence (lost time or non-lost time) concluded different causes. No statistics performed.
Salminen, Saari, Saarela & Rasanen [13]	Finnish accident research model of 14 factors applied to 20 fatalities and 79 serious accidents.	No, although to cover all bases, the paper also examines accident type (e.g. struck by object) and part of body injured.	Results support Petersen's different causation hypothesis more than identical causation hypothesis, based on Kolmogorov-Smirnov test comparing distributions i.e. the number of causes assigned to the different levels of severity.
Kaplan, Battles & Mercer [14]	Causal factors according to the classification of MERS-TM: Technical, Human and Organisational factors for near misses and actual events.	No	Authors' state this data supports the common cause hypothesis – but only under certain severity conditions. Conservative significance level chosen.
Wright [15]	Causal factors according to CIRAS: Technical, Proximal, Intermediate and Distal for near misses and unsafe acts.	No	Results based on preliminary analysis of data and comparison graphically. Differences noted between Technical and Organisational causes between the near misses and unsafe acts.

4. A CONFOUNDED VIEW

It is apparent that frequency, severity and causal mechanism have become inextricably linked. It appears that researchers have confused the causes of severity and frequency with the causes of accidents and incidents. Thus, if a ratio is established, and the data follow the pattern of the ratio found by Heinrich [1] or Bird [2], it is suggested that the common cause hypothesis is validated. Where the ratio is invalidated i.e. severe incidents do not occur at the expected frequency when compared with minor or no injury incidents the common cause hypothesis is discounted. These positions fail to take into account the fact that the ratio model (whether validated or not) has no bearing on the common cause hypothesis: causality has no bearing on the ratio relationship propounded by the iceberg model and vice versa.

5. A PROPER TEST DEFINED

A valid test of the common cause hypothesis should be based solely on causal patterns and not ratio data. Such a test should be determined by using data that has been analysed for causal factors and not be based simply on frequencies of accident severity. Causal factors should be assigned using a valid, theoretically based taxonomy, that includes technical, human and organisational factors as a minimum. Data collected and analysed for causal factors should conform to the severity levels as outlined by previous research into the ratio models i.e. fatality, major injury, minor injury, property damage and near misses. For the most methodologically valid test, the incidents should all be investigated in the same way and to the same depth (as it is difficult to determine the effects of depth of investigation on the number of causal factors assigned). We suggest that there are three possible ways in which the common cause hypothesis can be tested:

- comparing the actual occurrence of causal codes based on a dichotomy of causal codes being either present or absent
- comparing the actual frequency of causal codes contributing to the different incident outcomes
- comparing the relative proportions of causal codes contributing to the various incident outcomes.

Testing the common cause hypothesis using causal codes as either present or absent is the weakest method of testing the hypothesis. Each time a causal code occurs more than once it still only counts as having occurred once. Hence the results conceal the actual or relative frequency of occurrence of causes. Comparing the actual frequency of causes that contribute to the severity levels may actually be inappropriate in general and lead to confounding results. This is the case where the data arising from the different consequence levels is collected and investigated in different ways. Therefore the third way was used for a proper test.

6. DATA COLLECTED

The data used in this study were collected from one railway company in the UK. The railway industry provides a rich source of data on actual incidents (such as injuries and property damage) and also has a robust system for collecting near misses (CIRAS – the Confidential Incident Reporting and Analysis System), thus making it a good candidate for this research question. The incidents used for this study were investigated in three different ways, according to the railway standards in the UK: the more serious incidents were investigated via Formal Inquiries (where a panel of experts discuss the incident and interview all the staff involved), which consist of a greater depth of investigation than either Signals Passed At Danger (SPAD) investigations or voluntary CIRAS reports. Hence it is unclear in this case whether differences in the absolute numbers of causes assigned are due to the type of incident (e.g. serious events have more contributory causes than near misses) or to the investigation procedures. However, by using proportions to test the common cause hypothesis with this data set, the problems mentioned above are thus avoided. Table 2 below shows the data source and level of severity of the incidents used.

Table 2: Severity level and data source

Severity level	**Data source**			**Total**
	Formal Inquiry	SPAD investigation	CIRAS report	
Fatality/ injury	17	0	0	**17**
Damage	18	7	0	**25**
Near miss	11	81	106	**198**
Total	**46**	**88**	**106**	**240**

7.1 CIRAS Analysis

The data were analysed according to the University of Strathclyde CIRAS human factors model which is hierarchical (see Davies et al[16] for a full description of the system). According to this model individual causal codes are subsumed under one of four top-level categories: 'Technical', 'Proximal', 'Intermediate' and 'Distal' which shall be called the 'macro' codes. These macro codes each comprise an exclusive set of individual causal codes, which shall be termed 'micro' codes. Thus the common cause hypothesis can be tested on two levels: the more general level of macro codes, and the specific level of the individual micro codes. Both levels were tested: however in this paper only the results for the macro codes will used.

7.2 Inter-rater reliability

Inter-rater reliability is a vital (and often neglected) part of any analysis system. Data analysed via the CIRAS system are subject to periodic inter-rater reliability trials. Index of concordance was above 80% for each trial. To ensure the data used in this study were also reliably coded two independent raters (experienced in using the coding scheme) coded a total of 14 incidents from various classes of event used in this study. Number of agreements between raters was 29 and the number of disagreements was 8. This resulted in an index of concordance of 78.4%.

8. RESULTS

The Figure below shows how the 4 macro causal codes are distributed over the 3 levels of severity. A Chi-square test for proportions [17] showed no significant differences for all macro codes when compared at the different severity levels.

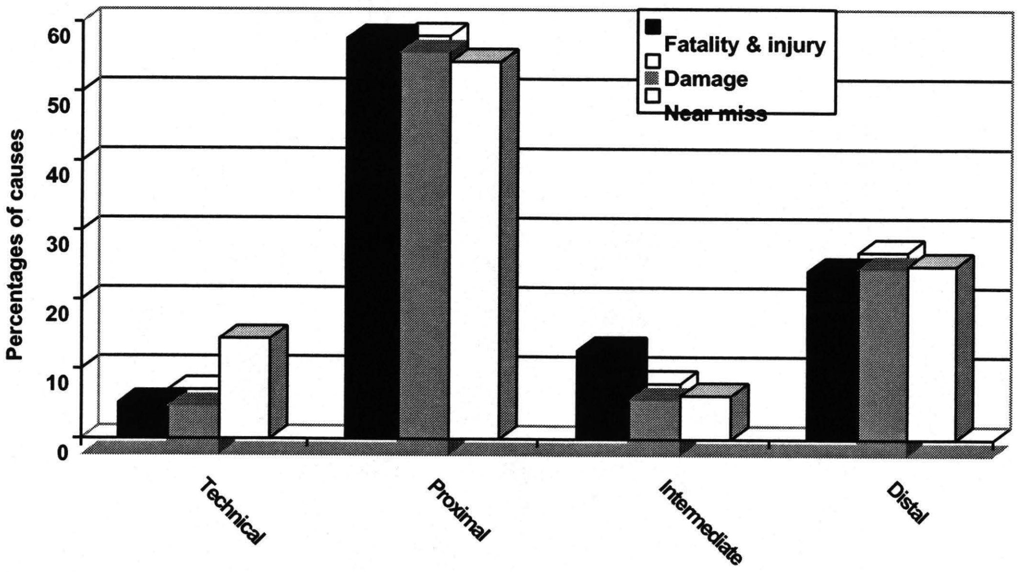


Figure 1: The distribution of macro causal codes

There are 21 individual micro codes in the CIRAS taxonomy. Of these 21 causal codes only three were significantly different.

9. CONCLUSIONS

Bearing the limitations of investigating the null-hypothesis in mind, the testing of the common cause hypothesis shows that at the general level of macro codes, there is qualified support for the common cause hypothesis as there are no significant differences between severity levels. At the more specific level of individual codes only three of 21 causal factors are significantly different. In all three cases 'fatality & injury' and 'near miss' incidents have different proportions assigned. Also in all three cases the proportions are greater for the more serious incidents. It is possible that these differences are due to real differences in causation, or that the differences are an artefact of the investigation techniques, or management priorities. Additional research to determine if these factors are identified for more serious incidents with a greater frequency than for minor events when investigation techniques remain the same would help to clarify this issue.

These findings provide qualified support for the common cause hypothesis within the railway domain. However, as this study was limited to one domain, using one type of causal taxonomy, it is recommended that further empirical tests of the common cause hypothesis be performed for a number of different domains, and with other types of taxonomies. This would provide more robust evidence of the applicability of the theory.

REFERENCES

[1] Heinrich, H. W. (1931) *Industrial accident prevention.* McGraw-Hill, New York.

[2]Bird, F. E. (1966) *Damage control.* Insurance Company of North America. Philadelphia

[3] Skiba, R. (1985) *Taschenbuch Arbeitssicherheit*, Erich Schmid VErlag, Bielefeld

[4] Petersen, D., C. (1971) *Techniques of safety management.* New York. McGraw Hill

[5] Petersen, D., C. (1978) *Techniques of safety management.* (2nd edition) New York. McGraw Hill

[6] Heinrich, H. W., Petersen, D. & Roos, N. (1980) *Industrial Accident Prevention.* (Fifth edition). McGraw-Hill, New York

[7] Svenson, O. (1991) The accident evolution and barrier function (AEB) model applied to incident analysis in the processing industries. Risk analysis, 11(3):449-507

[8] Van der Schaaf, T. W. & Kanse, L. (2000) Errors and error recovery. *In* Elzer, P. F. Kluwe, R. H. & Boussoffara, B. (Eds.), *Human error and system design and management.* (Lecture notes in control and information sciences, 253, pp. 27 – 38). London: Springer-Verlag

[9] Salomiemi, A. & Oksanen, H. (1998) Accidents and fatal accidents – some paradoxes. *Safety Science* 29 (1998) 59 – 66

[10] Lozada-Larson, S. R. & Laughery, K. R. (1987) Do identical circumstances precede major and minor injuries? In *Rising to new heights. Proceedings of the 31st annual meeting of the Human Factors Society.* New York. 1. 200 - 204

[11] Tinline, G. & Wright, M. S. (1993) Further development of an audit technique for the evaluation and management of risk. Tasks 7 & 8. Final report C2278. *A study for the Health and Safety Executive*, VROM & Norsk Hydro. London. Four Elements

[12] Shannon, H. S. & Manning, D. P. (1980) Differences between lost-time and non-lost-time industrial accidents. *Journal of Occupational Accidents*, 2 (1980), 265 – 272

[13] Salminen, S., Saari, J., Saarela, K. L. & Rasanen, T. (1992) Fatal and non-fatal occupational accidents: identical versus differential causation. *Safety Science* 15 (2) 109 – 118

[14] Kaplan, H. S., Battles, J. B., & Mercer, Q. S. (1998) The Hazard Analysis Action Decision Table, a Means to Limit Unnecessary Change in a Medical Event reporting System. In proceedings of *Enhancing Patient Safety and Reducing Errors in Healthcare,* Rancho Mirage, CA., November 8 – 10, 1008.

[15] Wright, L. (2000) Towards an empirical test of the iceberg model. *7th European Conference on Cognitive Science Approaches to Process Control*, 2000. Conference proceedings.

[16] Davies, J. B., Wright, L.B., Courtney, E. & Reid H. (2000) Confidential Incident Reporting on the UK railways: The CIRAS system. *Cognition, Technology & Work* Vol:2 Number 3, Springer-Verlag, London Limited

[17] Fleiss, J. L. (1981) *Statistical methods for rates and proportions.* New York, John Wiley & Sons

BUILDING EXPECTATIONS AND THE MANAGEMENT OF SURPRISES: STUDIES OF SAFETY-TRAIN OUTAGES AT A SWEDISH NUCLEAR POWER PLANT

Vincent Gauthereau
Quality and Human Systems Engineering, IKP
University of Linköping
SE-58183 Linköping, Sweden
vinga@ikp.liu.se

Erik Hollnagel
CSELAB, Dept. of Computer & Information Science
University of Linköping
SE-58183 Linköping, Sweden
eriho@ida.liu.se

ABSTRACT

In this paper we present results for two field studies conducted at a nuclear power plant in Sweden. Both studies were realised during safety-train outages and focus on the verification activities before the restart of the plant. More specifically, the focus of the paper is on how the plant manages the surprises that come together with operation. A few social practices are identified that seem to contribute to the ability of the plant to successfully cope with complexity. We show for instance how details about operation are first recognized and then reported to the shift supervisor. The paper also points out how the social identity of the operators plays a role in the construction of safety.

Keywords

NPP, Surprises, Expectations, Planning

1. INTRODUCTION

1.1 Operational Readiness Verification

Operational Readiness Verification (ORV) refers to the test and verification activities that are necessary to ensure that a system, such as a nuclear power plant (NPP), is able to provide its required function at the required time, by ensuring that all plant systems are in their correct functional state when the plant is restarted after an outage period. ORV is an important issue because NPPs, regularly have periods of outage and maintenance during which many systems are disconnected and disassembled. Throughout these periods the NPP is effectively off-line and the reactor is shut down. Before the NPP can be restarted and brought on-line, it is necessary to ensure that all systems are ready. This means verifying that repairs and modifications have been completed and have achieved the intended effects, that the systems have been assembled correctly, that they can be operated or activated according to their functional specifications, and that the power plant as a whole is able to function.

Over the years, deficiencies in ORV has been seen as the cause of a number of incidents, both in Sweden and in other countries [e.g. 9]. A research project was consequently started to identify the potentially problematic aspects of ORV. A literature review and interview survey at different NPPs around Sweden concluded in the need to understand better the context of work practice in NPP [3, 8].

1.2 Previous Study

A field study realized in November 2001, highlighted the importance of so-called Efficiency-Thoroughness Trade-Offs (ETTO) and of planning for ORV [7]. Human actions must always meet multiple, changing, and often conflicting criteria to performance. Humans cope with this complexity by adjusting what they do to match the current conditions. On the one hand people try to do what they are supposed to do and to be as thorough as they believe is necessary. On the other hand they try to do this as efficiently as possible, which means that they try to do it without spending unnecessary effort or wasting time. This is referred to as the Efficiency-Thoroughness Trade-Off principle [6]. The study found that the physical, social and historical context of the task were important factors influencing the ETTO on an individual level.

The essence of planning is prediction, and planning allows control to be based on feed-forward so that the anticipation of an action's outcome directs the action. This co-exists with feedback control, and striking the right balance between the two may be crucial for the efficiency and safety of an ORV. In order to understand how the organization "planned" and "conducted" the short outage, we therefore need to understand how organizational functions that are not canonically linked to "planning" (i.e. planning as defined by the organization in concern), participate to the different control layers. We needed to shift our focus from looking at "planning" as an activity in its own right, to a focus on how planning contributes to maintaining control of the ORV.

Since Perrow's [15] Normal Accident Theory, the literature concerned with safety in organizations has often looked for solutions that manage both centralization (as a solution to the problem of tight-coupling) and decentralization (as a solution to the problem of interactive complexity). A vital characteristic of High Reliability Organizations (HROs) is their capability to maintain both [16, 21]. Yet it has also been noticed that people in such organizations usually do not concern themselves with this centralization-decentralization conflict [17]. The findings in the study could be interpreted as a successful management of the two opposite: a central planning on the one hand providing coordination between individuals, and local improvisation of individuals actions on the other hand. Yet for the people involved the situation was not one of carefully balancing the two opposites, but rather one of going through a quite unproblematic process of 'planning' and 'performing'. By evoking multiple layers of simultaneous control proposed by the Extended Control Model (ECOM) [5], we have found an alternative way of looking at the initial dichotomy. In fact, having "centralization" and "decentralization" at the same time is a problem if you only have one type - or one layer - of control. But by describing the organization (or rather the 'organizing') on several layers at the same time, we can have centralization on one layer and decentralization on another without any conflicts.

1.3 Research Concern

Using the description of the organization on three simultaneous layers as a starting point, we attempt to analyse the different strategies developed at the plant for coping with the unexpected. For each of the three layers, we will identify both the canonical and the informal routines that allow the members of the organization to manage the complexity of the plant. Not only will we identify how the staff manages surprises, we will also examine how the staff builds expectations about the future. In fact, it would be quite incomplete to analyse "surprises" without understanding how the members of the organization prepare for the future since without expectations there would be no surprises.

2. THEORETICAL GROUNDS

Understanding adjustments and the management of variability on different layers, should start by a study of how the organization as a whole, and the individuals involved build up expectations. Surprises exist precisely because we usually have expectations about the "normal" course of action. The construction of this expected course of actions is both an individual and a group phenomenon. It is embedded in the physical, social and historical (cultural) context of behaviour. Not only will expectation guide the behaviour of the actors (seen either as individuals or groups), it will also guide their perception of events. This phenomenon has many names in the literature: some authors talk about tunnel vision, others about sense making [e.g. 20].

Surprises can basically be of two kinds: situational or fundamental [12], although this taxonomy has sometimes been extended further [in 23]. Weick & Sutcliffe [23] argue that high performance in HROs is a product of the successful management of the "unexpected". Studying how organizations manage the unexpected cannot be separated from the social process of building expectations. Using Weick's words [22], sensemaking is about authoring as well as interpreting, about creation as well as discovery.

2.1 Surprises

Lanir [12] differentiates between Fundamental and Situational surprises. A fundamental surprise, in contrast to a situational one, is an event that points out the incompatibility between one's perception of reality and one's environmental reality. Such surprises are, for instance, the launch of Sputnik for the US, and Prime Minister Olof Palme's murder for Sweden. Another example is the Harrisburg accident for the nuclear power industry [24]: Three-Miles Island revealed that failure could emerge from socio-technical systems and not only from purely technical factors or from pure "human error". Lanir's concept of Fundamental surprise is quite similar to Weick's cosmology episodes: "*A cosmology episode occurs when people suddenly and deeply feel that the universe in no longer a rational orderly system*" [20]. The individuals confronted by such surprises are not only surprised of the occurrence of a low probability event, but also feel deeply assailed because their previous view of the world no longer matches reality. Cosmology episodes reveal to individuals that their understanding of the world was incorrect.

Fundamental surprises put important burdens on individuals and organizations. They force them to re-conceptualise their view of the world. According to Lanir [12], the process of adjustment that follows a fundamental surprise is a reaction that flows "*over the boundaries of the event itself to include issues that have little to do with the triggering event-crisis*" [24]. Such crises can lead to fundamental learning. However, during this learning process there is often a risk of redefinition of the surprise itself from "fundamental" to "situational". Actors tend in hindsight to redefine the event as only a situational surprise. Because responding to situational surprises is much easier than responding to fundamental ones, actors tend to deny the authenticity of the fundamental surprise (Ibid.).

Situational surprises -that could be renamed to "expected surprises"- are more common. Weick & Sutcliffe [23] identify four types of surprises that are not fundamental surprises:

- "wrong direction",
- "too early",
- "duration",
- "amplitude".

The first kind of surprise is when something expected happens, but where the "*direction of the expectation*" is wrong. For instance, when out running, one usually expects his/her heartbeat to increase as fatigue sets in. And if running while tired, one would be surprised to see his/her heartbeat lower than usual.

The second kind of surprise is when something that was expected happens *too early*. For instance, one usually expects to be tired after one day at work. A person might expect to be tired around 5 o'clock. If tiredness appears around 3 o'clock instead, the person would be surprised.

The third kind is a surprising *duration* of some expected problem. A problem that was expected to be transient might turn out having enduring effects [23].

Finally the fourth kind of surprise is when the *amplitude* of the expected problem is not as expected. One example of this last kind of surprise is during shut-down of nuclear power plants. Employees are expecting to discover new problem in connection to the shut down: but while expecting approximately 30% of new work-orders they might sometime be surprised by a more important number of new maintenance tasks to be performed.

Studies in other domains than High Reliability Organizations seem to confirm the usefulness of this classification [19].

2.2 Building Expectations

We could focus our attention on surprises only, and try to find out how the organization copes with the complexity of the situation and thus with the surprising that arise in operation. Yet understanding surprising is really about understanding the construction of expectations. Without expectations, there cannot be surprises.

According to [14] expectations are "*beliefs about a future state of affairs*". Related concepts are "beliefs", "schemas" or "set". All expectations are actually derived from beliefs. Olson et al. identify three major sources of belief: direct personal experience, communication from other people (indirect experience) and other beliefs (through logical inference).

Not all beliefs and expectations have an effect on our behaviour in a similar manner. Olson et al. [14] identify four properties of expectations: certainty, accessibility, explicitness, and importance. Depending on these four properties, the impact of the expectancy on our behaviour will vary. One can clearly see that in order strongly to affect our behaviour, expectations will have to be quite certain, accessible, explicit and important. An uncertain, inexplicit and unimportant expectancy would probably have a minor effect on behaviour (even if easily accessible).

Weick [22] identifies seven properties of sensemaking:

- Sensemaking is grounded in identity construction
- Sensemaking is retrospective
- Sensemaking is enactive of sensible environments
- Sensemaking is social
- Sensemaking is ongoing
- Sensemaking is focused on and by extracted cues
- Sensemaking is driven by plausibility rather than accuracy

Most of these properties relate in one way or another to the construction of expectations. Most of them tell us something about how individuals, and organizations build expectations. Some properties are however more directly related to our current concern. Let us develop three of them.

"*Sense making is a process that is enactive of sensible environments*" [22]. While interpretation explains how people cope with '*entities that already exist*', sensemaking also explains '*how entities get there in the first place*'. Enactment is the idea that people create their own environments that in turn constrain their own actions.

Experience is ongoing: "*people are always in the middle of things, which become things when those same people focus on the past from some point beyond it*". Though experience is ongoing, interruptions, momentarily breaks of continuous flows, are important elements to understand.

The third characteristic focuses on how people extract cues from their situation and make sense of them to form an opinion on the whole situation. According to Weick the context of behaviour will affect both what a cue is, and how it will be interpreted.

2.3 Expecting the unexpected

In an environment such as a nuclear power plant, a belief that will strongly affect how people build expectations is the understanding of complexity. Perrow [15] differentiates between complex and linear interactions. When a system is interactively complex, it becomes harder for operators to understand how different subsystems relate to each other. "*Complex interactions are those in which one component can interact with one or more components outside of the normal production sequence, either by design or not*" [15]. Linear interactions are often expected, or at least visible. Complex interactions lead to unexpected sequences that are often not visible (or not directly understandable).

Expecting the unexpected is about understanding the power plant as interactively complex. Operators are, of course, trained to foresee, understand and cope with linear interactions. Running the plant is to a large extent linked to the possibility for operators to understand the linear complexity of the plant. Their competency, their knowledge of the operation sequence is vital to enable them to operate the plant efficiently.

On the other hand, operators need to be able to 'see' complex interactions. Of course fundamental surprises cannot, by definition, be expected. But operators can expect the unexpected to happen. How well operators expect the unexpected is presumably proportional to the operators' understanding of the plant complexity.

3. EMPIRICAL STUDY – METHODOLOGICAL ISSUES

3.1 The Setting

The study was realized at one reactor of a Swedish Nuclear Power Plant. The reactor, built in the early 1980s, belongs to the latest generation built in Sweden. It is a Boiling Water Reactor (BWR) that has the particularity (in comparison to older reactors) of having four independent safety-trains. Safety systems are divided into four sections of which two are necessary for plant safety. The sections (safety trains) are built as independently as possible from each other but are similar to each other: each section is basically composed of the same sub-systems, and fulfils identical safety functions. This particular design allows maintenance of the safety-trains while on a producing state (out of the three remaining safety-trains, failure in one of them would still not compromise plant safety). The reactor thus undergoes so-called safety-train outages, i.e. maintenance periods during which the plant is in a producing state. Safety-train outages allow shorter non-productive outages.

Figure 1: A view of the Main Control-Room

Work at the plant is basically organized into two main departments: the operation department, and the maintenance department. While the maintenance department takes care the maintaining the plant, on the order of the operation department, the operation department is in charge of running the plant, and ultimately in charge of plant safety. This separation into two canonical entities also reflects cultural differences, and is sometimes the source of minor conflicts at the plant. The present study focused on the operation department (and more precisely on control room personal) and on work-permit management personal who function as links between the two communities.

In the main control room work is organized in shift. Each of the seven operating shift-teams of the reactor is composed of at least six individuals. Three station technicians are in charge the different errands out in the plant (i.e. outside the control room): some systems cannot be manoeuvred from the main control room; recurrent rounds need to be performed, etc. The three other members of the shift-team are the three control-room operator. First, the turbine operator is in charge of the electrical turbine and electrical production of the reaction. Second, the reactor-operator is in charge of reactor safety. And finally, the shift supervisor has the overall responsibility. In addition to these six members, additional operators can be part of the team, as part of their education, training. Figure 2 shows the design of the main control room and the assigned position of the operators (see also, Figure 1)

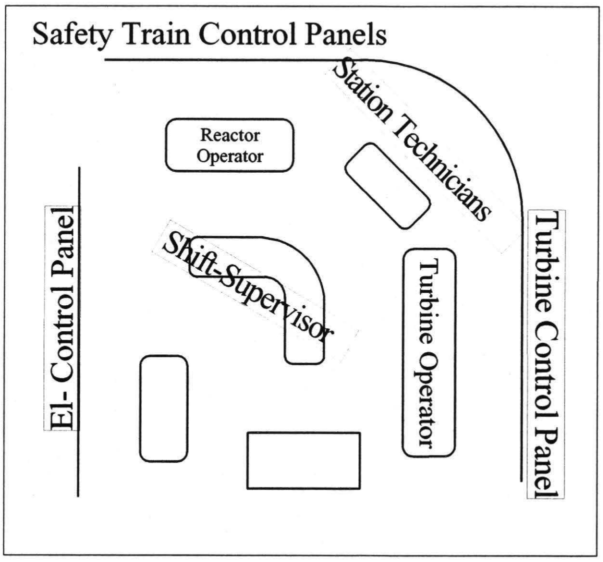


Figure 2: Schematic of the main Control-Room

3.2 Data Collection and Analysis

After a pre-study that including interviews of staff from almost every Swedish NPP (one NPP was excluded from the study because of its particular situation), a first field study was realized during the autumn 2001 at one NPP.

Further empirical material was gathered during another short observation period [see e.g. 18] at the same Swedish NPP during a safety-train outage. This study focused mainly on the main control room operators, and the supporting staff (like work-permit management staff, and planners). Two investigators realized the observations. Observations were reactive in the sense that the employees were aware of the presence of the observers [1]. They were however not too disruptive, in the sense that the observers tried to minimize their impact on the subjects. That is to say: observers did not ask the employees to describe their actions simultaneously as they go along with them. Instead a few retrospective interviews [10] were conducted. Semi-structured interviews also took place [11].

4. FINDINGS

4.1 Three Layers of Planning

Previous study [4] proposed an alternative view of planning through the ECOM model [5]. Through the analysis of an unusual event in the plant's life (the simultaneous occurrence of a short-outage and a safety-train outage), the activity of 'planning' was redefined in terms of control and broke down into three different, and simultaneous layers of control: targeting, monitoring, and regulating. (The fourth layer of the ECOM model, tracking, was beyond the scope of this study.) While the formal activities of planning were mostly related to higher layer of control (mainly targeting), lower layers of control were managed through work-order and work-permit management.

This distinction between three layers of control allowed us to understand how centralization and decentralization may happen simultaneously. While Perrow [15] denies the possibility of simultaneously managing in centralized and decentralized manners, breaking down the planning activity (understood as a control task) into three simultaneous processes enabled us to understand better how high-reliability organizations seem to successfully managed both centralization and decentralization, without perceiving any conflict [e.g. 17].

The highest layer of control referred to by [5] is denoted 'Targeting', or goal setting. Control is primarily achieved through anticipatory control. At the plant, targeting is about defining the frames for the safety-train outage. First the time frame is to be defined: both the start and the end of the outage are decided in advance. But the regulatory frame is also considered: documents are written that specify which actions have to be performed at the beginning and before the end of the outage. The targeting of safety-train outages does not have to take into account that many constraints. In contrast to non-productive outages that need, for instance, to take into account the demands of the electricity market, safety-train outages are only constrained by a maximum of sixty days per year. Sometimes, however, resource constraints affect the definition of this frame. Certain maintenance operations can only be performed by a small number of employees that may not be available during the period originally planned. In such a case, additional safety-train outages can then be planned later on. In other words, targeting is not only about taking higher level goals into account, it is also about being aware of the resources that will be needed to fulfil the goals.

Once the frame for the safety-train outage is defined, the maintenance tasks to be performed are identified and work-permits are created. Monitoring consists mainly of these two tasks. The identification of the tasks to be performed is a routine operation: safety-train outages are quite similar from one year to another. Moreover, "surprises" from previous outages are usually considered. Employees also try to stay informed of unexpected events at other plants. Employees also know that approximately one third of the activities can not be anticipated. The outage is often the occasion to discover new maintenance needs. Employees are thus prepared for having to create new work-permits during the outage itself.

Regulation is achieved through strict work-permit control. No activity is performed in the plant without a work-permit being issued. Control-Room operators are informed of all work permits delivered. Moreover, when it comes to maintenance operations on the control-systems (i.e. operations directly affecting what operators perceive in the control-room), operators ask for being informed right before the activity. Not only do maintenance technicians need a valid work-permit: they also need to inform the control-room operators personally.

It would be incorrect to understand these three layers of control as strictly consecutive to each other. Relations between the different layers are more complex than just breaking-down high level goals into sub-goals. As we saw when describing targeting: information from lower layers of control is needed in order to operate higher layer of control. The control loops are actually simultaneous.

Similarly to Clarke's study of contingency planning [2], complementary values of planning were also identified, but are hereby left aside. Instead we will now focus our attention on strategies that seem to help the plant to manage variety on the three layers of control identified previously.

4.2 "Attention to Details"

One fundamental strategy developed at the plant is not to rely on the plans. The operators say: "*it is reality that controls us*" (i.e. not the plans). During the outage, the overall plan has little influence on the work in the control room: operators instead focus on the "reality", i.e. on what gets done in the plant. While they check whether work goes as planned, they are not actively seeking to follow the plan.

One important thing for main control-room (MCR) operators to successfully follow "reality" is to be correctly informed of what goes on in the plant. Though facing control-panels representing the state of the plant, operators in the control-room are only indirectly connected to the plant itself. Information coming from station technicians thus contains central details that cannot be provided by the control-panels.

There seems to be a common understanding among the employees of the plant that paying "*attention to details*" [23] is central, and that every small detail should be reported to the MCR operators. Any small divergence from the 'normal' state should be reported to the shift supervisor. For instance, two persons from the work-permit management group were out for a short walk in the forest along the shore during lunchtime. They saw that the buoy marking the emplacement of the water inlet had been moved by significant ice-formations. As soon as they were back in the plant they reported this 'detail' to the MCR operators. Another example was a female station technician signalling to the operators the lack of ventilation in the women locker room.

What seems most interesting is not much the tendency of reporting small deviations as the sensitivity to notice deviations. Employees know the normal position of the buoy, though their work is not concerned with the buoy. In fact, every employee seems to know what the 'normal' state of the plant is supposed to be, and every employee seems to understand the importance of noticing the smallest deviations from this normal state. A question that needs to be answered is how employees decide which details are reported and which are not. In order to find that beginning of an answer, we need to comprehend how the social identity of the practitioners develop over time. It seems especially important to grasp how station technicians learn the behaviour of experts through everyday practice. Unfortunately the present work does not leave room for such an analysis.

MCR Operators are also attentive to details. Operators constantly try to know what is happening in the plant. For instance, operators can notice the activity of an emergency diesel engine close to the control-room without looking at the control-panels: there can hear / feel the engine (while the observer was unable to notice anything!). Operators constantly try to know who is where in the plant. And they usually succeed. For instance, the communication between the station technicians and the operators usually takes place via intercoms. Intercoms are fixed in the corridors / rooms of the plant. It was regularly observed operators in the control room guessing where the station technicians were in the plant and calling them on the right intercom.

4.3 Extracting cues

We previously saw that people make sense in an ongoing manner and by extracting cues from their context. Sensemaking is about interpreting cues, but first of all it is about selecting cues. Different persons will naturally extract different cues from their environment. Relating to the previous section: the different members of the operating team will pay attention to different details.

To avoid some sort of tunnel vision, the organization should benefit from a multiplicity of opinions. At the same time it seems important for the team to share a common understanding of the plant's state. With respect to that, one setting seems central. The operating team working shift, a shift meeting takes place three times a day. Each of these meeting has three phases: first the departing team gathers in order to summarise the past shift: what has happened? Then each operator meets his counterpart in the incoming team. Finally the incoming team gathers to discuss what has happened since the last time they were in charge and what is going to happen during their shift. The first and third phase look quite different from one shift team to another.

For some teams, the shift supervisor starts by telling his side of the story, and ask then each of the members of the team to complete his report. For other teams, stations technicians might first be asked to report, then supplemented by operators, and finally the shift operator completing and summarizing what the team members have said.

These two strategies have important implications for the participation of the members in the practice. While the first approach leaves peripheral participants (i.e. station technician in regard to operators and operators in regard to the shift supervisor[1]) some peripheral roles in the practice, the second approach allows peripheral participants an increased participation in the practice [13]. While the first approach spreads one perspective of the plant state, the second allows different perspectives to be shared. In other words, the second approach that supports situated learning to a wider extent than the first one, also seems to participate to safety by extending the operators' awareness of what actually goes on in the plant.

4.4 Telling Stories

In the previous sections we saw strategies for coping with unexpected events on the lowest layer of control – regulation. When it comes to monitoring, managing the unexpected seems mainly about predicting what could go wrong during the outage. We mentioned previously the common assumption that only two thirds of the work-orders of a specific outage can be prepared in advance and that the rest is usually discovered during shutdown. In fact employees seem to have a quite good idea of which systems usually pose problem.

When a problem occurs and is resolved, employees say: "now **we** know about it, but we need to update the procedures for the persons not yet working here". There seems to be a general assumption that everything that happens in the plant is automatically mediated to all the employees: "it has happened, so 'we' all know!" In other words: not only the persons present when the problem was discovered and then solved, but everyone employed. And in fact, employees seem to know!

One habit that seem to contribute to this sharing of past events is "story telling". Employees love technical puzzles and they also love to tell how puzzles were solved. One particular example seems actually quite representative. Another Swedish NPP had experienced a problem with a sensor mounted in a pipe that loosened and was push by the flow in the pipe so that it damaged the pipe further down. This particular system was not present at the plant under study, but it seemed to be on everyone's lips during a whole day. People went from meeting to meeting, telling the story to each other, drawing small schematics on white boards or on papers. Everyone seemed fascinated by the story, while at the same time everyone recognized that the same thing could not happen in the present plant because of a different design.

---

[1] Note that the conceptual description developed by Lave & Wenger is made explicit in the physicality of the main control-room (see Figure 2).

We could hypothesise that this story did not only spread because it was thrilling, that there must have been at least some unconscious utility. Such an habit of telling stories does not develop over night: it develops through years of practice, because it is useful to the organization. The present story could be seen as developing the employees' awareness of how systems can fail. It tells them about a generic failure mode, that is to say a failure mode not associated to a particular system. It increases their capacity of predicting potential failures. In cases of events that might not have been unknown, story telling might unable to the increase accessibility of this particular knowledge.

4.5 Possible Challenges to Reliability

In the previous sections we looked into two different social practices that seem to facilitate safe operation at the plant. In this section we will look at what seem to be two potential challenges to reliability. First we will try to understand the relation between the employees' social identities and their "questioning attitude". Second, we will try to further understand how employees might be "*experts*" while at the same recognizing their partial understanding of the plant's complexity.

4.5.1 Mindfulness and the social identity of expert practitioners

When relating of the employees' attention to details, we briefly mentioned that employees need to know what to report, and what not to report. Obviously, not every little detail can come up to the shift supervisor. When the shift supervisor interrogates a station technician about the completion of a procedure, he is not expecting the technician to give him every little details, but he surely wants to hear about any deviation from normality. Employees usually learn what the supervisor wants to hear through practice. Novice technicians listen to what expert-technicians report. Becoming an expert is not only about learning what to do, but also about learning to sound like an expert. One becomes an expert in the eyes of others when one talks like an expert.

At the plant, experts are expected to know what is taking place in the plant: experts know the details of operations. Expert are supposed to know the history of the components, experts are expected to know the plant's history. But experts are also expected to know the actual state of the plant. In a professional environment such as a Nuclear Power plant, being an 'expert' plays an important role in how others interpret the information one is spreading. Obviously, a shift supervisor will interpret differently the sentence "when I was in room 234, there was something out of the ordinary..." whether it comes from a novice, or an expert. While such an assertion is true in most environments, the complexity of nuclear power plants enforces further the need to interpret incoming information. In other words, once being accepted as experts, it is important for the employees keep this social status.

This importance of this social status even seems to sometimes lead to some face-keeping reactions. For instance, it was repeatedly observed individuals who would seem to acknowledge information during meetings, only to later go back to the person to ask more details. Interviewed persons repeatedly state that information meetings are not supposed to contain discussions, thus explaining why they did not ask further information. It however seems that misunderstandings occur because individuals too often presume that everyone knows everything!

4.5.2 Transforming the "unexpected" to the "expected", and the role of social identity

While reporting details of operation to the control-room operators is an important basis of safe operation, it is obviously not enough. Operators are expected to separate important information from noise, and to act upon it. Separating important information from unimportant one is however not that easy. While the operators' knowledge of the plant's normal production sequence may allow them to comprehend linear interactions, complex interactions are not easy to grasp. The operators' knowledge of the facility usually allows them to understand complex interactions due to the physical location of components. Moreover, as we saw previously how operators use stories from the plant and other facilities in order to mediate the "unexpected".

Coming back to the social identity of the expert practitioners, we could however wonder how an operator can both be seen as an expert while at the same time acknowledging his limited understanding of the plant. Rochlin [17] propose that operators in nearly failure-free organizations do not see the paradox of recognizing they are the best at what they do while at the same time recognizing that their understanding of the systems are (and will always be) incomplete.

Further analysis of our empirical data is needed in order to further comprehend this "paradox".

5. CONCLUSIONS

In the present work we identified and analysed some strategies developed to the NPP in order to cope with the complexity of operations. We saw operators trying to be as close as possible to the operations, trying to be informed of the details of the activities performed in the plant. We try to define how this "*sensitivity to operations*" [23] takes place in practice. We attempted to explain what operators define as "*it is reality that controls us*". In other words, we restrained ourselves to the lower layers of control defined in ECOM.

It is an important feature of ECOM, that the layers of controls are simultaneous. But in the present study it was not clear how the higher layers of control relate to the operations in the control room. While the time constraints are quite unimportant for safety-train outages (as the ones under study), the study of non-productive outages might reveal a fairly different side of the story. For non-productive outages, it is not only "reality" that controls the operators' work: production pressures are a reality! And while the effects of deviations on the original planning seem quite unimportant in the empirical material at hand, situations during non-productive outages might reveal a more complex balance between Efficiency and Thoroughness [6].

6. ACKNOWLEDGMENTS

The work reported here was carried out as part of the second and third tasks in the study of "Operational Readiness Verification: A Study of Safety during outage and restart of nuclear power plants" funded by the Swedish Nuclear Power Inspectorate, SKI. The authors would like to thank Per-Olof Sandén, Ola Svensson, Martin Fridleifer, and Jan-Åke Fredriksson, for valuable and inspiring discussions throughout the project, as well as staff members at the plant under study for so willingly having given their time and efforts to the study.

7. REFERENCES

[1] Bernard, H.R. Research Methods in Anthropology: Qualitative and Quantitative Approaches. AltaMira Press, Walnut Creek, CA, 1995.

[2] Clarke, L.B. Mission Improbable: Using Fantasy Documents to Tame Disaster. The University of Chicago Press, Chicago, 1999.

[3] Gauthereau, V. and Hollnagel, E., Operational Readiness verification: A study of Safety during Outage and Restart of Nuclear Power Plants. in CSAPC'01, 8th Conference on Cognitive Science Approaches to Process Control: The cognitive Work Process: Automation and Interaction, (Neubiberg, Germany, 2001), 25-32.

[4] Gauthereau, V. and Hollnagel, E. Planning, Control, and Adaptation at a Swedish Nuclear Power Plant: Stepping away from the Centralization/Decentralization Struggle.

[5] Hollnagel, E. Cognition as Control: A Pragmatic Approach to the Modelling of Joint Cognitive Systems. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans (Special Issue "Model-Based Cognitive Engineering in Complex Systems").

[6] Hollnagel, E. Of Mountains and Molehills. in Bogner, S. ed. Human Error in Medicine, to appear.

[7] Hollnagel, E. and Gauthereau, V. Operational Readiness Verification, Phase 2: A field Study at a Swedish NPP during a Productive-Outage, Statens Kärnkraft Inspektion (SKI), Stockholm, 2003.

[8] Hollnagel, E. and Gauthereau, V. Operational Readiness Verification: A study of Safety During Outage and Restart of Nuclear Power Plants, Statens Kärnkraft Inspektion (SKI), Stockholm, 2001.

[9] IAEA / NEA. Erroneous Safety System Status Control After Outage, International Incident Reporting System, 1999.

[10] Klein, G.A. Recognition-primed decisions. in Rouse, B. ed. Advances in Man-Machine System Research, JAI Press, Greenwich, Connecticut, 1989, 47-92.

[11] Kvale, S. Interviews : an introduction to qualitative research interviewing. Sage Publications, Thousand Oaks, Calif., 1996.

[12] Lanir, Z. Fundamental Surprise. Decision Research, Eugene, Oregon, 1986.

[13] Lave, J. and Wenger, E. Situated Learning: Legitimate Peripheral Participation. Cambridge University Press, Cambridge, 1991.

[14] Olson, J.M., Roese, N.J. and Zanna, M.P. Expectancies. in Higgins, E.T. and Kruglanski, A.W. eds. Social Psychology Handbook of Basic Principles, Guilford, New-York, 1996, 211-238.

[15] Perrow, C. Normal Accidents: Living With High-Risk Technologies. Basic Books, Inc., New-York, USA, 1984.

[16] Rochlin, G.I. Informal organizational networking as a crisis-avoidance strategy: US naval flight operations as a case study. Industrial Crisis Quarterly, 3 (2). 159-176.

[17] Rochlin, G.I. Safe operation as a social construct. Ergonomics, 42 (00).

[18] Schwartzman, H.B. Ethnography in Organizations, 1993.

[19] Suparamaniam, N. Renegociation of Authority in the Face of Surprises: A Qualitative Study of International Disaster Work Graduate School for Human-Machine Interaction, Division of Industrial Ergonomics, University of Linköping, Linköping, Sweden, 2003.

[20] Weick, K.E. The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster. Administrative Science Quarterly, 38 (December). 628-652.

[21] Weick, K.E. Organizational Culture as a Source of High Reliability. California Management Review, 29 (2). 112-127.

[22] Weick, K.E. Sensemaking in Organizations. SAGE Publications, Inc, Thousand Oaks, CA, 1995.

[23] Weick, K.E. and Sutcliffe, K.M. Managing the unexpected: Assuring High Performance in an Age of Complexity. Jossey-Bass, San Francisco, CA, 2001.

[24] Woods, D.D. Risk and Human Performance: Measuring the Potential for Disaster. Reliability Engineering & System Safety, 29. 387-405.

Modelling a cooperative medical decision making in an intensive care unit

L. Pellegrin
MTCD- UPRES EA 2672,
UFR de Médecine, Université de la Méditerranée
Marseille, France
liliane.pellegrin@medecine.univ-mrs.fr

N. Bonnardel
PsyCLE (EA 3273),
UFR Psychologie et Sciences de l'Education, Université de Provence
Aix en Provence, France
nathb@up.univ-mrs.fr

N. deVictor
LPC
UFR Psychologie et Sciences de l'Education, Université de Provence
Aix en Provence, France
nathaliedevictor@hotmail.com

C. Bastien
LPC
UFR Psychologie et Sciences de l'Education, Université de Provence
Aix en Provence, France
bastien@aixup.univ-aix.fr

H. Chaudet
MTCD-UPRES EA 2672,
UFR de Médecine, Université de la Méditerranée
27, bd. Jean Moulin 13385
MARSEILLE Cedex 05
Service d'Information Médicale, Hôpital Nord,,
Assistance Publique des Hôpitaux à Marseille,
MARSEILLE France
lhcp@acm.org

ABSTRACT

This research deals with modelling the collective decision-making in a specialised medical situation: the patient management of multiple trauma and neurological injuries in an intensive care unit. Our focus was first upon the activity of physicians specialized in intensive care in a French public hospital. This activity was of special interest since we hypothesized that it was based on a supervised collaboration with the other caregivers, members of the trauma resuscitation team. Our goal was to build a method useful for both observing and representing the collective activity of management, which should be re-usable by the team members in order to prepare themselves to official procedures of accreditation. The field study presented in this paper allowed a first finalization of this tool. It consists of two elements: an observational method of the medical staff activities developed upon patient management, and a specific representational method. This last one is constituted by an ontology, which describes both actors and observed events related to patient management by a temporal flowchart. The obtained results allowed us to identify specific features of this complex and time-constrained situation, especially about the strong collaborative activities between members of the patient-care team.

Keywords

Cognitive ergonomics, cognitive task analysis, medical decision-making, collaboration.

1. INTRODUCTION

The formalisation of strategies and procedures to strengthen quality measurement and improvement in medical practice is one of the major topics of interest for national agencies of health systems accreditation in various industrialized countries, such as USA or Canada. In France, accreditation goals of the national agency for evaluation and accreditation of Health Care organizations (ANAES) consist in assessing the quality and safety of care and the ability of health care organizations to ensure continuous improvement in the quality of overall patient care [1, 2]. These agencies are allowed to evaluate professional practices in all medical specialties with a focus on the analysis of performance in medical decision and practice, using for example, Evidence-Based Medicine methods, formalized by the medical community itself [3].

We planned to use the acquired experience in cognitive ergonomics about task and activity analysis methods to study decision-making in a specialised medical situation: the patient management of multiple trauma and neurological injuries in an intensive care unit (ICU). First, work situations in an intensive care unit are similar to other ones occurring in dynamic and complex environments in industry (nuclear power plans, flight regulation, etc.). This orientation is comparable to other researches in the field of anaesthesiology [4, 5]. Secondly, multiple traumas and neurological injuries are particularly representative of the kinds of medical problems that are encountered in ICU, and especially represent the safety-critical characteristics of these problems. Moreover, in our studied unit, these cases constitute a homogeneous and important set of patients.

In the project introduced in this paper, our aim was to build a method of observation and representation, describing the collective activity of management of patients in an intensive care unit. Our focus was first upon the activity of physicians specialized in intensive care in a public hospital. This activity was of special interest for us since we hypothesized that it was based on a supervised collaboration with the other caregivers, members of the trauma resuscitation team. Moreover, this method had to be usable by the unit staff (physicians and/or nurses) themselves to describe and to analyze their activities in the aim of identifying clinical adverse events, such as potential incidents, dysfunctions and near miss events during patient managements [6]. The final method had to be robust and reproducible to allow further qualitative and quantitative analyses as well as comparative studies. At the end, it would have to help to improve the quality of patients' care and the reliability of medical decisions associated with other methods as medical decisions algorithms [7, 8, 9].

This method consists of two elements: an observational method of the medical staff activities developed upon the patient management, and a specific representational method. This one is constituted of :

- An ontology, which describes both actors and observed events related to patient management,
- A temporal flowchart inspired by sequence and collaboration diagrams in Unified Modeling Language.

A first field study was conducted to acquire observation data on the medical cases resolution during the patient's management. This paper is organized as follows. First, the context of this study is characterized (section 1). Then, we present the method we used for analyzing the activity of specialized physicians (section 2) and the type of results we obtained, i.e. an ontology of the concerned domain specifying and structuring the observables, actors and actions (section 3). Finally, we describe an application of this ontology on a set of real patients' management observed in this intensive care unit (section 4).

2. POSITIONING AND METHODOLOGICAL BACKGROUND

The work introduced here is an analysis and a formal description of a situation work in line with others methods of tasks analysis and description such as Analytical Description Method (MAD), Groupware Task Analysis (GTA) or ConcurTaskTrees [10, 11, 12] but with an important restriction: the present method is not a generalist one and could only be applied to specialized medical areas. Our own task modelling approach could be more accurately classified in the descriptive task category (or "Task model 1") proposed in GTA method, than in a prescriptive task model (or Task model 2), which describes the envisioning of the new situation work incorporating a specific (and users adapted) technology [13]. However, our ultimate goal is not to introduce a new technology in a yet-existing situation but to give an adequate tool for observing and evaluating individuals' activities in a work situation. We plan to introduce a tool in the future task situation as one of possible answers from this medical unit to requirements of the national agency in terms of health care quality evaluation and accreditation.

Our aim is not to suggest the final and definitive modification of the current situation to a new one but a recurrent adaptation of the current situation to new constraints:

- To satisfy the required official specifications, tasks performed by the actors will be regularly evaluated (every 5 years) and, if necessary, re-designed to fit these recommendations in new procedures.
- Technological aspects, which contribute to the complexity of the task' environment will be also regularly modified (new drugs recommendations, new materials),

Thus, the main characteristics of such tasks are the activity of decisions and actions undertaken upon the patient by the all the experts (physicians and nurses) in a non-static professional environment.

The resuscitations' activities in patients care units present several characteristics:

- A collaborative work, implying several caregivers (with or out of the patient's bed): physicians of different medical specialties (resuscitators, neurosurgeons, radiologists, orthopedic surgeons), nurses and other contributors (fire brigade, traffic police, emergency services, mobile accident units);
- Changes in patients' health states, with major risks implying temporal constraints;
- Complex patient charts due to multiple traumatisms based upon numerous indicators, each of them with specific constraints of evolution.

We focused our attention on collaborative aspects between physicians and other contributors during patient management time period.

Medical Intensive Care is a situation described as a coordinating heterogeneous work where collaboration occurs between various kinds of practitioners evoked above [14]. This work situation is heterogeneous because each actor is focalised upon a single patient with different activities, motivations and concerns. The main goal that is asked for this medical practice is to stabilize the vital functions of the patient. The actors need different kinds of information elements extracted from different sources:

- The patient through direct observation,
- Medical data obtained in situ or from other services (mainly medical imagery and biological analysis) by physicians or nurses,
- Preliminary information about the accident scenario and the injury state of the patient transmitted by the mobile units,
- On-time vital data from various monitors.

Moreover, the major temporal constraints and complexity of the cases to manage involve that individuals cannot carry out the required tasks. As in other medical situations, both individual and collective actions have to be performed according to shared or even independent goals. Such a team work could be described as a collective task for the following reasons: some tasks would be dependent or interdependent or different from others ones and their role in the global resolution would be different. In this context, we hope to identify elements of the situation actors' activity.

3. DESCRIPTION OF THIS STUDY

This method is made up two phases: data collection then formalization of the results. The first phase of our work consisted in collecting, during a 4 months' first study, observational data in situ in the anesthesia-resuscitation department. Towards this end, a direct observation grid was built to collect actions, decision-making, communications between caregivers, and explicit information needs. The data we gathered dealt with the management of 12 cases of neurological and multiple traumas occurring during this study. These observations were completed by semi-directive interviews based upon MAD methods in post-task verbalizations from the physicians about the observed cases and their own perception about their work.

The second phase consisted in developing a formal representation of observed cases based upon a first version of an ontology describing situation objects and actors. This formalization of the observed domain will allow us to represent, in a final representation sheet, the main elements of this activity. We underline that it is based on the resuscitators' point of view, even if other caregivers were implied in the resolution situation.

Moreover, our final goal is to propose this tool to the members of the department to allow them to describe themselves their collective activity in cases of patient management. A second study was conducted during last year (2002-2003) to specify the modalities of this method.

4. ELABORATION OF THE METHOD

4.1 An oriented-domain ontology of actors and objects

The ontology we propose belongs to the "information ontologies", according to Falasconi and Stephanelli's classification [15]. An information ontology specifies the organization of observables and actions belonging to the considered domain. Applied our specific medical domain, observables may include clinical findings from the patient (e.g. the intracranial pressure measures, the "Glasgow coma scale score") and actions which can be performed to evaluate and change the patient's health status, e.g the intracranial pressure monitoring device and treatment, the radiological and scanner examinations or sedation protocols.

Moreover, our ontology was built essentially to represent the task world defining relevant concepts and relationships for the purpose of a task analysis. According to this point of view, two meta-concepts were defined to structure the formal description of the observed task situation called "described objects" and "descriptive objects".

- The "described objects" level includes observable elements extracted from the work situation. It comprises 3 classes as actors, time and spatial localization. The "actors" class includes all animated or inanimated objects belonging to the activity world and having a role to play during the patient management as the "human agents" category (patient, physicians, nurses and so on), and "paraclinic information sources" (monitors and scopes, radiographies and scanners, the patient medical record, biological examinations). This part of the ontology could be extended, if necessary, by the adjunction of medical knowledge as medical acts and diagnosis classifications (WHO International Disease Classification or the reanimation and resuscitation scores as TISS-Therapeutic Intervention Scoring System). The time class is defined by a temporal axis with minutes and seconds. The localisation indicates where are located the patient and the others actors in possible medical services: the intensive care unit, the scanner and radiology unit, the out-of-hospital setting and the others units of final patient's delegation.
- The "descriptive objects" introduce a domain-independent level to define generic concepts necessary to build the final representation sheet of the patient management tasks. The main classes were the sheet structure (horizontal axis represents actors' actions and communications and vertical axes present time and localization as well as each actor implied in the situation resolution), the graphics forms and the labelled texts. Graphics forms include several subclasses of descriptive objects whose roles are to qualify the observations as actions, information communications or explicit decisions between collaborative agents (axes, arrows, segments and columns). Each of them is decomposed into descriptive sub-classes of objects. The labelled text describes observations of actions from the actors, time points and patient localisations. A first dictionary of these labels was proposed, related to the "described concepts" of the ontology.

All the elements of this ontology will be used to build the ultimate "patient management observational sheet".

4.2 Application of the method on a set of patient management cases

During the first study, which will be introduced here, this method was applied on 12 observed patient management cases. The main features of theses cases, introduced the table 1, are the following ones:

- the patients' various ages (from 19 to 91 years old),
- injuries causes consisting frequently in road accidents,
- the identification of multiple and neurological traumas, at the patients' admissions in the service.

The final representation highlights on a single sheet the information transmissions, actions upon the patient, information needs and explicit decisions of actions. The graphical principles are based upon UML sequence diagrams with two main dimensions: a vertical dimension representing time and localisation, an horizontal dimension representing different objects [16, 17]. This representation is task-centred, described by predefined labels on segments linking between resuscitators and radiologists (or others specialists), or others agents (the crew, others specialized physicians and the patient himself). This representation is also temporal-oriented to obtain step outlines in actions sequences and tasks scheduling.

Table 1. Characteristics of the 12 observed patients' managements

Year 2001	Age	Trauma cause	Delegation Unit
Case 1	30	Traffic Accident	Deceased in the Unit
Case 2	51	Traffic Accident	Operating Block
Case 3	86	Traffic Accident	Referral error/Orthopedic Unit
Case 4	55	Falling rocks	Orthopedic Unit
Case 5	43	Convulsion crisis, falling.	Operating Block
Case 6	82	Stairs falling	Deceased in Scan Unit
Case 7	20	Traffic Accident	No Delegation
Case 8	19	Traffic Accident	No Delegation
Case 9	58	Traffic Accident	Orthopedic Unit
Case10	24	"Defenestration"	Deceased in the Unit
Case11	31	Traffic Accident	Operating Block
Case12	75	Airplane crash	Orthopedic Unit

Fundamental elements of the patient management activity may be classified into 4 main categories:

- Information transmissions between actors (oriented or collective),

- Undertaken actions to the patient,
- Information acquisition (by the way of monitors or radiological results),
- Explicit decisions of actions (as decisions of examination or transfer) (cf. figure 1).

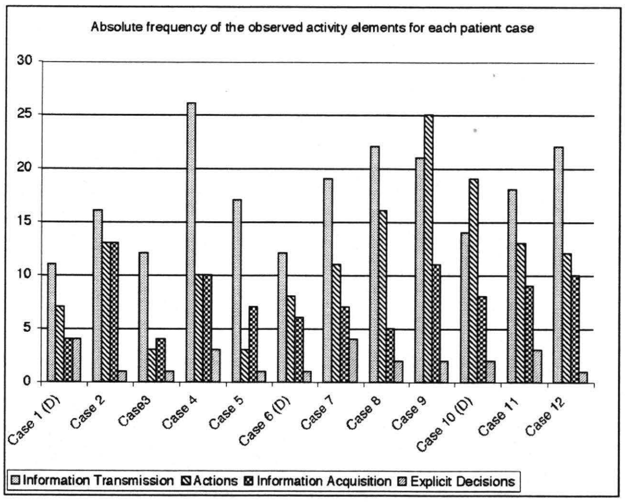


Figure 1. Frequency of observations in the four predefined categories of activities

If we group all observations into these categories, we observe that information transmissions between reanimation physicians and other health-care providers constitute the most frequently observed activity (mean: 17,5), the actions performed upon the patient are in second place (mean: 11,7), observed and explicit information needs are less important (mean: 7,8) and the explicit decision for actions are the less numerous (mean: 2,1) as it is shown in the figure 1. If we group all information management (transmission and acquisition), these observations are congruent with another publication illustrating the importance of information needs to patient management by the medical team in a Surgical Intensive Care Unit [18].

Based on the data represented in the ultimate "patient management observational sheet", we identified elements of collaborative information management between the medical team members (clinicians, nurses and outside medical consultants) and interaction of information management and the undertaken actions. Each sheet illustrates more explicitly the elements of this distributed support for decisions for the case resolution. Especially we observe this cooperative activity for specific tasks as comments upon radiological results, generally done between resuscitators and radiologists or others specialists.

We present below presented extracts from two cases corresponding to different aspects of a common patient management between the team's members. Each of these cases is illustrated by a figure showing an extract of the final sheet.

In the first example, the patient is a 19 years old man sent for a traffic accident, (falling in the river and cutting free from vehicle) by the SAMU at 19h30 (emergency medical units). The resident supervises the initial management (Total time patient management: 3h40). In this case, the team has to solve different problems before sending the patient to the scanner, which the main goal to reach in accordance with the medical decision protocol. With the elements of the current situation, the "supervisor" physician main task is to determine the delay time before sending the patient to the scanner unit and to prepare him to his transfer and examination. In the same time, he performs several decision-making and actions with the non-members unit physicians and nurses to stabilize the patient.

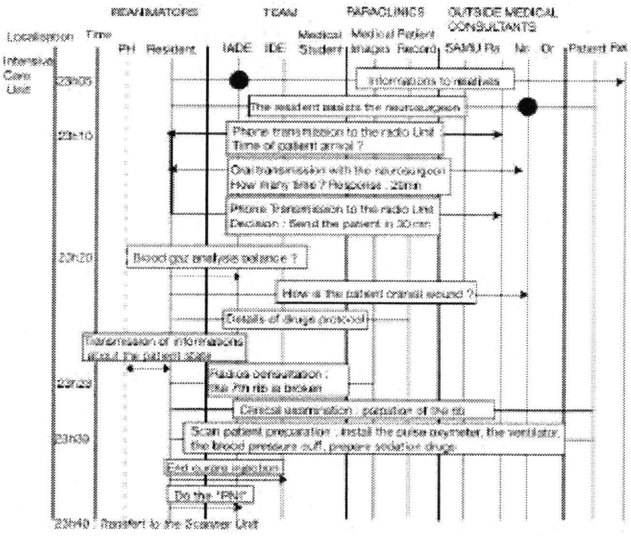


Figure 2. Patient management (case n°8)

In the second example, the patient is a 24 years old man, addressed by the SAMU 13, for having fallen out of a window in the second floor. The inpatient physician supervises initial management assisted by two residents (Total time patient management: 0h55). In this case, the vital status of the patient is critical and only the unit's staff (physicians and nurses) performs numerous and recurrent actions and information managements. In this example, the density of actions is more important than in the previous example. The main goal is the stabilization of the patient and it is expressed by a main emergency medical act, cardiac massages, performed successively by the physicians. The supervisor leads the undertaken actions, investigates information upon the patient state evolution (transmission with the SAMU physician, observe patient monitors, performs clinical examinations) but gives little explicit recommendations to the others crew' members while acting himself. This sequence of actions illustrates the workload distribution between the physicians with parallel, successive and alternated actions and information seeks during a short timestamp, especially between the anaesthetist senior and the residents.

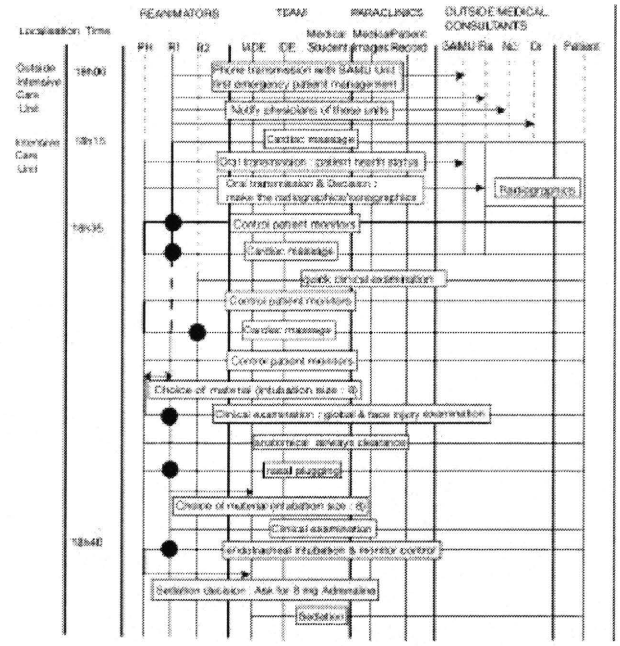


Figure 3. Patient management (case n°10)

These first description shows that, even if the trauma patient resuscitation management is not a fully distributed decision, actions and decisions of responsible physicians are based upon an active collaboration and information transmissions between all participants as it was observed in previous studies [19, 20].

Some problems emerged from these first results. They are related to the quality of final sheet' representation. The lack of visibility to differentiate actions from information transmissions and decisions has lead us to represent explicitly these classes of activity in the sheet by different forms :

- Therapeutics and medical acts (old "actions" category) are represented by ovals,
- Clinical and monitoring investigations (old "information acquisition" category) are represented by round corner rectangles,
- Information communication (all kind of information transmission between actors belonging to old "information acquisition" category, which are not clinical or monitoring or radiological investigations, and old "information transmission" category) by rectangles.
- Explicit and declared decisions of actions are represented by rhombs.

These new classes of observations can be compared with the some of the supra categories of Manser & Wehner (ibid.) in the anaesthesia field: monitoring, manual task (or measures) and communication. In the same way, other elements of description were simplified and we introduced, in a representational formalism inspired by the package notion in UML, 4 elements describing the main procedures of the official decision protocol: patient arrival in the unit, examinations and medical acts to evaluate the patient health status and stabilize the patient, sending the patient to the scanner for more precise evaluation and finally delegating the patient to the adequate unit.

The final result was applied upon the patient management cases observed during the 2002-2003 session. The new descriptive objects of our ontology are introduced in figure 4 by an extract of another patient management: a 48 years old man sent to the UCI by the SAMU 13, for a traffic accident with multiple and trauma injuries. This new final sheet introduced in the following example (figure 4) is actually a work-in process validation.

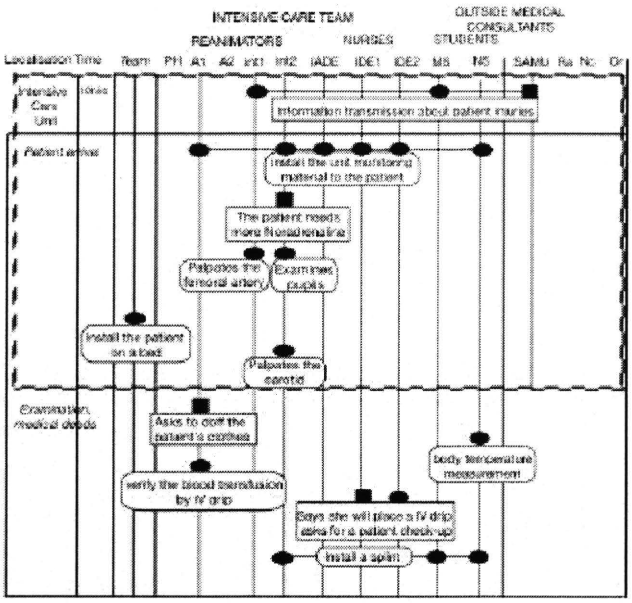


Figure 4. Application of the new formalism upon a patient management extract

5. Conclusion

Thanks to the observation sheet, it would be possible to describe some features of a cooperative activity in this situation [22].

- This activity concerns both a set of individual performances of the tasks due to the implication of several medical specialties and management of all the required tasks by the actors belonging to the team (cooperation in action).
- The case of cooperative activity observed here could be described as a supervised cooperation implying planning and allocation of the future tasks and actions between staff members (cooperation in planning).
- Each member of the medical team (physicians and nurses) have knowledge of the common tasks to perform and the common goal to reach which allows them to interfere with the others actors' procedures.

The definition of cooperation in cognitive ergonomics expressed by F. Decortis [23] is applicable on the work situation. We observed during this study the jointly performance of tasks by a set of caregivers "according to some objectives and on the basis of a mutual and distributed understanding of intentions and actions".

By the way of the final "patient management observational sheet", some elements of this cooperation are outlined:

- The actors jointly perform a set of structured and complex processes.
- The allocation of some tasks to the in-present actors are different (i.e. parallel or successive actions) according to the complexity and the emergency of the situation.

- In the case of critical situations, shared knowledge and a mutual and distributed understanding of intentions and actions between actors are observed by the few of information transmissions between them and an increase of mutual actions undertaken upon the patient.

In the same way, this cooperative activity is occurring upon a spatial and temporal axis where the actors could be spatially distributed (in the unit or outside in others places), with interactions between them through verbal communication via the telephone or these agents cooperate in the same place, near the patient by the way of collaborative actions and communications. Moreover, the time pressure has some consequences upon the global management by the agents of the successive or simultaneous tasks to perform. The physician, responsible for the patient management, supervises coordinated and synchronized actions of the team, even if they should be heterogeneous.

In the future, we plan to extend a final version of the model activity tool with respect to the following specifications:

1) to have an explicit representation of the components of physician activity during the patient case management.
2) to be used by members of the medical staff to prepare the official evaluation and accreditation procedures by the national agency ANAES.

In addition, the ultimate goal of the representation sheet is to compare the observed patient management to the prescribed medical recommendations in clinical guidelines and decision algorithm of this medical specialty, to highlight the difficulty of cases management due to diagnosis severity, complexity of the situation and time constraints. [24, 25].

6. ACKNOWLEDGMENTS

Thanks to all members of the Intensive Care Unit in the Hôpital Nord (Assistance Publique de Marseille) especially Pr. C. Martin and Dr. J. Albanese for their fully participation to this project. We want also to underline the personal implication of the ergonomist' students, V. Murillo C. Dobigny and C. Boureau, who have collaborate to this project. Thanks also for Dr. Elie Sanchez who has carefully reviewed this paper.

This project ("Analysis and Evaluation of the activity of medical decision-making in a dynamic situation") was granted by a national research program: the Hospital Clinical Research Program (PHRC-2000) from French Heath Department for the years 2000-2002 and supervised by Dr. H. Chaudet, Medical Information Service (Hôpital Nord-Assistance Publique de Marseille) with the collaboration of UPRES EA 2672, faculty of medicine (University of the Mediterranean) and the laboratories belonging to UFR Psychologie et Sciences de l'Education, University of Provence).

7. References

1. Agence Nationale d'Accréditation et d'Evaluation en Santé. *Manuel d'accréditation des établissements de santé*, ANAES Paris, 1999.
2. Agence Nationale d'Accréditation et d'Évaluation en Santé. *Principes méthodologiques pour la gestion des risques en établissement de santé.* ANAES Paris, 2003.
3. Sackett, D. Rosemberg, W. Muir Gray, JA. Haynes, BR. Richardson SW. Evidence-Based Medicine: What it is and what it isn't. *British Medical Journal*, 312, (1996), 71-72.
4. Xiao, Y. Hunter, W. Mackenzie, C. Jefferies, N. & the LOTAS GROUP. Task complexity in Emergency Medical Care and its implications for team coordination, *Human Factors, 38,4*, (1996), 636-645.
5. Gaba, D.M., Howard, S.K. Situation awareness in Anesthesioloy, *Human Factors, 37,1*, (1995), 20-31.
6. Busse, D.K., Johnson, CW. Identification and Analysis of Incidents in Complex, Medical Environments, in Johnson C.W. eds, *First Workshop on Human Error and Clinical Systems; Glasgow*, UK (April 1999), 101-120.
7. Albanese J, Arnaud S. Traumatisme crânien chez le polytraumatisé. In Sfar, eds. *Conférences d'actualisation. 41ème congrès d'anesthésie et de réanimation.* (Elsevier Paris 1999), 737-63.
8. Bullock, R., Chestnut, RM., Clifton, G. Guidelines for the management of severe head injury. *Journal of Neurotrauma*, 13 (1996) 643-734,
9. Bullock, R. & Joint Section on Trauma and Critical Care of the American Association of Neurological Surgeons and the Brain Trauma Foundation. *Guidelines for the Management of Severe HeadInjury*. American Association of Neurological Surgeons, Park Ridge Ill, 2000.
10. Sebillotte, S. *Méthodologie pratique d'analyse de la tâche en vue de l'extraction de caractéristiques pertinentes pour la conception d'interfaces*, Rapport technique N° 163. INRIA, Rocquencourt, France 1994.
11. Paterno, F. *Model-based Design and Evaluation of Interactive Applications*, Springer Verlag eds, Heidelberg, 1999.
12. Van Welie, M., Van der Veer, GC., Eliens, A. Ontology for Task World Models, in *5th International Eurographics Workshop on Design Specification and Verification of Interactive Systems DSV-IS98*, 3-5 June 1998, Abingdon, UK, (1998), 57-70.
13. Van Welie, M., Van der Veer, GC. Ontologies and Methods in Interdisciplinary Design, In *Computer Science Education: Challenges for the New Millenium, Casa Cartii de Stiinta, Cluj*, (1999), 143-158.
14. Reddy M.C., Dourish P., Pratt W. Coordinating Heterogeneous Work Information and Representation in Medical Care, In *Proc. of 7th the European Conference on Computer Supported Cooperative Work -ECSCW'01* (Bonn, Germany, September 2001), 239-258.
15. Falasconi, S., Stefanelli, M. A library of medical ontologies. In Mars, NJI (eds), *Proceedings of the ECAI94 Workshop Comparison of Implemented Ontologies*, (Amsterdam, Holland May 1994), 81-91.
16. Object Management Group Inc. *OMG Unified Modeling Language Specification,* Version 1.5 formal/03-03-01, March 2003.

17. Rumbaugh J., Jacobson I., Booch G. *The Unified Modelling Language Reference Manual*, Addison Wesley Longman, Inc., Reading, 1999.

18. Reddy, M., W. Pratt, P. Dourish, and M.M. Shabot.. Asking Questions: Information Needs in a Surgical Intensive Care Unit. In *Proceedings of American Medical Informatics Association Fall Symposium (AMIA'02).* (San Antonio TX. November 2002). 657-651.

19. Cicourel, A.V. The Integration of Distributed Knowledge in Collaborative Medical Diagnosis, in *Galegher, J., Kraut, R.E., Egido C., Eds, Intellectual Teamwork,* Lawrence Erlbaum Associates, Hillsdale, NJ., (1990) 221-242.

20. Grosjean, M., Lacoste, M. *Communication et Intelligence Collective: le travail à l'hôpital*, Coll. "Le travail humain" Presses Universitaires de France, Paris, 1999.

21. Manser, T., Wehner, T. Analysing action sequences: Variations in Action Density in the Administration of Anaesthesia, *Cognition, Technology & Work*, 4, (2002), 71-81

22. Hoc, JM., Carlier, X. Role of common frame of reference in cognitive cooperation: sharing tasks between agents in air traffic control, *Cognition, Technology & Work,* 4, (2002), 37-47.

23. Decortis, F. Noirfalise, S., Saudelli, B. *Cognitive ergonomics as framework for cooperative,* http://www-sv.cict.fr/cotcos/pjs/Theoretical Approaches/DistributedCog/DistCognitionpaperDecortis.htm, 2000.

24. Sougné J., Nyssen A.S., De Keyser V. Temporal reasoning and reasoning theories a case study in anesthesiology, *Psychologica Belgica*, 33, (1993) 311-328.

25. Xiao, Y., Milgram P., Doyle DJ. Planning Behavior and Its Functional Role in the Interaction with Complex Systems. *IEEE Trans. on Syst., Man, and Cybern., Part A: Systems and Humans*, 27,3, (1997), 313-324.

Individual differences and distributed cognition: the case of troubleshooting diagnosis

Julien Cegarra
CNRS – University of Nantes
IRCCyN, PsyCoTec, B.P. 92101
F-44321 NANTES CEDEX 3

Julien.Cegarra@irccyn.ec-nantes.fr

Jean-Michel Hoc
CNRS – University of Nantes
IRCCyN, PsyCoTec, B.P. 92101
F-44321 NANTES CEDEX 3

Jean-Michel.Hoc@irccyn.ec-nantes.fr

ABSTRACT

When one experiences a telephone breakdown, one option is to contact the telephone company's customer service line. In French telephone centers, an increase in the number of such calls has led the CNRS engineering laboratory at Valenciennes (LAMIH) to design a tool to automate some of the activities of the service's operators. An experiment carried out to validate this tool, using expert operators, has shown one unexpected result: that there is an important variability in performance quality, from 20% to 90% of successful scenarios. Initially, it is possible to see this result as being due to differences in skill levels. In this study, however, we will show that some features of skill strategies (functional and structural representations, low cognitive workload) are not always associated with the best performances. We will then use the concept of *cognitive compromise* [1] to explain these results. Cognitive compromise aims to produce an acceptable performance whilst at the same time safeguarding cognitive resource. Moreover, we will show that differences in the way this cognitive compromise is managed can lead to differences in performance, with some operators preserving their resources whilst others boost their efficiency. Finally, we will conclude with the need to take individual differences into account when designing human-computer interactions.

Keywords

Diagnosis activities, strategies, individual differences, cognitive compromise.

1. INTRODUCTION

In the Seventies, the development of computer science opened up a whole new field of research within Human-Computer Interaction (HCI): Cognitive Ergonomics. Whilst it is still part of HCI today, it cannot, however, be entirely constrained within this field. Indeed, the introduction of Cognitive Psychology of Work concerns has broadened its interests to include interactions between humans and their cognitive work environment [9]. In a similar way, the theory of distributed cognition has shown that cognitive interests could apply to a broader unit of analysis than the individual, as, for example, in the aeroplane cockpit [17].

In Cognitive Ergonomics, diagnosis has become a significant issue during recent years, particularly where there are multiple interests at stake: health interests for patients in the medical field, financial interests relating to the cost of errors made, and finally, legal interests due to operators' responsibilities. However, diagnosis activities are still often reduced to nothing more than comprehension activities, with no links being made between these activities and the underlying decision-making action. Moreover, studies are still often restricted to an individualized cognition, without taking into account the importance of the resources and materials in the environment.

In this paper, a study of distributed cognition between humans and machines within the framework of telephone service troubleshooting will enable us to study the human strategies in diagnosis.

2. TROUBLESHOOTING DIAGNOSIS

In France, when customers encounter a breakdown in the telephone service (for example a line cut due to bad weather), they can contact their telephone company's customer service line (e.g. France Telecom). The operators of this service have to identify the origin of the problem, which then makes it possible to determine if it is necessary to send a technician straightaway. Because the operators have to deal with a large number of calls, and because they must carry out their diagnosis within a limited period of time to avoid saturating the telephone center, it becomes necessary to design a tool to automate part of their activities.

The CNRS engineering laboratory at Valenciennes (LAMIH) designed a tool to assist these operators. The internal model of the tool associates 49 principal failures (syndromes) with more than 500 breakdowns (symptoms) in a hierarchical form according to various representations (functional, topographic, structural, etc.). The screen displays all the possible breakdowns at any one moment according to these various representations and the operator can validate or invalidate a breakdown, in this way reducing the search space which is automatically updated.

To evaluate the contribution of this new distribution of diagnosis activities, an experiment was carried out to test ten operators, considered as experts by their peers, in fifteen successive scenarios. An engineer playing the role of the customer exposed a problem and then used a document that gave answers to the questions that the operators asked. The results of this experiment indicate that when both human and machine contribute to performance, the rate of good diagnosis is 82% on average against 64% on average for an expert alone. The results are not homogeneous, however, with the performance of operators varied from 20% to 90% of successful scenarios.

With this type of result, distributed cognition remains a more descriptive than explanatory approach. As a consequence, it is appropriate to study more precisely the diagnosis activities from a human viewpoint.

3. DIAGNOSIS ACTIVITIES

Originally introduced in the field of medicine, the concept of diagnosis, in its common definition, concerns the process of comparing symptoms to a syndrome. If the use of the term is extended to different fields (process control, debugging, etc.), several common characteristics can be identified from one diagnosis situation to another [15]:

- First, the diagnosis integrates an activity of comprehension, to organize elements into a meaningful structure.
- This organization is oriented by the operator towards decisions relevant to actions. The diagnosis is thus finalized by an action decision.
- An action consists of organizing the elements in accordance to the objectives. The operator orients the diagnosis towards an acceptable performance according to the objectives whilst preserving a satisfactory comprehension of the situation.

From these three points, one can retain the general definition of diagnosis by Hoc [10], who considers it to be "a comprehension activity relevant to an action decision". And this comprehension activity can be carried out because the operator builds a representation of the situation that enables him/her to determine the state of an object, even if information relates to elements that are not directly observable. Hoc and Amalberti [12] put forward the term *current representation*, which is compatible with the diagnosis as previously defined, and includes three points: a representation of the process and its goal, a representation of possible actions and a representation of the available resources. The content of this representation is, however, not uniform and can be presented in various formats.

3.1 Types of representations

Rasmussen [23] [24], in his work on diagnosis, suggested a classification of diagnosis strategies that could be decomposed into two main categories [15]:

- Strategies guided by the characteristics of the correct functioning of the installation, based on the physical structure for a *topographic search*, or based on a more abstract representation for a *functional search*.
- Strategies guided by the characteristics of abnormal functioning (Rasmussen's term is *symptomatic strategies*); for example, the search by hypothesis-testing guided by knowledge of syndromes.

These categories come from several studies, however, and as Bainbridge [5] noted, they are sometimes ambiguous, mixing various aspects. It is, then, possible to make a finer distinction according to their contribution to the representation [15]:

- Type of representation of the device [24]. At a topographic level, the operator considers the physical form and the configuration of the variables, such as apparatus and plugs. On a functional level, the operator works with more abstract values, such as providing the user with a dial tone.
- The type of knowledge of the device [24]. Knowledge can relate to the correct functioning of the device or to its breakdowns and failures; in other words, abnormal functioning.
- Representation precision [22]. Representations can relate to structural knowledge (syndromes) or to factual knowledge (symptoms).

3.2 Internal and external factors of performance

As Roth and Woods [25] noted, human performance varies according to the match between cognitive demands and the capacities of the agent (and his/her support). The difficulty in meeting cognitive demands depends on the current representation. Skills and task complexity could then be considered as external and internal factors of performance.

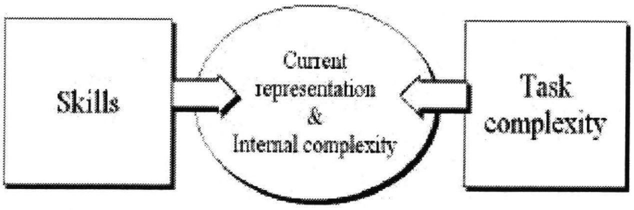


Figure 1. Internal complexity is determined by skills and task complexity.

There is an established school of though that considers internal complexity (cognitive workload) to be determined by skills in conjunction with external complexity (task demand) (e.g. [29]), as presented in Figure 1. But because these ideas have different definitions, we have to define them more clearly.

Skills, developed by experience, allow operators to get a better representation of the situation [6]. We could define a skill as the effective use of an adapted behaviour, developed by experience [4]. One notes within the framework of diagnosis studies, that expertise, together with underlying skills, involves more functional representations [24]. And one knows that expert operators have a more exhaustive comprehension of the situation because they are using a broader framework of analysis [11]. Structural representations are thus associated with the skill level of the operators [22].

Moreover, internal complexity (taken here in the sense of cognitive workload) will be modulated by expertise. It is associated with the external complexity of the studied process, which determines the work requirements. It is generally considered that the amount of data to be taken into account as well as the intrinsic complexity of the variables and their interactions [30] are the principal factors of external complexity. Bainbridge [3] [4] demonstrated that internal complexity decreases with acquisition of skills.

The results presented in Part 2 demonstrate significant differences between operators considered by their peers to be experts. These differences could then be explained by skills differences. Some operators perform better because they are using more "expert" strategies; that is to say, they are using more structural and more functional representation whilst having a less important cognitive workload. This is the hypothesis we are going to test.

4. METHODOLOGY

From the experiment, we have gathered several types of data: verbal reports between operators and the virtual customer, as well as the actions of the computer assistant.

4.1 Independents variables

- Type of scenario: {Low complexity, Strong complexity}

For the 15 scenarios, we determine the number of possible breakdowns (translating the number of possibilities) according to the failure to be diagnosed. This number makes it possible to select the three most complex scenarios and the three least complex scenarios.

- Type of participant: {Least skilful operators, Most skilful operators}

The 10 participants of the experiment were recognized as being experts by their peers. We selected six: the three smaller percentages of success (46% of success on average) and three higher percentages of success (86% of success on average).

4.2 Dependants variables

- Type of representation: The aim here is to identify whether the representation is topographic (concerning the physical structure of the device) or functional (concerning the functions fulfilled by the device).
- Representation precision: Precision is taken into account according to whether it relates to a failure (which means a structural representation) or a breakdown (a factual representation).
- Internal complexity: This is evaluated using the number of remaining failures, indicating the number of breakdowns that must be memorized, because these breakdowns must be tested to identify the single failure.

4.3 Protocol analysis

To work on these individual protocols, we used the method developed by Hoc and Amalberti [2] [13] [14], starting from individual protocols to infer cognitive activities, which are then formalized in the form of a "predicate-argument" structure; activities constitute the predicates and specifications constitute the arguments. The coding scheme we use for diagnosis activity analysis was elaborated by Hoc and Carlier [15] and a sample of a protocol is presented in Table 1. Coding is done with the help of the MacSHAPA software developed by Sanderson *et al.* [27].

Table 1. Sample of a coded protocol.

Raw protocol	*Coded protocol*
Operator: How many apparatus are connected to this telephone number?	*HYPGEN(hypothesis number, time concerned by this hypothesis, representation type, representation precision, internal complexity, object, variable, value, condition, goal)*
	HYPGEN(1, present, topographic, structural, 49, apparatus, number, several, computer check, enrichment)
	IGG(means, time concerned, representation type, representation precision, internal complexity, object, variable, value, condition, goal)
	IGG(client, present, topographic, structural, apparatus, 49, number, ?, hypothesis, test)
Client: On this number, there is only one.	*IG(means, time concerned, representation type, representation precision, internal complexity, object, variable, value, level of interpretation of the value, condition, goal)*
	IG(client, present, topographic, structural, 48, apparatus, number, one, basic, hypothesis, test)
	TEST(hypothesis number, means, time concerned, representation type, representation precision, internal complexity, object, variable, value, issue)
	TEST(1, client, present, topographic, structural, 48, apparatus, number, several, invalidation, rejection)

Note. In each cell of the right column, the structure of the predicate is specified in italics before coding.

5. RESULTS

The results presented associate the significance test with the observed bilateral level of significance, a Bayesian judgment on the parent effect (δ) and the guarantee associated with this judgment (γ)[1] [19] [26].

5.1 Type of representation

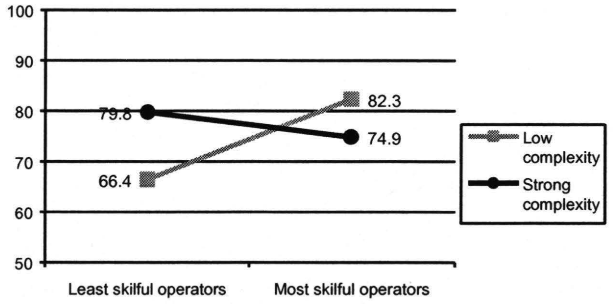


Figure 2. Effect of operators skills on the percentage of functional representations according to the level of complexity of the situation.

When complexity is low and with a skill increase, we could note an increase in the percentages of functional representations. This difference is not significant but it is notable (dl=4 ; tobs=1.331 ; N.S. ; p>.2 ; δ>4.6%, γ=.80)[2]; the most skilful operators thus manage the representations on a higher abstraction level than the least skilful operators in the situation with low complexity.

When complexity is strong we cannot infer an effect of the skill increase on the percentage of functional representations, and it is not possible to inferentially conclude on the importance of this effect (dl=4 ; tobs=0.42 ; N.S. ; p>.6).

5.2 Representation precision

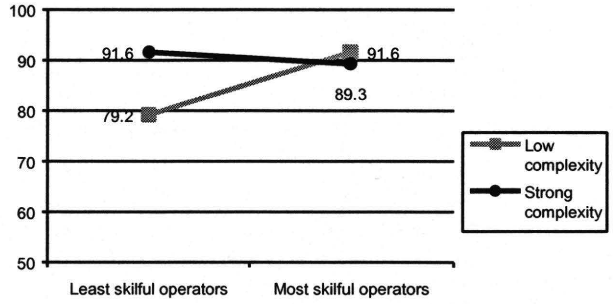


Figure 3. Effect of operators skills on the percentage of structural representations according to the level of complexity of the situation.

When complexity is low and with a skills increase, we could note an increase in the percentage of structural representations,

---

[1] Non informative: assuming a maximum uncertainty before the observation.

[2] A non significant result cannot be interpreted in terms of effect size. However, even with a reduced guarantee (here 80%), it is sometimes possible to give an indication of the size by Bayesian inference.

and this difference is notable although not highly significant (dl=4 ; tobs=2.355 ; N.S. ; p>.05 ; δ>7.4%, γ=.80); in the situation with low complexity, the most skilful operators work on a more structural representation than the least skilful operators.

When complexity is strong, we cannot distinguish the percentage of structural representations with the skills increase, the difference is not significant and we cannot inferentially conclude on the importance of this effect (dl=4 ; tobs=0.521 ; N.S. ; p>.6).

5.3 Internal complexity

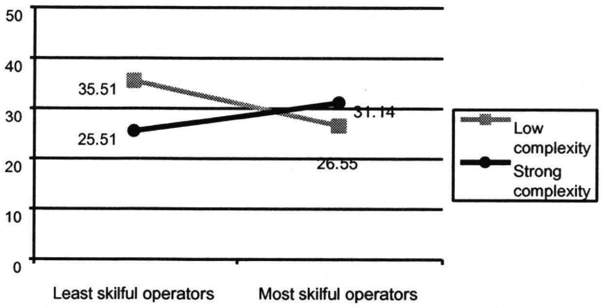


Figure 4. Effect of operators skills on the internal complexity (cognitive workload) according to the level of complexity of the situation.

In the least complex situation, we could note a reduction in internal complexity with an increase in skill, this effect is not significant but it is notable (dl=4 ; tobs=1.888 ; N.S. ; p>.1 ; δ>4.49 ; γ=.80). When complexity is low, the most skilful operators have a less important internal complexity than the least skilful operators.

In the most complex situation, we can also note a significant and notable difference of internal complexity with the skills increase (dl=4 ; tobs=2.778 ; S. ; p<.05 ; δ>3.72 ; γ=.80). When complexity is strong, the most skilful operators process a higher internal complexity than do the least skilful operators.

6. DISCUSSION

When the scenarios are slightly complex, the results presented in the literature are confirmed: the most skilful operators use more functional and more structural representations and their mental workload is less significant.

But when the scenarios are strongly complex, we cannot differentiate the two groups of participants with regard to their diagnosis strategies, and the least skilful participants present a less significant cognitive workload in an unexpected way. If we consider that skill acquisition leads to more functional and more structural representations, we could not explain the results using this approach.

This second result could then be explained by bringing it closer to the work of Leplat [20], who suggested the concept of *effectiveness*: the ratio between mental workload (cognitive cost) and performance quality (efficiency). Here, we find ourselves in a paradoxical situation; one where the operators' effectiveness is similar because the operators choose to preserve their resources (for the least skilful operators) or to boost their performance (for the most skilful operators).

This result demonstrates the management of a cognitive compromise in the sense of Amalberti [1]. During diagnosis, operators permanently evaluate the external risks of the situation (in the sense of not being able to complete the task), and the internal risks (to saturate the resources). This compromise aims for an acceptable performance whilst preserving resources.

This compromise is guided by former experience of diagnosis which enables the operators to judge their own performance (i.e. reflexive knowledge). This meta-knowledge must be taken into account because it can be decisive in resolution tasks (see for example the study of Lee and Moray [21]) and, in particular, in the "feeling of difficulty" (which means an internal risk of setting the cognitive capacities in failure due to cognitive resources saturation) and the "feeling of incomprehension" (which mean an external risk of loss of control of the situation). Therefore, the emergence of these feelings, developed by experience, influences the operators' strategies. This meta-knowledge controls the management of the cognitive compromise by either protecting resources or by monopolizing them in the task. This is one way of explaining performance differences between operators.

This is a mechanism that can be noted in other studies, such as those undertaken by Dörner [7]. In a laboratory task which was devoted to the study of complexity management, he used individual differences to distinguish between weak and strong performances. According to Dörner, the subjects that displayed weak performances are characterized by a weak judgement of their own capacity to act; and this judgement would exist a priori, that is.to say it is not due to the experiment. This weak judgement, can be placed alongside the emergence of a feeling of difficulty which leads operators to preserve their resources.

For these weak performances, Dörner raises several errors types. *Thematic vagabonding* implies that subjects change the topic under consideration, without supplementing it truly, in order to protect their resources. *Enkystement* seems to indicate an opposite effect: subjects process small details very precisely, but in fact are working in areas of the problem that are the least problematic and of the least importance. If one considers that these subjects show a cognitive compromise directed towards the resources, it should be added that errors presented by Dörner are visible in weak performances, thus confirming our study of a cognitive compromise that can differenciate individuals.

7. CONCLUSION

In diagnosis activities, as in problem-solving tasks generally, individual differences are often confused with operators' skills; indeed, it is usually considered that training can reduce all differences between subjects [8]. Our study shows that we have to take individuals' differences into account because they are not strictly related to operators' skills.

A tool cannot be introduced without first carrying out a performance evaluation of the human-machine joint system. However, this tool introduces task modifications, allowing new practices that were not previously taken into account to arise [28]. In our case, the introduction of the tool revealed the importance of the management of the cognitive compromise. If we consider that human strategies have to be used to design support tools [18], we have to consider individuals' differences in order to elaborate the representation format.

The use of distributed cognition is beneficial if we want to know which feature of the activities or artifact is relevant for the achievement of the task. And as Hollan, Hutchins and Kirsh [16] state (p.177): "From the perspective of distributed cognition, the organization of mind – both in development and in operation – is an emergent property of interactions among internal and external resources". We could add, then, that as we do not consider that computers and humans have an identical internal process, we have to take into account the fact that differences between humans means there are different ways to manage external resources.

8. ACKNOWLEDGMENTS

We thank David Jouglet for his authorization to use the data he had collected for another purpose.

9. REFERENCES

[1] R. Amalberti. *La conduite de systèmes à risques [Risky process control]*. Presses Universitaires de France, Paris, 1996.

[2] R. Amalberti and J.M. Hoc. Analyse des activités cognitives en situation dynamique : pour quels buts ? comment ? [cognitive activity analysis in dynamic situations : why and how ?]. *Le Travail Humain*, 61:209-234.

[3] L. Bainbridge, Diagnostic Skill in process operation. Paper presented at the International Conference on Occupational Ergonomics. Toronto, May 1984.

[4] L. Bainbridge, Development of Skill, Reduction of Workload. In *Developing Skills with Information Technology,* L. Bainbridge and S.A.R. Quintanilla, Eds. Taylor & Francis, London, 87-116, 1989.

[5] L. Bainbridge, *The relation between the categories in 'types of skill' and in Rasmussen's 'skill-rule-knowledge based' schema.* Unpublished manuscript, University College London, Department of Psychology. Available at: http://www.bainbrdg.demon.co.uk/Papers/RasmusSRKB.html.

[6] J.M. Cellier, H. Eyrolle and C. Mariné. Expertise in dynamic environnements. *Ergonomics*, 40:28-50, 1997.

[7] D. Dörner, On the Difficulties People Have in Dealing with Complexity. In *New Technology and Human Error*, J. Rasmussen, K. Duncan and J. Leplat, Eds., New York : John Wiley & Sons, Ltd., 97-109, 1987.

[8] K.D. Duncan, Representation of Fault-Finding Problems and Development of Fault-Finding Strategies. *Programmed Learning and Educational Technology, 22*:125-131, 1985.

[9] T.R.G. Green and J.M. Hoc. What is cognitive ergonomics ? *Le Travail Humain,* 54:291-304, 1991.

[10] J.M. Hoc. *Cognitive psychology of planning.* Academic Press, London, 1988.

[11] J.M. Hoc, Strategies in controlling a continuous process with long response latencies : needs for computer support to diagnosis. *International Journal of Man-Machine* Studies, 30:47-67, 1989.

[12] J.M. Hoc and R. Amalberti, Diagnosis: Some theorical questions raised by applied research. *Current Psychology of Cognition*, 14:73-100, 1995.

[13] J.M. Hoc and R. Amalberti, Analyse des activités cognitives en situation dynamique: d'un cadre théorique à une méthode [Cognitive activity analysis in dynamic situations: from a theoretical framework to a method]. *Le Travail Humain*, 62:97-130, 1999.

[14] J.M. Hoc and R. Amalberti, Modelling NDM cognitive activities in dynamic situations: the role of a coding scheme. In *How professionals make decisions*, H. Montgomery, B. Brehmer and R. Lipshitz, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, In press.

[15] J.M. Hoc and X. Carlier. A Method to Describe Human Diagnostic Strategies in Relation to Design of Human-Machine Cooperation. *International Journal of Cognitive Ergonomics,* 4:297-309, 2000.

[16] J. Hollan, E.L. Hutchins, D Kirsh. Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. *ACM Transactions on Computer-Human Interaction*, 7:174-196.

[17] E.L. Hutchins. How a cockpit remembers its speed. *Cognitive Science,* 19:265-288, 1995.

[18] U. Konradt. Strategies of failure diagnosis in computer-controlled manufacturing systems : Empirical analysis and implications for the design of adaptative decision support systems. *International Journal of Human-Computer Studies,* 43: 503-521, 1995.

[19] B. Lecoutre, M.P. Lectoutre and J.M. Grouin. A challenge for statistical instructors : Teaching Bayesian inference without discarding the « official » signifiance test. In *Bayesian Methods with Applications to Science, Policy and Official Statistics.* Office for Official Publications of the European Communities, Luxembourg, 311-320, 2001.

[20] J. Leplat. Regards sur l'activité en situation de travail, contribution à la psychologie ergonomique [Glances on activity in Work Situation, contribution to ergonomic psychology]. Presses Universitaires de France, Paris, 1997.

[21] J. Lee and N. Moray. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics,* 35:1243-1270, 1992.

[22] J. Patrick. Cognitive aspects of fault-finding training and transfer. *Le Travail Humain*, 56:187-209, 1993.

[23] J. Rasmussen. Strategies for state identification and diagnosis in supervisory control tasks, and design of computer-based support systems. *Advances in Man-Machine Systems Research,* 1:139-193, 1984.

[24] J. Rasmussen. *Information Processing and human-machine interaction: an approach to cognitive engineering.* North-Holland, Amsterdam, 1986.

[25] E.M. Roth and D.D. Woods, Aiding Human Performance I: Cognitive Analysis. *Le Travail Humain,* 51:39-64, 1988.

[26] H. Rouanet, Bayesian methods for assessing importance of effect. *Psychological Bulletin*, 119:149-158, 1996.

[27] P. Sanderson, J. Scott, T. Johnson, J. Mainzer, L. Watanabe, and J. James. MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA). *International Journal of Human-Computer Studies,* 41:633-681, 1994.

[28] K.J. Vicente. *Cognitive Work Analysis.* Lawrence Erlbaum Associates, Mahwah, NJ, 1999.

[29] A.T. Welford, Mental work-load as a function of demand, capacity, strategy, and skill. *Ergonomics*, 21:151-167, 1978.

[30] D.D. Woods. Coping with complexity : the psychology of human behaviour in complex systems. In *Tasks, errors and mental models,* L.P. Goodstein, H.B. Andersen, and S.E. Olsen, Eds. Taylor & Francis, London, 128-147, 1988.

Session 4 Technical Analytical Tools and Simulation

A Method for Defining and Inferring Team Situation Awareness in Cooperative Activity Based on Mutual Awareness

Yufei Shu
Department of Quantum Engineering and Systems Science
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo
113-8656, Japan
shu@cse.k.u-tokyo.ac.jp

Kazuo Furuta,
Institute of Environmental Studies
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo
113-0033, Japan
furuta@k.u-tokyo.ac.jp

Keiichi Nakata
Institute of Environmental Studies
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo
113-0033, Japan
keiichi.nakata@i-u.de

ABSTRACT

Team Situation Awareness (TSA) is a critical contributing factor in team cooperative activity. Our research program is finding out the underlying mechanism of TSA to reflect team cognitive process in a way consistent with team cooperative activity, and focusing on how to achieve mutual understanding, how to effectively incorporate between human teams and artifacts. In this paper, we argue that the earlier models of TSA, which have been discussed as an intersection of situation awareness (SA) owned by individual team members, are inadequate for study of sophisticated team reciprocal process. We suggest that the definition of TSA is necessary to integrate the notion of individual SA to cooperative team activity. We propose a new notion of TSA, which is reducible to mutual beliefs as well as individual SA in three levels. Further, we develop a feasible TSA inference approach and discuss human competence and system-related factors required to build TSA through field study. We try to demonstrate how TSA is actively constructed via the inferring practices. We also develop the principle to assess appropriateness of TSA from two aspects: the soundness and completeness of mutual belief. Comparison of evaluating results indicates that team uses the information from external environment to establish effective TSA based on internal knowledge.

Keywords

Team situation awareness, mutual awareness, cooperative activity, inference, team-machine interaction, cognitive process.

1. INTRODUCTION

Along with progress of information technologies, sophisticated automation is being introduced into complex process systems, unprecedented problems have arisen in terms of relation between humans and machines. Situation Awareness (SA) is a key concept that is widely used to assess performance of human-artifact interaction in a number of application domains, such as industrial process control (e.g., nuclear power plant, or process industry), aviation (e.g., cockpit or air traffic control), ship navigation, etc. In many of these operational contexts, operator crews are embedded in a dynamic environment: they continuously recognize their environment and change it by performing some operational action. Heavily explored by philosophers, psychologists and engineers, Team Situation Awareness (TSA) is becoming more broadly recognized as a critical contributing factor of team-machine interaction design and assessment. The purpose of providing team situation awareness is to establish and maintain a common context and to allow activities associated with one user to be reflected in another's mind.

Several studies have already been completed on TSA for these purposes (Salas, 1995, 2001). These works mainly discussed TSA as an intersection of SA owned by individual team members. However in a sophisticated team reciprocal process, such intersection is too simplistic in relation to TSA, which has a multi-layered structure of individual SA (ISA), mutual awareness (MA), and beliefs among team members. Although individual SA is adequately discussed in various frameworks (Endsley, 1995; Adams, 1995), modeling of TSA involved in a collaborative activity requires a number of additional notions. In particular, understanding of mutual awareness is a necessary element. Furthermore, the ability to infer mutual awareness in TSA is an important aspect to establish collaborative relation between team members or team and machines.

The aims of this study are to develop a new notion of TSA, which is reducible to mutual beliefs as well as individual SA, and to propose a TSA inference approach for team-machine cooperative activity.

2. CONCEPTUAL FRAMEWORK OF TEAM SITUATION AWARENESS

2.1 Characteristics of Situation Awareness

We followed the lead of human factors researchers who focused on SA as knowledge created through interaction between a person and her environment. The basic characteristics of SA are as followings: it is knowledge of a dynamic organization structure, maintained through perceptual information gathered from the environment. Various definitions of SA have been proposed. A general definition of SA is "the up-to-the minute cognizance required to operate or to maintain a system" (Adams et al 1995). Endsley (1995) focuses more on the process, proposing a three-staged definition:

level 1SA(SA1): perception of the elements in the environment,

level 2SA(SA2): comprehension of the current situation,

level 3SA(SA3): projection of future states.

We adopted the above notion of ISA, and then tried to construct a framework of TSA.

2.2 Conceptual Framework of TSA

Various definitions of team SA have been proposed. Salas (1995) defined team SA as "at least in part the shared understanding of a situation among team members at one point in time". Endsley (2001) suggested that team SA consists of both the situation awareness required for each team member and the overlap in situation awareness that is necessary among team members particularly for coordination.

Although we do not fully disagree with these definitions, we would like to put more emphasis on the interactive nature of TSA and reflecting the inferring fashion that human does in every day life. We adopt Endsley's three-staged definition and used the background on individual SA and earlier philosophical studies on cooperative activity to guide our definition of team situation awareness.

We considered TSA as a partly shared and partly distributed understanding of situation among team members. We view TSA as an extension of individual SA, which is not only aware of the environment but also other members in the team. The definition of TSA is:

Two or more individuals share the common environment up-to-the-moment understanding of situation of environment and another person's interaction with the cooperative task.

First, TSA is an awareness of the distributed context and how it changes within a team. Second, TSA is an awareness of the team member and how they interact within a distributed context. In a collaborative situation, people must prepare to undertake other's task, that of collaboration, and therefore their TSA must involve both the domain and the collaboration. Then we can say TSA includes two basic elements: mutual awareness and individual situation awareness.

In a broad sense, mutual awareness refers to the awareness that individuals of a cooperative entirety have of each other's activities, beliefs, intentions and so on. We view mutual awareness at different levels, which refer to different human cognitive process: attention, perception of events or facts, comprehension of what perceived, etc. We think that mutual awareness includes following interrelated elements: mutual access, mutual responsiveness (Bratman, 1992), and knowledge and understanding of each other's work.

3. DEFINITION OF TSA

We attempted the use of precise logical formula to express a definition of TSA, which specified a set of heuristic rules. Such rules allow us to reason about TSA, particularly the potential awareness that individuals might have of each other's mental states.

As mental phenomenon, we attempt to employ a sort of mental states, such as belief and expectation for other members as evidences to describe TSA. We use modal logic to define TSA, which is similar to that employed in an earlier AI study on joint action (Levesque, 1990). We proposed some formulations to describe the mental processes, such as modal operators, e.g., SA, TSA to interpret awareness, BEL, EBEL and MBEL are used to denote an individual belief, the conjunction of individual beliefs and mutual beliefs, which are adopt from previous AI study (Rao, etc. 1992) respectively, and some predicates, e.g., Perceivable, Comprehensible, and Projective to imply mental processes and Hold, State, Symptom to denote system states. These formulations have the usual connectives of a first order logic language (? :negation, ? : conjunction, ? : disjunction, ? : implication, =: equivalence). The definition of SA in three levels is:

SA1 (m, P)=BEL(m, Hold(P, now)) ? Symptom(P)

SA2 (m, P)=BEL(m, Hold(P, now)) ? State(P)

SA3 (m, P)=BEL(m, Hold(P, after(now))) ? State(P)

The definition of TSA in three levels is:

TSA1(g, P)= $?_{m?g}$ $?_{Pi?P}$ (SA1(m, Pi)? MBEL(g, Hold(Pi, now)))

TSA2(g, P)= $?_{m?g}$ $?_{Pi?P}$ (SA2(m, Pi)? MBEL(g, Hold(Pi, now)))

TSA3(g, P)= $?_{m?g}$ $?_{Pi?P}$ (SA3(m, Pi)? MBEL(g, Hold(Pi, after(now)))).

Where Pi={P|SA(m, Pi)} and P=$?_{m?g}$ Pi.

4. METHODOLOGICAL FRAMEWORK for ASSESSING TSA

From the logic definition, we try to find out the underlying structure of TSA. For simplicity, we assumed the dyadic case where the team consists of two members, A and B. The nesting of mutual beliefs can repeat up to the infinite depth theoretically, but considering two or three levels might be enough in practice. From the definition of TSA, TSA(g, p) for the dyad case is simplified as:

TSA(g,P)= SA1(A, Pa)? BEL(A,BEL(B, Hold(Pb, t)))? BEL(A, BEL(B, BEL(A, Hold(Pa, t)))

holds in A's mind and its counterpart holds in B's mind:

SA1(B, Pb)? BEL(B, BEL(A, Hold(Pa, t)))? BEL(B,BEL(A, BEL(B, Hold(Pb, t))),

where t=now for TSA1 and TSA2, and t=after(now) for TSA3. It shows that A's TSA consists of three basic components: her own SA (SaPa, or SaCa, SaFa), belief on B's individual SA (BaPb, or BaCb, BaFb) and belief on B's belief on her (A) own SA (BaBbPa, or BaBbCa, BaBbFa) in each of the three levels. Here, Pa/Pb is used to denote A's/B's perception, Ca/Cb is used to denote A's/B's comprehension, Fa/Fb is used to denote A's/B's projection, and Ba/Bb is used to denote A's/B's belief. The later two parts of the formula correspond to mutual awareness. This structure is similar to that described in previous philosophical studies on we-intention (Tuomela and Miller, 1987), and it corresponds to that shared cooperative activity (Bratman, 1992), which has three conditions to obtain: commitment to the joint activity, commitment to mutual responsiveness and commitment to mutual support. Fig.1 shows an illustrative view of the above analysis.

Since what are in A's mind do not necessarily coincide with what are in B's mind, different appearances of Pa and Pb in the above expression are not necessarily equivalent. As for Pa, there are three appearances, and each of them defines a different version of Pa.

Pa={p | SA(A, p)},

Pa'={p | BEL(B, BEL(A, Hold(p, t)))},

Pa"={p | BEL(A, BEL(B, BEL(A, Hold(p, t))))}.

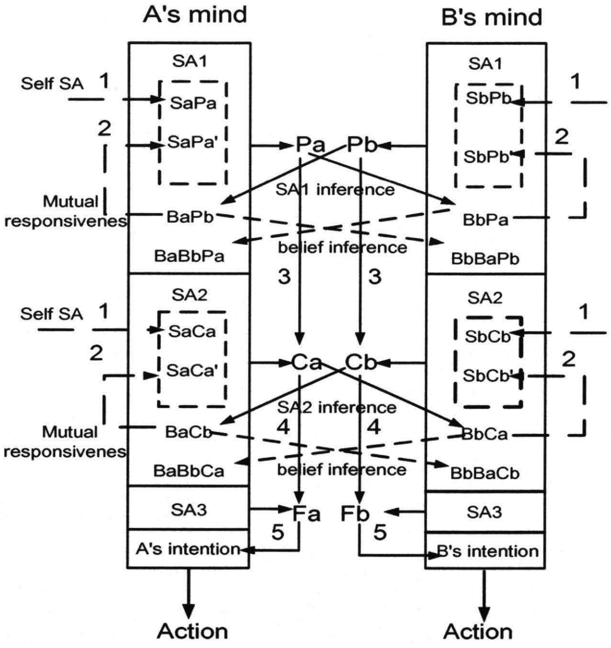


Fig.1 Illustrative structure of TSA and abilities required for its establishment

The team has the correct and complete TSA, if and only if the above three versions coincide with each other, i.e., Pa=Pa'=Pa". If any proposition p in Pa is not included in Pa' or Pa", the TSA is incomplete. If any proposition p in Pa' or Pa" is not included in Pa, the TSA contains wrong belief. The difference between one's real SA1 and other's belief on it creates the space for evaluating individual difference in the ability to achieve complete and correct TSA. The soundness of inferred SA depicts the degree of inferred SA, which matched with real SA. Completeness of inferred SA implied the degree of real SA, that matched with inferred SA. For example, the soundness of B's belief on A's SA1 can be assessed as follow:

Soundness of Pa' =count(Pa n Pa')/ count(Pa')

The completeness of B's belief on A's SA1 is:

Completeness of Pa' = count(Pa n Pa')/ count(Pa)

A similar claim is applicable to the three versions of Pb and SA2, SA3. By assessing the difference among Pa, Pa' and Pa", we can assess the appropriateness of TSA in cooperative activities. Additionally, by comparing the difference between the real TSA and the expected TSA, we can assess the design of team-machine system or team performance.

5. METHOD OF TSA INFERENCE

Mutual awareness is a basic element in TSA, however, how does one know the other's SA? We proposed an inference algorithm of TSA in this study to identify the problem of individual SA and belief inferring. When team members cooperatively work together, there are two ways to achieve mutual awareness:

1) One is obtained as a result of explicit communication, as in direct face to face verbal exchange or via artifacts;

2) Other one is inferred from implicit verbal communications or from observing the external action of another without explicit verbal exchanges (it is called non-verbal communication).

For communication, we defined communication channels, which are decided by organizational structure, role allocation and member's knowledge, rule or skill map, prescribed in the knowledge base. We also considered three communication modes: 1) broadcasting 2) ask and answer 3) instruction. However, verbal communication is not the usual mode of human mutual awareness, because humans do not think aloud in ordinary situations. Instead humans can infer others' SA feasibly from their external behavior based on knowledge about them.

For the inference method, we identified a set of inferred SA and belief that satisfy the condition SaPa=BbPa=BaBbPa and SaPb=BaPb=BbBaPb for TSA1. In the above, BbPa/BaPb is used to denote that B/A inferred result of A's/B's perception (SA1) and BaBbPa/BbBaPb to denote that A/B inferred result of B's/A's belief on A/B respectively. It is similar for TSA2 and TSA3. TSA is described as a set of individual SA and mutual beliefs. Fig.2 shows an overview of the proposed approach. The solid line arrows correspond to the process of self SA, the broken line arrows correspond to the process of inferring other's SA, and the dotted broken lines arrows correspond to the process of inferring other's belief. Upper side shows Operator A's mental process and down side shows B's. We will explain SA and belief inference that comprise our proposed inferring method in detail as followings.

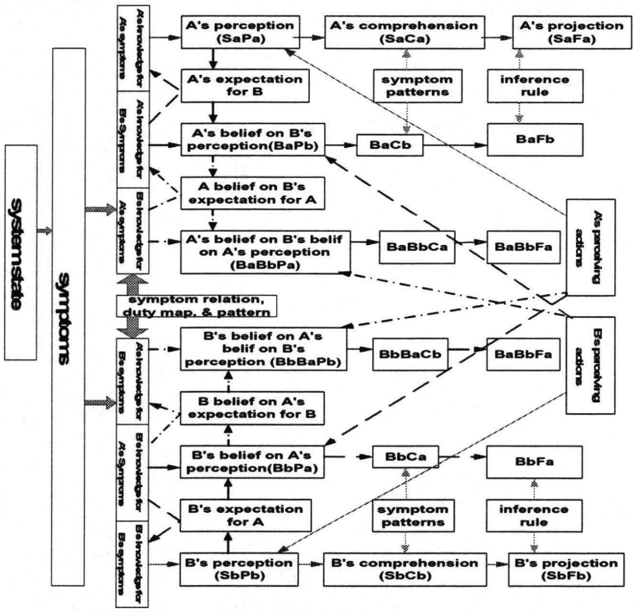


Fig. 2 Overview of TSA inference approach

5.1 SA Inference

As we know, one makes up beliefs from what she senses. The information that is useful for SA inference is what can be observed from the environment, system interface and other team members. In the prescribed system scenario, if we assume that there are two members A and B. When A perceives the symptom from the system, she will form the first part of her SA1 autonomously in light of her knowledge on the symptoms. Then she will use the mutual knowledge to infer other's perception.

The method used in inferring other's perception requires strong mutual knowledge. Firstly, all possible current system state is generated from system symptoms, which are perceived by operator A. This process is modeled as forward similarity matching.

Secondly, symptoms that an operator might be expected to pursue are listed from the causality map that is derived by means-ends analysis by consideration of the current system states. This process is backward similarity matching.

Thirdly, procedures to perceive these symptoms are listed in action plan, which prescribed the relation between perceiving actions and all possible symptoms to be perceived.

Then, comparing these perceiving action related symptoms with the expected perceived symptoms, candidates of the other's perception are identified.

Finally, for each candidate, we use pattern-matching method to calculate its confidence. In this inference method, it is not always the case that one symptom corresponds to one possible state. Under the different system state, the maximum confidence is selected for a certain symptom.

If there is shared display in the control room, operator can observe other's symptom easily. In this case, one can infer the other's SA by observing her perceiving action and observing the same symptom. This is because we assume that team members have same knowledge to deal with that particular information.

Operator puts the result of perception inference into working memory and uses it as an input of inferring comprehension (SA2). Here, we assume that A believes that B has Pb, also A believes that B has the knowledge to deal with Pb, and A knows this knowledge. Then operator A uses this inferred result (BaPb) to infer Cb with the knowledge as B has.

5.2 Belief Inference

B's belief on A's SA (BbPa) is not exchanged explicitly often either. Operator A needs to describe her belief on B's belief (BaBbPa) to infer the B's belief to form her third SA component. We surmised that B's belief could be inferred from an interpretation of A's own SA, in light of A's beliefs on B's SA that is responsive or in support of A's situation assessment. This corresponds to Bratman's (1992) theory. He thought it is characteristic of shared cooperative activity that each participating agent attempts to be responsive to the intentions and actions of the others, and called this feature "Mutual Responsiveness". We extended this theory so that not only in the intention, or action stage, are the participating agents responsive to each other, but also in the symptom perception and state diagnosis stages. The belief inference procedure is based on ISA inference. Accordingly, the current system state is no longer derived from system symptom perceived by Operator A, but is from the description of B's perceived symptoms, which is formed from A's point of view (BaPb).

The inferred belief on perception is an input to infer belief on comprehension. It is also similarity-matching process as previously discussed.

5.3 TSA Inference Approach

SA inference and belief inference are a parallel mental process, but belief inference relays on the result of SA inference. Both inferring processes based on the same state-symptom patterns and action plans in the long term memory and shared perception candidates in each inference process. In addition, belief inference process starts from the result of SA inference.

As a result of both individual SA and Belief inference, the set of an individual SA and beliefs on the other member is obtained. The inference engine searches for the combinations of them that satisfy the condition: SaPa=BbPa=BaBbPa and SaPb=BaPb=BbBaPb for the first level TSA, and defines them as TSA1 as an out put. Similar condition for generating second level TSA is SaCa=BbCa=BaBbCa and SaCb=BaCb=BbBaCb and for generating third level TSA is SaFa=BbFa=BaBbFa and SaFb=BaFb=BbBaFb. The TSA inferring method proposed here is domain independent. It can be used under any situation to achieve TSA. However, the method should use domain dependent knowledge for a concrete inferring process

6. APPLICATION of TSA INFERENCE APPROACH

6.1 Plant Simulator

The proposed method was applied to operate DURESS (Dual Reservoir System Simulation) to generate individual SA, belief and test data, in which team operators must deal with all symptoms to achieve the mission goal

DURESS is a thermal-hydraulic process simulation (Vicente K, and Rasmussen J, 1990). Fig. 3 shows the plant configuration of DURESS. It consists of two redundant feed-water lines that can be configured to supply water to either, both, or neither of the two tanks. Each tank has associated with it an externally determined demand for water that can change over time. The work domain purposes are: to keep demanding outlet flows/temperatures at Valve5/Valve6, and maintain each tank water level from 20% to 80%. On the interface panel, the power of each pump and heater, level of each tank, inlet flow of each tank, outlet flow and temperature, demanded flow and temperature and other 10 other alarms (Tank A/B water level High/Low, tank A/B temperature High, Flow A/B different, Flow A/B temperature different) are indicated. All of them are available symptoms perceivable by Operator A/B to have TSA.

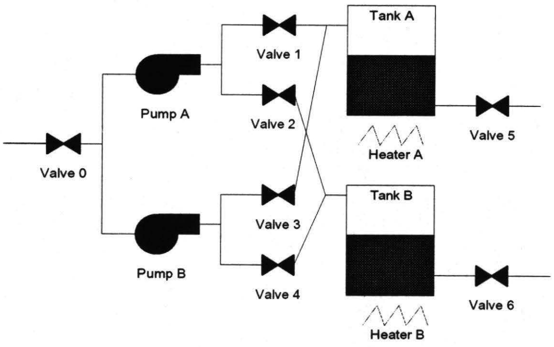


Fig. 3 DURESS configuration

6.2 TSA Simulator

Operators' missions are to perceive symptoms, recognize what happened within the simulator plant, and what the other team members perceived, and comprehended. Each member is assigned a role to monitor specific parameters of process simulation plant shown in Fig. 3. He/she not only can monitor the parameters assigned to the other member, if he/she is not busy and there is shared display, but also can observe others' actions. All event scenarios are prescribed.

The TSA simulator emulates the operator's information process, situation assessment and interaction with others based upon information received from the environment and supported by an internal mental model of the simulated scenario. Each individual operator can be thought of as an agent, that can perceive and recognize autonomously based on information obtained from the environment and other agents in addition to its own knowledge. TSA simulator is a knowledge-based system. It includes the information processor and the situation assessor subsystem, which perceives signal from environment or others and stores all necessary knowledge for processing perceived information.

The architecture of the TSA simulator consists of two distinct but tightly coupled modules: 1) The TSA inference engine, and 2) the knowledge base. The data used for the inference (symptom and perceiving action) is put into the inference system, with various kinds of knowledge prescribed in the knowledge base, the inference engine generates its inference result and outputs are individual SA, belief and TSA. The inference engine was developed in a logic programming language (Prolog), and integrated with a process plant simulator. Fig. 4 shows the architecture of TSA simulator

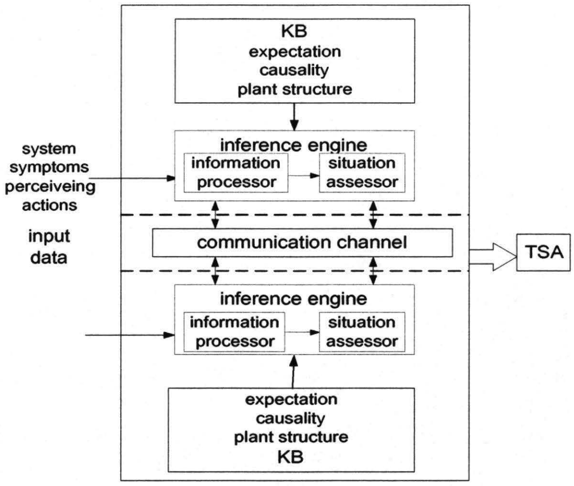


Fig. 4 Architecture of TSA inference system

6.3 Result and Discussion

6.3.1 Scenario

We will prescribe here contents of operational scenario used in this study. Each scenario describes initial plant event, role allocation, communication channels, each member's mental model, control panel structure (shared of not), and control room setting (can observe each other or not). Two groups including 12 scenarios were prepared by changing various factors. First group of scenario 1-9 starts with Pump A trip and the second group of scenario 10-12 with Tank A leakage.

The 12 scenarios are summarized in Table 1.

Table 2 is an example of operator's perceiving actions revealed in Scenario 1-9, which can or cannot be observed by the other team members depending on the scenarios. The starting point of the time corresponds to when the abnormal symptom is detected by the team, such as Tank A level is low. The second column is Operator A's perceiving action, and the third one is B's. For example, Operator A checks Tank A level at time interval T2. Each test scenario is divided into several intervals (e.g. T1, T2, ...) by the time points when new information such as perceiving new symptom or observing other's action is entered into the inference system.

Table 2 Perceiving actions (Scenario 1-9)

Time point	Perceiving action	
	Operator A	Operator B
T1	Quantify Tank A inlet	Quantify Tank A inlet
T2	Check Tank A level	Check Tank A level
T3	Quantify Tank A level	Quantify Tank A level
T4	Check Tank A temp	Check Tank A temp
T5	Check Line A outlet	Check Line A outlet

6.3.2 Result

Table 3 shows the example result of the proposed TSA inference approach, which represents the result for Scenario5 at assessing point T4. It shows the individual SA1, inferred belief on other's SA1, individual SA2 and inferred belief on other's SA2. The Operator A's SA1 is "tankA_level is decreasing with a certainty factor of 0.7" and "tankA_inlet is zero with a certainty factor of 1.0". Operator B's belief on A's SA1 is "tankA_inlet is zero with a certainty factor of 1.0". All of these A's SA1 or inferred belief on A's SA1 are interpretation of A's SA1 from a different aspect. However, there is only one inferred result on A's perception, because B's belief on A is more restricted by action plan described in the knowledge base. For example, the result of inferred A's SA1 must be based on observed A's perceiving action. If B did not monitor A's perceiving action yet, she cannot infer A's SA. Another reason is that inference is based on one's own state recognition. If B did not have enough information to recognize the current state, she cannot infer A's SA either.

Table 1 Scenarios used in the test simulation

Scenario	Failure Event	Role allocation	Observation	Display	Communication	Mental model
1	Pump A trip	Type 1	No	Shared	Yes	Same
2	Pump A trip	Type 1	A ↔ B	Shared	No	Same
3	Pump A trip	Type 1	A ↔ B	Shared	Yes	Same
4	Pump A trip	Type 1	No	Shared	No	Same
5	Pump A trip	Type 1	A → B	Shared	No	Same
6	Pump A trip	Type 1	A ↔ B	Shared	No	Different
7	Pump A trip	Type 1	A ↔ B	Shared	Yes	Different
8	Pump A trip	Type 1	A ↔ B	Separate	No	Same
9	Pump A trip	Type 1	A ↔ B	Separate	Yes	Same
10	Tank A leak	Type 1	A ↔ B	Separate	Yes	Same

11	Tank A leak	Type 2	A ↔ B	Separate	Yes	Same
12	Tank A leak	Type 2	A ↔ B	Separate	No	Same

Table 3 Result of TSA inference (Scenario 5 at T4)

Operator	SA1	Inferred SA1	SA2	Inferred SA2
A	**1. tankA_level is decreasing (0.7)** 2. tankA_inlet is zero (1.0)	**tankA_level is decreasing (0.7)**	**1. TankA level is decreasing (0.7)** 2. TankA_inlet is zero (1.0) 3. Trip pump A (0.54) 4. LineA outlet is constant (0.54) 5. tankA_temp is increasing (0.54) 6. tankA_level is low (0.54)	**TankA_level is decreasing (0.7)**
B	**1. tankA_inlet is zero (1.0)** 2. tankA_level is decreasing (0.7)	**tankA_inlet is zero (1.0)**	**1. tankA_inlet is zero (1.0)** 2. tankA_level is decreasing (0.7) 3. Trip pump A (0.54) 4. tankA_level is low (0.54) 5. LineA outlet is constant (0.54) 6. tankA_temp is increasing (0.54)	**tankA_inlet is zero (1.0)**

6.3.3 Evaluation and discussion

Various factors that affect TSA were evaluated by the proposed assessing method from two view-points: soundness and completeness of inferring results.

Communication and Observation

Fig.5 presents the completeness score of inferred belief on A's SA1 of different five teams corresponding to Scenarios 1-5, along time line respectively. It shows that Team 1 and Team 3 have high scores of mutual belief on SA1 and the score did not significantly vary across Team 1 and 3. Team 1 can communicate each other, but cannot observe each other, while Team 3 can communicate and observe each other. This result indicates that even if the team cannot observe each other, they can have good TSA by communication. Comparison of Team 2, who can observe but cannot communicate each other, with Team 3 reveals the score of Team 2 is significantly lower than Team 3. It indicates that that even if the team has the same mental model and a shared display and they can observe each other, the team cannot achieve good mutual understanding without communication.

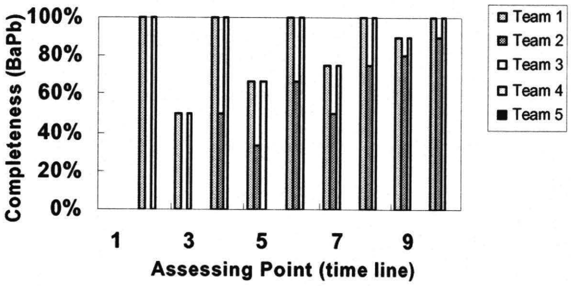


Fig. 5 Score of BaPb in different scenarios

Since Team 4 can neither observe nor communicate each other, they have no mutual understanding at all even with the same mental model and a shared display. It shows that verbal communication or mutual observation is an essential element for inferring others' SA1.

In Team 5, both A and B have no mutual communication, A can observe B, but B cannot observe A. The score of belief on each other's SA1 is therefore definitely different. A's inference on B's SA1 is much better than B's inference on A's SA1 as shown in Fig. 6, since B cannot observe A's action, which is key input data for inferring A's SA without communication. As for the B's beliefs inferred, both Team 4 and 5 totally have no mutual understanding, because they have no input data for inferring.

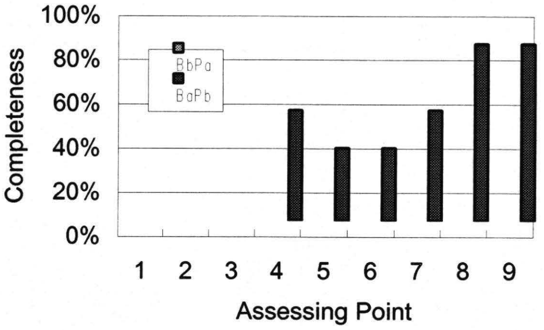


Fig. 6 Score of inferred SA1 in Scenario 5

Fig.7 shows the score of inference result of B's SA2. We can find that the score of Team 2 is better then Team 5, even though A can observe B's action in the both cases. It is because the second level SA depends not only on mutual knowledge, but also on input information from the previous inference stage. Even if they have the same mental model, they can have different scores of SA2 inference. Team 1 and 3 have the best score, because they can communicate at any time. Team 2 has a low score at very beginning, but the more evidence is obtained, the better mutual understanding they achieves. Team 4 has the worst score, because they can neither communicate nor observe each other.

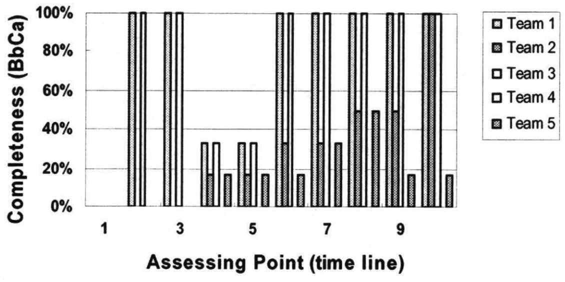


Fig. 7 Score of BaCb in different scenarios

Shared Displays

Fig.8 shows the result of Scenario 2, 6 and 8. Firstly, we try to find out the effect of display by comparing results of Scenario 2 and 8. Team 8 has no shared displays, while Team 2 has a shared display. The result shows that Team 2 has better scores than Team 8. The shared display is useful for the team to build up mutual understanding.

Furthermore, we compared inferred result of Scenario 3 and 9. Team 9 has no shared displays, while Team3 has a shared display. The result shows that Team 3 and 8 have almost the same level of mutual understanding. It indicates that usefulness of the shared display can be substituted by other means of communication, in this case, verbal communication.

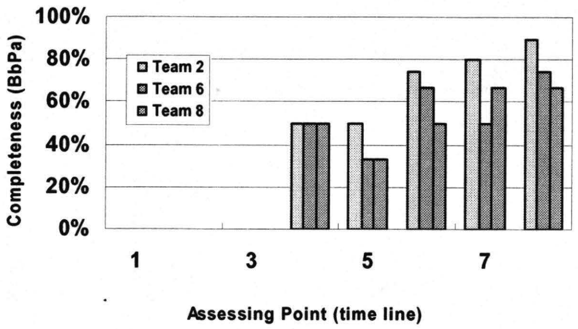


Fig. 8 Score of inferred SA1 in Scenario 2, 6 and 8

Mental Model

Now, we try to check out the effect of mental model by comparing results of Scenario 6 and 2, shown in Fig.8. Team 6 has different mental models, while team 2 has the same mental model. The result shows that the team with the same mental model has better mutual understanding.

Also we compared inferred result of Scenario 7 and 3. The result shows that the team has the same level of mutual understanding, since even the mental model is different in Scenario 7, communication will help team members to achieve the same level of mutual understanding as Team 3.

By Different TSA definition to evaluate result of TSA

We assessed the simulation results by different definition of TSA to show the validity of our TSA model. We fixed event scenario and role assignment. Fig.9 shows the completeness of TSA in Scenario 2. We find that TSA defined as the intersection of all individual SAs (XSA) resulted in better scores than TSA based on mutual belief (MSA). Since the team members have the same mental model, it is easy to have the common elements of ISA.

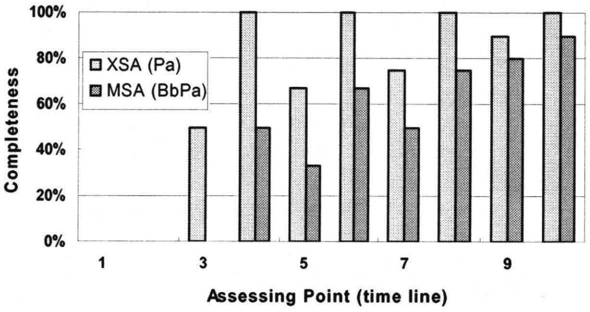


Fig. 9 Score of inferred SA1 in Scenario 2

Comparing the result of Scenario 7 shown in Fig.10, however, we will find that MSA has better scores than XSA. If the team has different mental models, it is hard to have the common elements of ISA, but the team members can establish mutual understanding by communication.

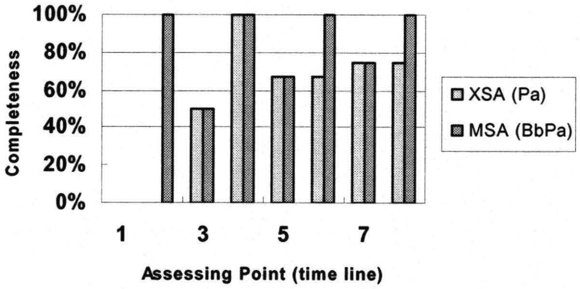


Fig. 10 Score of inferred SA1 in Scenario 7

In the result shown in Fig.11 for Scenario 6, the scores for MSA are lower than those for XSA, since the condition did not allow them to build up mutual understanding, which is achieved either by communication or by the common mental model. Comparing Fig.10 with 11, we find that the sores for MSA in Fig.10 are significantly better than Fig.11 at any assessing point, while those for XSA do not change. It seems communication has much more favourable effects for MSA than XSA, which is decided only by a shared mental model.

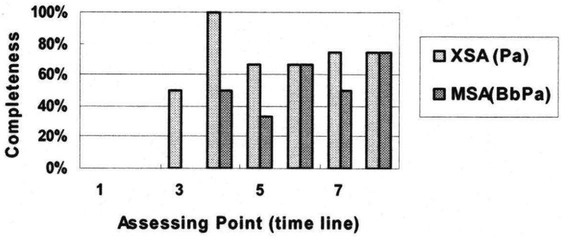


Fig. 11 Score of inferred SA1 in Scenario 6

We also tested different event cases and role allocation, but the results are very similar. Through application of proposed method to DURESS simulator, we verified that our TSA model reflects team cooperative process more appropriately than the conventional notion of TSA defined as the intersection of ISAs. This study also supports the hypothesis that TSA can be enhanced by providing a team with common information using shared displays as well as shared mental models. Contrary to expectations, however, good TSA can be achieved even without shared displays in some other ways of communication.

7. CONCLUSIONS

This work provides a conceptual and theoretical framework for describing and assessing notion of team situation awareness (TSA) in cooperative activities. Based on previous studies on SA, intention, belief, and joint activity in philosophy, AI, and human-machine systems, we have proposed a new notion of TSA and given its formal definition. It is assumed TSA is reducible to individual SA and mutual belief, and it consists of three levels of perception, comprehension and projection following the individual SA definition. We have also proposed the principle to assess appropriateness of TSA on this framework.

Combining this TSA simulator to proposed assessing method, we can evaluate the TSA dynamically as it is enhanced or degraded with application of new type interface in the simulator. The use of this evaluating method allows us to consider the operator's basically cognitive capability and limitation, and integrate these operator-related factors with critical system-related factors, such as information quality and amount of control panel, the control panel layout format and way to supply the information of human-machine interface. The inference simulator is free from any cognitive limitation, which is superior to human assessor at this point.

The TSA discussed in this paper is of a narrow sense, which is just related to the observation and interpretation stages of human cognition. TSA in broad sense, however, includes aspects of planning and execution, and we have to consider additional mental constructs such as shared goal, shared plan, and shared intention to discuss cooperative activities. We have already proposed a formal definition of team intention in our previous work to infer team intention and to detect inconsistency of team intention for cooperative human-machine interaction. Combining the team intention of our previous work and TSA of this work can provide the definition of TSA of a broad sense: all aspects of TSA therefore can be covered. It will be a challenge of next step work.

8. REFERENCES

[1] Adams M., Tenney,Y., situation awareness and the cognitive management of complex systems, Human factors, 37(1),85-194, 1995.

[2] Bratman M., Shared cooperative activity, 101(2) P327-341, the philosophical review, 1992

[3] Cooke N.J., Stout R. J., Salas E., A knowledge elicitation approach to the measurement of team situation awareness, 114—139, New trends in cooperative activities, human factors and ergonomics society, 2001

[4] Endsley M.R Towards theory of situation awareness in dynamic system. Human factors, 37(1), 32-64, 1995

[5] Levesque H.J., Cohen P.R. and Nunes J.H.T., On acting together, Pro. AAAI-1990,P94—99,1990

[6] Rao A.S., Georgeff M.P., and Sonenberg E.A., Social plans, Decentralized AI-3, P55—76, Elsevier Science Publishers, 1992

[7] Salas E., Prince C. Situation awareness in team performance, Human factors, 37(1),123-136, 1995

[8] Tuomela R., Miller K., We intentions, philosophical studies, 53, P367—398, 1988

[9] Vicent K and Rasmussen J., The Ecology of Human Machine Systems II. Ecological Psychology, 207-249

Pre-operational validation in safety critical environments: the challenge of Advanced Shadow Mode Trials

Paola Lanzi
Deep Blue s.r.l.
via Basento 52/D
00198 Rome, Italy
+39 06 8554801
paola@dblue.it

Patrizia Marti
University of Siena – Dep. of Communication Science
Via dei Termini 6, 53100
Siena, Italy
and Deep Blue s.r.l.
via Basento 52/D 00198
Rome, Italy
+39 0577 270335
marti@unisi.it

Paolo Battino
Deep Blue s.r.l.
via Basento 52/D 00198
Rome, Italy
+39 06 8554801
paolo@dblue.it

Simone Pozzi
University of Siena – Dep. of Communication Science
Via dei Termini 6, 53100
Siena, Italy
and Deep Blue s.r.l.
via Basento 52/D 00198
Rome, Italy
+39 06 8554801
pozzi@media.unisi.it

ABSTRACT

This paper presents a new approach to the validation of pre-operational systems in safety critical environments, based on Advanced Shadow Mode Trials.

Shadow-mode is a technique for validating pre-operational prototypes in real work settings.

In Air Traffic Control (ATC), the technique is usually implemented using a prototype system that is fed with live data from the environment (e.g. radar and other data), but that is not actually used to control or influence the live traffic. The prototype is operated by an off-duty controller or other trial staff and measurements and observations are recorded for the following off-line analysis. However, recently the level of use of the prototype has been varied depending on the maturity and reliability of the prototype itself and the objective of the validation. In the paper, we discuss a new application of the technique in which the traffic is entirely managed by the trial team using the prototype system whilst operational controllers work in parallel as back-up unit.

Keywords

Validation, Shadow Mode Trials, Air Traffic Control, Safety Critical Environments, Medium Term Conflict Detection.

1. INTRODUCTION

This paper is a methodological and experimental contribution to the debate on pre-operational validation. It addresses the problem of how to validate pre-operational tools or systems designed to be introduced in safety critical environments like Air Traffic Control (ATC). By definition the pre-operational validation should not be targeted to investigate aspects such as usability, reliability or compliancy of the tool with its original requirements. Achieving the pre-operational status means that usability, reliability and compliancy with user and system requirements have been already analysed and addressed in previous stages of the development/validation process. What is still missing to assess is the adequacy to the socio-cultural and technical environment in which the tool is going to be introduced. In other words, what misses is the analysis of its context suitability.

It is proved that the context of use plays a fundamental role in the way a new tool is adopted and used in a certain working environment [5 and 12]. Consisting of an original and dynamic combination of physical, social and cultural entities, the context of use cannot be considered as just the setting where human actions take place. It is a sort of framework that supports the human perception of the world, and is meaningful in itself. It deeply shapes the way the new tool is perceived, adopted and used. But, the context is in turn modified and re-shaped by the introduction of the new tool, that brings on changes in the relations among its physical, cultural and social components.

The dynamic and mutual relationship between context of use and integration of new tools and procedures is crucial in determining the success of a new system. The literature on system validation [9] reports few validation techniques that, at different levels, allow to assess whether a tool is suitable for a given context of use and therefore likely to be successfully introduced in the real working environment. The most popular techniques are Real Time Simulations (RTSs) and Passive/Active Shadow Mode Trials (SMTs). In both cases, the pre-operational validation is based on simulations, that can take different formats, according to the objectives of the validation, the maturity and reliability of the tool itself, the nature of the tasks supported by the new technology and the activity context.

In RTSs the operators interact with the prototype system in a simulated but realistic environment. This makes RTSs especially suited for validation, with a focus on human performance. In this technique, controllers and other participants interact with and react to the simulated traffic samples in purposely designed scenarios.

As a matter of fact, RTSs are fundamental to evaluate the use of a pre-operational system with respect to a large set of situations. They allow to investigate many aspects of the new tool in a controlled and safe way. However what they do not provide is the possibility to deal with unexpected and complex events, that are almost unlike to happen in a structured simulated environment, where everything is planned in advance. Moreover, even if effort is spent to make the simulated environment as much similar as possible to the operational one, it can pretend to be realistic, but not real. Several aspects of the context of use, including the variability of social and cultural components, are almost impossible to reproduce in a simulation room. Therefore Real-Time simulations present advantages and disadvantages. On the one hand they allow to control the events to be tested, permitting to explore the way the system could be used in critical and rare situations, that hopefully are difficult to observe in the everyday activity [8]. On the other hand they are not able to reproduce the extremely rich and dynamic context of use in which the human activity is carried out, producing limited results on the actual suitability of the tool for the real working environment.

In Passive/Active SMTs (see next paragraph for details), the operators act in the real working environment, managing a prototype tool that is fed with live data, but is not actually used to control or influence the live traffic. The pre-operational tool is operated by an off-duty controller or other trial staff and measurements and observations are recorded for an off-line analysis. In this case the context of use is not simulated, but real. What is realistic, but not real, is the human activity. Controllers in fact simulate the traffic management with the prototype tool, but they are not entitled to instruct the pilots. As consequence, even if the new tool supports them in defining and implementing new and more effective strategies, they cannot see the effects of their decisions. Therefore even the SMTs have both advantages and disadvantages. On the one hand they allow to assess the effect of introducing the new pre-operational tool in the real working environment. On the other hand they do not allow to observe the actual interaction with the tool: even if introduced in the real working environment, the pre-operational tool cannot be used to manage the traffic.

In the paper, we present an alternative approach that seeks to overcome the limitations of RTSs and SMTs. This approach is based on a very promising and innovative technique, named Advanced Shadow Mode Trials. What distinguishes Advanced Shadow Mode Trials from RTSs and Passive/Active SMTs is that the simulation is completely abandoned and the trial is performed in the real working environment with real traffic and entirely managed by a Trial Team. This Team actually uses the pre-operational tool, and it is supported (in case of need) by the Operational Team working in parallel on the normal platform as back-up unit. The Operational Team mainly monitors the situation ready to intervene when needed. The potentialities (and the risks) of this technique are quite evident. What is less evident is how to apply this kind of technique in completely safe conditions and how to combine and integrate the results with those of other validation techniques like the RTSs.

This paper contributes to answer these issues giving full evidence of the validation work conducted during the first Advanced Shadow Mode Trials carried out within the European ATC community in May 2003. The case study presented was conducted as part of the Medium Term Conflict Detection (MTCD) pre-operational validation project, funded by the European Commission and co-ordinated by Eurocontrol. The objective of the project is to validate MTCD as it moves from the development to the pre-operational phase. Specifically the project aims to validate a pre-operational implementation of MTCD in "Passive Shadow Mode", "Active Shadow Mode" and "Advanced Shadow Mode" with different traffic densities and levels of complexity. For this reason three different sessions of trials have been scheduled to be carried out respectively in Malmö (Sweden), Rome (Italy) and Maastricht (Netherlands). This paper refers to the "Advanced Shadow Mode" Trials held in Rome ATC Room in May 2003.

2. SHADOW MODE TRIALS

Shadow Mode Trials are usually performed to test prototype systems in real operational conditions. They represent a way to conduct relevant validation steps for a pre-operational prototype, designed and developed to satisfy operational standards through an incremental development process. There are three main kinds of Shadow Mode Trials, that are defined respectively Passive, Active and Advanced Shadow Mode Trials [3]. These are all carried out in the real operational context, but define different roles and responsibilities for the actors involved (in this paper defined "Trial Team" and "Operational Team").

Passive Shadow Mode (PSM) is characterised by the non-interfering use of the new system and/or platform in the operational context. For this reason there is no interaction between the Operational and the Trial Team. The Trial Team is enabled to listen to the operational R/T communication and to monitor the activity of the Operational Team, without acting on the traffic, or communicating with pilots or other controllers. The main aim of PSM operations is to make controllers acquainted with the prototype system and to figure out how current operations could be achieved.

Active Shadow Mode (represented in Figure 1) is also characterised by the non-interfering use of the new tool and/or platform in the operational context.

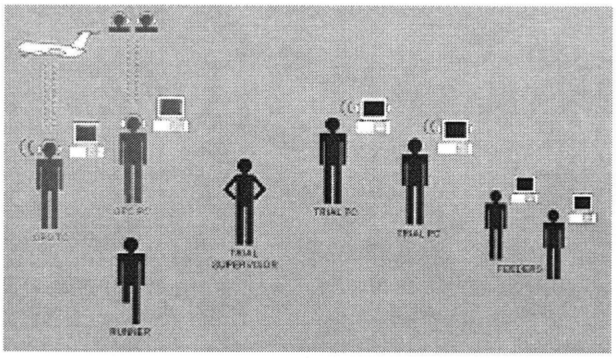


Fig.1 ACTIVE SHADOW MODE ENVIRONMENT

What distinguishes Active Shadow Mode from PSM is the possibility of interaction between the Operational and the Trial Team. As in PSM conditions the Trial Team is enabled to listen to the operational communications and to monitor the activity of the Operational Team without acting on the traffic, or communicating with pilots and controllers. In addition, using the prototype system, the Trial Team can anticipate potential problems and suggest solutions to the Operational Team. This collaboration can be direct, by telephone, or indirect, involving another controller that "runs" between the two teams assuring timely communications. It is important to highlight that in Active Shadow Mode conditions the Trial Team is always off-duty and the Operational Team is responsible for the provision of the ATM service, even in case of application of the Trial Team suggestions. The purpose of Active Shadow Mode is to give controllers the opportunity to appreciate the new functionalities of the prototype system and to compare the performances of the new system with respect to the operational one.

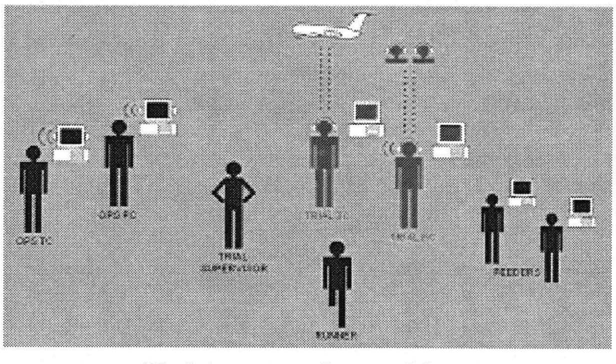


Fig.2 ADVANCED SHADOW MODE

Advanced Shadow Mode, defined also "Hot Shadow Mode" or "Shadow Mode +", adopts a different philosophy. As represented in Figure 2, it is based on the use of the new tool

and/or platform in the operational context, the traffic being entirely managed by the Trial Team. The Operational Team remains as a safety back up, monitoring the communications and the activity of the Trial Team, ready to intervene in case of system failure. The purpose of Advanced Shadow Mode is to eliminate all simulation effects integrating the prototype system in the real context of use.

3. THE CASE STUDY

In this paper we present the first Advanced Shadow Mode Trials conducted within the ATC European community. It concerns the pre-operational validation of a new planning tool for ATC, the Medium Term Conflict Detection (MTCD), whose introduction is expected to bring on radical changes in the standard and traditional traffic management.

Providing the Air Traffic Controllers with the possibility to anticipate the traffic evolution, it is designed to enlarge the temporal horizon of the work, that is commonly set at 10 minutes from the boundary of the controlled sector. With MTCD, this temporal horizon is extended to up 20 minutes. This means that the tool, can support the detection of potential critical situations that may happen in the sector in a temporal range of 20 minutes. This is likely to introduce a radical change in the ATC philosophy, evolving the controller's work from reactive to proactive.

Currently, the controllers are supported by the Short Term Conflict Alert (STCA), that detects cases in which two or more aircraft may collide in two minutes. In a so restricted time frame, controllers do not have enough time to study the traffic configuration and to define a strategy: they may also react applying a quick and effective strategy of conflict avoidance, that not necessarily is also the optimal one. On the contrary, with MTCD the perspective is different. The critical situation may occur in 20 minutes so the controller has time to analyse the situation and to define the optimal strategy to adopt. It is a proactive way to manage the traffic, that is likely to have a deep impact on the controllers' roles and tasks, modifying their activities and their behaviour. Moreover, since each sector is managed by a team of two controllers, working respectively as Planner (PC) and as Tactical (TC), also the cooperative work among them may be affected.

In what follows, we will not discuss the quality of MTCD but the strength and weakness of Advanced Shadow Mode Trials in validating tools like MTCD. Therefore we will concentrate on the adequacy of this kind of validation technique; the effectiveness of the Trials schedule; the correctness of the Validation plan; the safety of the Trials conditions.

The following paragraphs present a detailed description of the MTCD Shadow Mode Trials.

3.1 Trials schedule

The MTCD Shadow Mode Trials lasted 9 working days, from the 6th to the 16th of May 2003 and were preceded by a four-day training (on April, 28, 29 and 30 and May, 5).

The macro-structure of the daily schedule was defined in agreement with the controllers, with the aim to maintain a realistic working pace and to guarantee a sustainable level of workload. According to this:

- each day started with a briefing conducted by the supervisor, in order to provide controllers with information about traffic flows, military areas activation, procedures of the day, etc.
- two runs a day were planned, lasting one hour and a half each. This solution was preferred to the one generally applied during the RTSs, based on three runs a day, lasting one hour each, because considered more coherent with the actual working pace.
- turn-shift every 45 minutes, in order to rotate positions and to allow each controller to maintain a high level of situational awareness and to rest as necessary (in fact, in conditions of high traffic density it would have been too tiring for the controllers to work for one hour and a half).
- two post-session debriefings per day, one after each run, were conducted in the control room, projecting the radar screen dump of the run. This was useful to discuss relevant events on the basis of objective data.
- a focus group per day, lasting around one hour, concerning the MTCD key issues. It was generally based on video clips filmed during the runs and selected in advance by the Validation Team.
- a supervisor interview per day, concerning the main events occurred during the day and their potential impact on the effectiveness of the Trials.

The typical daily schedule proposed in the Validation Plan is represented in the following table. Even if the macro-structure was always applied, several modifications were introduced to this daily schedule in order to deal with traffic flows and local conditions, as for instance sector split due to traffic density.

09.00 – 09.30	Supervisor briefing
09.30 – 11.00	Run
11.00 – 11.20	Post-session Debriefing
11.20 – 11.50	Break
11.50 – 13.00	Focus group
13.00 - 14.30	Lunch
14.30 - 16.00	Run
16.00 - 16.20	Post-session Debriefing
16.20 - 16.40	Supervisor interview

Due to the delicacy of the Trials environment and to the need to have a very high level of mutual and situational awareness among the Trial Team and the Operational Team, it was decided to start each trial session in Active Shadow Mode conditions, in order to allow the Trial Team to get involved in the ATC activity and to acquire the necessary situation awareness. When the operational conditions were judged appropriate by the supervisor, the Advanced Shadow Mode was started. At the end of the trial, the reverse transition was applied to transfer again control back to the Operational Team.

3.2 Trials location

The MTCD Shadow Mode Trials were held at Rome Area Control Centre (ACC), inside the Control Room. The configuration of this room consists of:

- two operational Controller Working Positions (CWPs), based on the OPEN system currently used at Rome ACC and developed by ALENIA;
- two Shadow Mode CWPs, based on the Pre-Operational Validation Environment (PROVE,

www.eurocontrol.fr/projects/prove/), equipped with MTCD;

- two feeder CWPs, based on the PROVE platform and special support tools.

In order to ensure a high level of safety, the operational platform was kept independent from the Shadow Mode platform. This means that the Shadow Mode CWPs were able to receive data from Rome ACC platform, but not to modify data on it. In this way the operational platform was always ready as a safety back up in case of Advanced Shadow Mode CWPs failure.

Figure 3 represents the Control Room layout and the location of the island (circled in figure) where the Trials were conducted. The Trial Team managing the Shadow Mode CWPs was located within the Control Room, not far from the Operational Team managing the Operational CWPs. This choice proved to be extremely effective during the Trials, since the two Teams were able to look at each other and to easily communicate, maintaining mutual awareness.

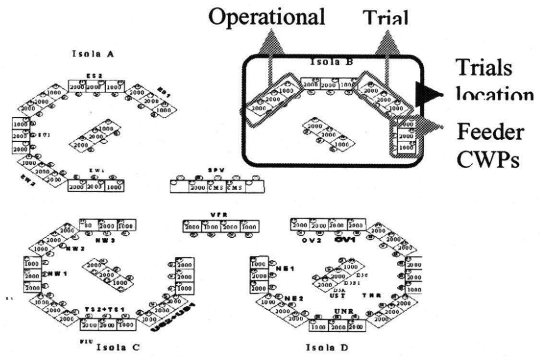


Fig.3 TRIALS LOCATION

3.3 People involved in the Trials

Apart from the Validation Team and the Platform Manager, a Team of eight persons was involved in the Trials:

- a Supervisor from the Italian ATC Service Provider, ENAV (Ente Nazionale Assistenza al Volo) with a large experience of RTSs and experimental environments.
- five air traffic controllers from ENAV, all operational who knew very well the characteristics of the sector chosen for the Shadow Mode Trials. Four of them had been previously involved in RTSs in which the same experimental interface was used. This was a pre-requisite to minimise the effect of a short training on the new platform.
- Two feeders from EUROCONTROL, that is expert personnel managing incoming traffic to the trial sector.

The supervisor was responsible for the Trials conduction, for the management of the transitions, for ensuring that all actors were appropriately briefed and that there was no ambiguity concerning the distribution of roles and responsibilities. He had also to constantly monitor that in Advanced Shadow Mode conditions the Operational Team was correctly duplicating the Trial Team's actions, maintaining a complete awareness of the situation. He was supported by a set of checklist, developed by ENAV and reviewed by EUROCONTROL, listing all the conditions to pass from Active to Advanced Shadow Mode and viceversa. It is important to point out that these transitions could either be "normal" or "forced" i.e. they could take place as part of the trial objectives or be dictated by external events (e.g. failure of the shadow mode (PROVE) platform). A range of potential "failure states" was identified using the HAZOP technique [6] in order to ensure the application of the correct procedures in case of "forced" transition".

The supervisor was also in charge of mediating the communication between the Trial Team and the Operational Team and to support the Validation Team in comparing the conflict detection strategies of the two Teams.

The other controllers alternated in covering the different roles. Two alternative strategies of controllers involvement were discussed:

- make the five controllers work as Operational Team, Trial Team and runner, thus covering all the five positions but risking to overload the controller and make them tired;
- ask the Operational Planner Controller in normal shift (briefly trained on the experimental setting and responsibilities) to keep covering this role, allowing one of the five controllers of the MTCD Team to have a 45-minute break.

The controllers preferred the second strategy. The roles rotation defined with the controllers is shown in Figure 4.

3.4 Roles and responsibilities

Roles and responsibilities in both Active and Advanced Shadow Mode were defined as follows.

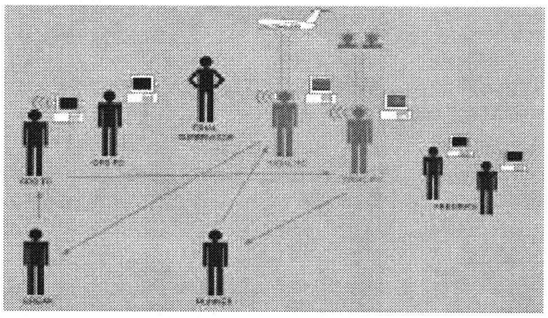


Fig.4 CONTROLLERS ROLES ROTATION

The **Operational Team** was in charge of managing the Operational CWPs, using the Rome ACC operational equipment. At the beginning of each run, in Active Shadow Mode, they were responsible for the normal operational tasks associated to the sector management. On the contrary, after the transition to Advanced Shadow Mode, they had to mirror the actions of the Trial Team, monitoring the activities of the Trial Team and duplicating their actions on the HMI in order to maintain the system updated in case of Advanced Shadow Mode platform failure. The Operational Team was connected to the Trial Team via a dedicated communication line, and to the other sectors of the Rome ACC via the operational LAN.

The **Trial Team** was in charge of managing the Advanced Shadow Mode CWPs, equipped with MTCD. A dedicated communication line linked the Trial Team with the Operational Team. In Advanced Shadow Mode conditions, the Trial Team was responsible for the normal operational tasks associated to the sector management. The Tactical Controller was in charge of communicating with the aircraft and issuing appropriate clearances. Changes made by the Tactical Controller to the flight labels on the HMI did not affect the labels on the Rome operative platform, for this reason this task of actions duplication was assigned to the Tactical Controller of the Operational Team. The Planner Controller

used the telephone to coordinate with planner controllers of other sectors. It was the duty of the Operational Planner Controller to duplicate the Trial Planner Controller's inputs on the operational platform. The Supervisor was in charge of assuring and monitoring that the Operational Planner Controller was correctly duplicating the Trial Planner Controller 's inputs.
At the beginning of each exercise (which, as already stated, started in Active Shadow Mode) the Trial Team had to mirror the Operational Team, in order to be ready for the transition to Advanced Shadow Mode and to detect differences between the way in which the activity is currently carried out, and the way it could be managed with the MTCD support. In case of potential conflicts detected by the MTCD, the Trial Team communicated the criticality to the Operational Team through a dedicated line.

One of the five controllers was requested to work as **runner/feeder assistant controller.** His role mainly consisted in supporting the supervisor, especially in mediating the relationship between Trial and Operational Teams and among them and the feeders. He was responsible for checking the consistency of all flight data and code callsign correlation between Trial the Operational and Trial working positions. This was an essential task to mitigate potential hazards. Being also feeder assistant, he also provided the Feeders with advices on aircraft routings and co-ordination.

The **feeder controllers** were in charge of managing the feeder CWPs, that were respectively correspondent to east and west feed sectors. They played the role of "false receivers" for Exit Flight Level and Exit Point co-ordinations to verify the behaviour of the system in detecting potential problems.

3.5 Validation Objectives

The validation objectives of the MTCD Advanced Shadow Mode Trials included::

- technical usability
- user acceptability
- domain suitability [4].

In addition to these three categories also safety was taken into account and assessed during the validation process..

3.5.1 Technical Usability

Technical usability concerns the perceptual and physical aspects of the HMI such as display formatting as well as anthropometric characteristics of the tool.
According to ISO 9241 "Ergonomic Requirements for Office Work with Visual Display Terminals", technical usability relates to the question of whether a particular system enables specific users to perform their activity efficiently, effectively and satisfactorily in a particular context of use.
As part of the technical usability assessment the following aspects of the interaction with the MTCD were analysed:

- MTCD usability
- coherency of the experimental interface used for MTCD
- controllers performances.

MTCD usability assessment was articulated in two different sessions, conducted respectively before and during the Shadow Mode trial at Rome ACC. A preliminary expert evaluation was conducted before the trial. It was based on the application of discount methods of usability assessment, like heuristic evaluation and cognitive walkthrough [2, 7, 10, 11]. The results were then discussed with controllers. Over-the-shoulders observations and debriefings were adopted during the Trials to explore the relationship between usability and operability. During the Trials the results of the preliminary usability assessment (and their expected impact on the work) were used to guide the observation and the discussion.

3.5.2. User Acceptability

User acceptability concerns the ease of use and suitability of the system in supporting cognitive task requirements. It examines the relationship between the design and the cognitive requirements of the system domain, highlighting the difference between usability and suitability. For example, it is possible for a system to be usable but not suitable for domain tasks, or vice-versa. User acceptability is often one of the factors that determine the success of a technological innovation. In fact a poor user acceptability is likely to result in under-usage or even total rejection of the tool, while a high level of user acceptability is likely to result in large usage, even of system with evident usability defects.
User acceptability was explored in terms of:

- ease of learning
- ease of use
- trust in the system
- benefits perceived
- job satisfaction.

These aspects of the activity were investigated through:

- preliminary expert evaluation,
- participation to the training session
- over-the-shoulders observation during the training and the trial
- post-session debriefings and focus groups;
- interviews with the supervisor.

3.5.3 Domain Suitability

Among the assessment categories adopted, domain suitability was probably the most interesting to be investigated in a Shadow Mode environment. In fact, whilst technical usability and user acceptability can be evaluated through standard user-centred methodologies, domain suitability poses problems with respect to the «cognitive adaptations» of the human activity after the introduction of a particular technology in a robust working environment. The introduction of technological changes and/or innovations produces adaptations in the way this activity is carried out, with a consequent impact on the operational performance. It is assumed that domain suitability is strongly related to the quality of the controllers' activity, affecting aspects of their work as quality of service, capacity and safety.
The assessment of domain suitability requires a deep knowledge of the context of use of the new tool, that is strictly dependent on factors like: the users, their goals; their working methods (including roles allocation and tasks distribution), their subjective strategies in specific situations, their interactions with the working environment, their practices and procedures, the communication with others.
In the MTCD Shadow Mode Trials this knowledge of the context of use was gained through a preliminary session of field observation, whose results were then used as baseline to analyse the controllers behaviour during the Trials. The knowledge of the way the activity is usually carried out in the normal working environment, allowed the Validation Team to detect deviations from the common behaviour and to analyse their possible dependence on the MTCD introduction.
It is important to highlight that a deep knowledge of the tool

was also necessary to evaluate the potential impact on the working environment.

Since controllers cooperation was expected to have a special impact on domain suitability, the following aspects of the their activity were deeply analysed:

- mutual awareness (that is crucial to maintain a high quality of service);
- situational awareness (to avoid situations of event-driven behaviours, with no planning ahead).

Mutual and situational awareness were directly explored during the debriefings, requiring controllers to compare their mental walkthrough of the same situation and to explain their respective strategies.

3.5.4 Safety

In the MTCD Shadow Mode trials, the safety assessment was mainly targeted to analyse the potential impact of introducing MTCD in a robust working environment. Since the impact on safety is strictly dependent on behavioural adaptations due to MTCD, it is evident that a close connection exists between safety and domain suitability assessment.

A key indicator of safety [15, 16] is the overall quality of the system (including environment, procedures, tools and people). However this kind of assessment requires a long lasting monitoring of system performances and therefore was almost impossible to be evaluated during the Trials. Error analysis, that is largely applied to this purpose, is generally effective in cases in which productive processes are well structured and rigid, and the context can be maintained under control. But it is difficult to be applied in ATC systems where the high level of complexity implies that the same actions can be successful or, can result in an accident depending on local and combined situations rapidly evolving. In such situations, it is very difficult to define a correct behaviour ex ante, since the human operator is always adapting to local conditions and past events. Moreover error analysis is generally perceived as putting pressure on the controllers. Thus they feel judged, and instead of helping the analyst in discovering weaknesses in the process, they try to hide them. What was suggested in the MTCD Shadow Mode trials was to consider errors as symptoms of weak areas in the activity, strongly dependent on local conditions and often extremely difficult to reproduce. From this viewpoint, errors were taken as the key measurements of the safety. They were not measurements of incorrect behaviours, but indicators of weaknesses in the working processes. In addition to these techniques, safety assessment was also based on collective envisioning sessions, carried out during the focus groups. Controllers were requested to identify potential hazardous situations, that may be positively or negatively affected by MTCD. They were supported in defining the perceived hazardous situations in terms of (past or future) scenarios, identifying tools, procedures and people involved. These scenarios were means to detect weak areas in the activity, not isolated problems, that are generally extremely difficult to detect during a trial session.

3.6 Validation Techniques applied

As already anticipated in previous paragraph, three classes of validation techniques, namely expert evaluation, direct observation and user feedback collection, were employed at different stages of the validation process, with different aims. The following table summarises the set of validation techniques applied, their aims and the stage of the validation when they were adopted.

Validation techniques	Validation objectives	Phase
Expert evaluation		
Cognitive Walkthrough	Technical usability	Before the trials
Heuristic Evaluation	Technical usability	Before the trials
Direct observation		
Field observation	Technical usability User Acceptability Domain Suitability Safety	Before the trials
Over-the-shoulders observation	Technical usability User Acceptability Domain Suitability Safety	During the trials
User Feedback Collection		
Semi-structured interviews (with the controllers)	Technical usability Domain Suitability	Before the trials
Semi-structured interviews (with the supervisor)	Domain Suitability Safety	After the trials
Debriefings, focus groups	User Acceptability Domain Suitability Safety	After the trials

The validation process was articulated in two main phases strictly interrelated, that were conducted respectively before and during the Trials.

A deep knowledge of the social, cultural and physical aspects of Rome control room was considered crucial to formulate hypotheses on the MTCD context suitability and on its potential impact on the operational working environment. For this reason a preliminary validation session based on field study was organised in order to gather information about the work done in Rome ACC and more in particular in the sector selected for the Trials. The analysis consisted of field observation and interviews with controllers, whose main purpose was to collect information about current working methods and practices, strategies generally followed, tasks distribution between Planner and Tactical controllers, etc. Moreover a session of expert evaluation (based on consolidated cognitive walkthrough and heuristic evaluation) was conducted on MTCD in order to investigate its usability, in relation to the particular context of use of Rome ACC.

The preliminary study allowed to formulate a meaningful set of hypotheses, concerning user acceptability and domain suitability of MTCD that were then checked during the Trials, by means of over-the shoulders observations, debriefings, interviews and focus groups. Over-the-shoulders-observations were based on video and audio recording, whose use had a double purpose. On the one hand they were means to keep objective tracks of the runs, avoiding to found the analysis on biased and subjective data, derived from the analysts' interpretation (or in some cases, misinterpretation). On the other hand they were means to support the controllers discussion. In fact user feedback collection sessions were all based on screen dumps and video clips recorded during the runs and purposely selected depending on the issues to be discussed. On the basis of video recordings, controllers were mainly asked to: discuss about MTCD performances, acceptability and domain suitability; reason about their activity and discuss different strategies; make a comparison among the activity carried out with and without MTCD; explain their behaviour on the basis of personal strategies and experiences.

It is evident that the approach adopted in both phases of the MTCD Advanced Shadow Mode Trials was not new. Instead it was based on a consolidated approach, being generally applied to RTSs and Active/Passive Shadow Mode Trials [8].

What made it different and innovative is the kind of live data to which it was applied. Since the work was done in the field a body of extremely rich and reliable data was collected, that lent to be analysed at different levels:

- phenomenological, concerning the overt human behaviour and patterns of actions;
- cognitive, concerning the cognitive effort associated to the use of the MTCD;
- emotive, addressing aspects of the interactions like satisfaction, engagement, frustration and confusion;
- socio-cultural, concerning the cultural impact of the context of use and the level of support provided in cooperative activities (like knowledge sharing, communication and collective memories).

It is evident that even if the validation techniques were the same of RTSs, the outcomes were different, addressing not only phenomenological and cognitive aspects of the activity (that are typical of RTSs), but also socio-cultural and emotive ones that are very difficult to gather in simulated working environments. In this respect the combination of over-the-shoulders observation and debriefings proved to be a valuable means to come up with this multidimensional validation approach, highlighting personal strategies of dealing with the traffic, emotional triggers and consolidated practices not officially defined but commonly used.

4. RESULTS AND CONCLUSIONS

In the paper we presented the first case of Advanced Shadow Mode Trials carried out within the European ATC community, giving full evidence of the validation work done. As a matter of fact the outcomes of the Trials are very positive from both the safety and the organisational point of view.

In 9 working days, a total amount of 25 hours and 06 minutes were conducted in Shadow Mode. Among these:

- 7hrs16 were spent in Active Shadow Mode;
- 12hrs50 were spent in Advanced Shadow Mode;
- 5hrs were spent in Active Shadow Mode with the sector split in two due to the high traffic density.

In 25 hours only one Trial CWP failure occurred, bearing witness of the robustness of the Trials architecture and organisation. It happened in Advanced Shadow Mode, just after the turn-shift and was mainly due to a defect of data transfer between the experimental platform and the operational communication line of Rome ACC. The failure was immediately and effectively managed putting into evidence the high level of situational and mutual awareness of both Teams. Discussing the events with the controllers involved in the MTCD Team, all of them declared to feel perfectly confident in the Trials conditions, considering the Advanced Shadow Mode environment completely safe, being four controllers managing a sector instead of two. One of the controllers involved said: "*I would have never accepted to work without the support of the Operational Team that was in charge of monitoring my work and mostly the traffic flows to ensure that the experimental platform provided a complete picture of the traffic inbound and outbound*". From an organisational point of view, they also remarked the effectiveness of roles rotation, which allowed them to feel always in control of the situation.

As general conclusion Advanced Shadow Mode Trials proved to be very effective in validating pre-operational tools, as MTCD, to which the case presented in this paper was applied. Overcoming the limitations of simulated environment, this kind of Trials allowed the collect reliable data on the actual use of the tool in the operational environment. Since the tool was used to manage real traffic, controllers were observed to be very focused on their activity and goals. With high traffic density the technology was observed to fade into the background becoming almost invisible [13]. This allowed to observe reliable interactions with the tool, that are expected to be quite similar to those actually performed after introducing the tool in the control room.

Working methods offer an evident example of the possibility to observe credible interactions. During the MTCD Shadow Mode Trials controllers were deeply trained on the working methods designed to use the MTCD components. In particular the idea of MTCD as a planning tool, not a tactical tool, was particularly stressed during the whole training and also in the following days. From a purely theoretical point of view controllers were perfectly aware of it. Nonetheless, after few days of practice a slow evolution of these working methods was observed, according to which some components of MTCD were used as planning tools, while others were used as tactical oriented assistance tools. What is important to highlight is that this evolution was presumably due to socio and organisational aspects of Rome ACC related to the particular distribution of roles and tasks between planner and tactical controllers. But controllers were totally unaware of it, looking amazed when faced with video clips proving evidence of their behaviour.

This example demonstrates the advantages of testing pre-operational tools by means of Advanced Shadow Mode Trials. Nonetheless also some limitations emerged during the MTCD Advanced Shadow Mode Trials as associated to this innovative technique. They mostly refer to the intrinsic context suitability that is necessary to validate the pre-operational tool in the real working environment and to the evident risks associated to this kind of technique. A couple of recommendation results from these limitations, whose main aim is to drive the organisation of future Advanced Shadow Mode Trials, in a completely safe and effective way.

Pre-operational tools must be suitable to the context of use in which they are going to be validated.

This is the fundamental and quite intuitive requirement to conduct Advanced Shadow Mode Trials. If the tool is not adequate to the working environment its user acceptability will be low and the Trials will be tout-court ineffective.

On the contrary, if the pre-operational tool is adequate to the context of use it will be possible to benefit by the advantages already discussed, concerning the reliability of data collected. Moreover a further advantage emerged during the MTCD Advanced Shadow Mode Trials. Since the tool to be validated was coherent with the operational environment in which controllers were used to working, a high level of satisfaction was discovered in the controllers. In general they were very keen to be involved in the Trials, considering them a concrete step towards the enhancement of their current working environment. The Trials location and the possibility to actually use the pre-operational tool to manage the traffic gave them the impression to be actively part of the ATC evolution process. One of them, previously involved in several large scale RTSs, claimed to feel for the first time able to concretely support the research in ATC, and to contribute to the improvement of the controllers' daily work.

Never forget that Advanced Shadow Mode Trials are risky undertakings.

There are two aspects of Advanced Shadow Mode trials that make them risky: on the one hand the architecture and the organisation of the validation environment itself; on the other hand the interaction with the host system. The first issue can be easily faced with a great attention of the staff in charge of the environment preparation to all the aspects of the Trials, from the platform, to the controllers training. From this viewpoint the MTCD case was successful, but it raises questions related to the accuracy of the organisation and the conduction of further Advanced Shadow Mode Trials, once this technique will be consolidated and popularly used to test pre-operational tools in safety critical environment.

The second issue is even more tricky, since it addresses aspects of the Trials related to the interaction between controllers, the experimental platform and the host systems used in the real working environment. It can be exemplified discussing two interrelated aspects. The former is the condition of functional separation that was established between the host system and the experimental platform. In other words the Shadow Mode CWPs were able to receive data from Rome ACC, but could never modify data on the operational system. This condition remained true also from an operational point of view: MTCD tools were considered additional tools to the usual functionalities of a CWP and they were switched off whenever the controllers decided to do so (e.g. controllers often reported to prefer to deactivate MTCD in case of high-density traffic). This possibility of personalisation cannot be assured when tools are shared among many agents, and for managing communication flows. This double-sided segregation should be always ensured when the operational system works as a back up system, otherwise the lack of robustness of the Trial platform may seriously hamper the normal activity in unexpected ways. Nonetheless, it should be highlighted that the back up functionality offered by the host system and the operational controller may provoke an unnoticed “normalisation and acceptance of risk”. Controllers involved in the Trial may start to accept risks simply because they trust the operational system as a back-up system to ensure safety. The more they accept these risks without any dangerous outcomes, the less they consider them as safety relevant. At the end of this process unacceptable risks may be included (almost unnoticed) in everyday activities, unless a strict safety monitoring is put in place.

The second aspect of the interaction between controllers, experimental platform and operative system can be well illustrated by a concrete example. It may happen that the host systems have defects that controllers are perfectly able to manage. During the MTCD Advanced Shadow Mode Trials, for instance, the host systems had problems of ghost tracks in certain areas of the sector: this was due to the replication of the aircraft signal and resulting in the representation of a twin track (named ghost track) with the same code of the original flight, but without the callsign. Generally, being the two tracks quite close a false conflict alert was displayed. Controllers were perfectly aware of it and were used to disregard these phenomena, that disappeared automatically in a while. Since the two platforms shared the same data source this problem was present also in the experimental platform, where ghost tracks and false conflict alerts were displayed. What is important to highlight is that in some cases different events were misinterpreted as cases of ghost tracks, whilst they were due to feeders’ lack of aircraft correlation. The risk associated to this kind of events is evident. Interpreting the new problem (lack of aircraft correlation due to the feeders) as known problems (ghost tracks and subsequent false conflict alert), controllers took as false a potential conflict that was instead real and particularly risky, because involving an unidentified flight, whose callsign and destination were unknown. Such cases should be studied and documented in advance in order to decide whether the Advanced Shadow Mode is worth or not, and in case it is, to train the controllers as necessary.

5. ACKNOWLEDGMENTS

A special thanks goes to EUROCONTROL and ENAV for having made it possible to set up and run the first experiment of Advanced Shadow Mode Trial in Europe. Their collaboration was a key element for the realisation of the work described in this paper.

6. REFERENCES

[1] Cardosi, K. M., Murphy, E.D. (1997), Human Factors Checklist for the Design and Evaluation of Air Traffic Control Systems, DOT/FAA/RD-95/3.1. Research funded by the Federal Aviation Administration’s Office of the Chief Scientific and Technical Advisor for Human Factors (AAR-100).

[2] Cockton, G., Woolrych, A. (2002), Sales Must End: Should Discount Methods be Cleared off HCI’s Shelves?, Interactions Vol 9, Issue 6, pp. 13-18.

[3] EUROCONTROL, (2001). A methodology for validating EATCHIP III MTCD using shadow mode trials. www.eurocontrol.be

[4] Halverson, C. (1995), Inside the Cognitive Workplace: New Technology and Air Traffic Control, San Diego: University of San Diego, Doctoral Dissertation.

[5] Hutchins, E. (1995). Cognition in the Wild. MIT Press.

[6] Kletz, T. (1999) Hazop and Hazan: Identifying and assessing process industry hazards; Hemisphere Publications, 4th Edition.

[7] Lindgaard, G. (1994), Usability testing and system evaluation, London: Chapman & Hall.

[8] Marti. P., Lanzi, P., Pucci, F. (2001). Evaluating safety and usability of ATM systems. Proceedings of the 4th International seminar on ATM R&D, FAA and Eurocontrol, Santa Fe, USA, December 2001.

[9] Marti, P., Scrivani, P. (2003). The representation of context in the simulation of complex systems. Cognitive Technology, vol.8, n.1, spring 2003, pp. 32-42.

[10] Nielsen, J., Landauer, T.K. (1993). A mathematical model of the finding of usability problems. Proceedings of ACM/IFIP INTERCHI’93 Conference (Amsterdam, The Netherlands, April 24-29), pp. 206-213.

[11] Nielsen, J., Molich, R. (1990), Heuristic Evaluation of User Interfaces. Proceedings CHI ’90 Conference on Human Factors in Computer Systems, ACM, NY.

[12] Nardi, B. (1996). Context and Consciousness: Activity Theory and Human Computer Interaction. Cambridge, MIT Press.

[13] Norman, D. (1998). The invisible computer. The MIT Press Cambridge, Massachusetts.

[14] Norman, D. and Draper, S.W. (eds.) (1986). User Centred System design: New perspectives on Human-Computer Interaction, Hillsdale NJ: Lawrence Erlbaum Associates.

[15] Reason, J. (1990). Human Error. Cambridge University Press, New York.

[16] Reason, J. (1997), Managing the Risks of Organisational Accidents, Aldershot, UK: Ashgate Publishing company

Evolutionary Optimization of Graphical Human-Machine Interfaces for Controlling Technical Processes

Dipl.-Ing. Andreas Völkel
University of Kassel
Mönchebergstraße 7
34123 Kassel/Germany
a.voelkel@uni-kassel.de

ABSTRACT

In this paper a new methodology for optimizing graphical human-machine interfaces (HMI) of industrial processes will be presented. This methodology uses abstractions of biological evolutionary mechanisms to adapt user interfaces to the needs of individual operators.

Keywords

Human-Machine Interfaces, Cognition Related Process Visualization, Evolutionary Optimization.

1. INTRODUCTION

The development of industrial processes comprises more complexity and, thus, an increasing strain for the operators of such plants. Additionally, there is an increasing level of automation due to increasingly powerful and cheaper automation hardware. One result is the decreasing number of operators for driving a process with concurrent rise of information to be observed by each separate operator [1],[2]. The raised automation level involves a loss of manual skills in guiding a process while the responsibility for the operators increases [3],[4]. You can say that communication between human and machines should be designed preferably understandable and transparent [5]. The machine should be adapted to human needs and not vice. An appropriate configuration of the human-machine interface, especially of the systems which perform the information transfer, can achieve the adaptation intended.

With the designed method process visualizations can be adapted to individual operators. Conventional and cognition related views will be used as a basis for the optimization. In the first optimization step the cognition related views gains an adaptation of the visualization to the universally valid needs of man in information processing. The second new step adapts the process visualization to the needs of an individual user by involving the user's actions and evaluations of the Graphical User Interfaces (GUI) as part of the Human Machine Interface (HMI).

2. USED MODELING TECHNIQUES

The first step for optimizing the graphical user interface is to adapt the representation of information to human needs. For that purpose, several additional visualizations (MFM, EID, Virt3D, see below) will be provided for the users beside the common modeling technique (TOP).

2.1 Topological View – TOP

The topological view (TOP) is the standard visualization of industrial processes. It is well known and used in many hierarchical representations. This view is equal to the whole-part-decomposition of systems to different subsystem levels. At the lowest level, there are the components of the system, at the highest the entire system. The whole-part-model of systems is connected to the topologically orientated description which yields different zooms in the graphical user interface. Figure 1 shows the symbols for the topological view and an example of a possible graphical user interface.

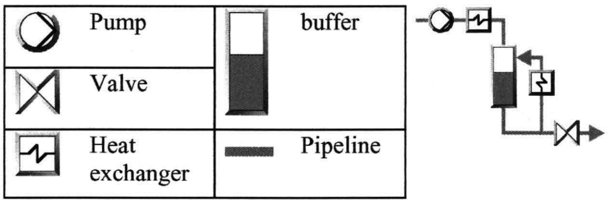


Figure 1: Topological Symbols and an example

2.2 Virtual 3D Process Visualization -Virt3D

The Virtual 3D Process Visualization (Virt3D) developed by Wittenberg [6] owns the same structure as the topological view. The difference lies in the graphical representation of the components in the interface. Whereas the symbols of the topological model already mean an abstraction, the Virtual 3D View shows the components very close to reality. This includes the representation of component connections as they are in reality. Figure 2 shows the virtual objects and the example from figure 1 in the Virtual 3D View.

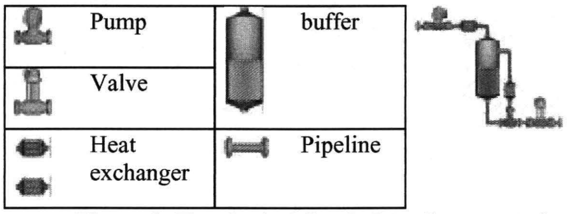


Figure 2: Topological Symbols and an example

2.3 Ecological Interface Design - EID

The Ecological Interface Design (EID) was developed by Vicente [7]. With EID it is possible to build cognition related human-machine interfaces. The basis for EID are the action models for human information processing by Rasmussen [8]. They allow the design of a function-orientated interface which support the user in unexpected and unpredictable situations of the system.

The behavior of a process can mostly be described by explicit physical laws. However, for modeling a system by EID, first of all a system has to be described by several equations for the interested values. The equations in figure 3 exemplarily indicate the generation of the mathematical description of a system.

This knowledge about a process is visualized in EID by geometric figures of mathematical relations between the relevant process variables. These mappings show the influence of the different variables in a particular convenient manner. Users can extract rules which result in appropriate actions. By this kind of representation cognitive functions of humans are better and more easily activated and their capacity to act improves. Figure 3 shows the construction for an EID model and again the example.

Mass balance	Temperature	Energy balance
$\frac{dh}{dt}=\frac{1}{\rho A}\left(\sum_{i=1}^{N_{in}}\dot{m}_{in_i}-\sum_{j=1}^{N_{out}}\dot{m}_{out_j}\right)$	$T=\frac{Q(t)}{A\,h(t)\,\rho\,c_p}$	$\frac{dQ}{dt}=A\rho c_p\left[(T-T_0)\frac{dh}{dt}+h(t)\frac{dT}{dt}\right]$

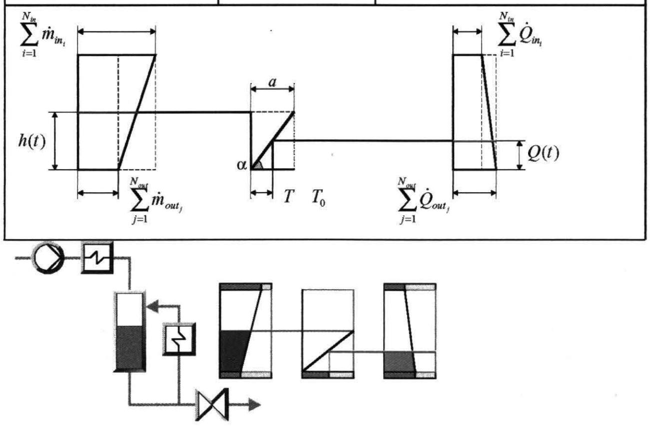


Figure 3: EID Modeling and example

2.4 Multilevel Flow Modeling – MFM

The basis for the Multilevel Flow Modeling MFM suggested by Lind [9] is the idea to model a system as an artifact. MFM deals with the representation of a system by multiple descriptions at different abstraction levels.

Symbol	Indication
●	Goal
◉	Source
◈	Transport
⊗	Sink

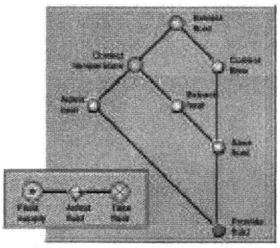


Figure 4: MFM-Modeling (Part) and example

A system is described by terms of goals, functions, and physical components. At the same time, each of these descriptions can be treated on different hierarchical levels of decomposition. The goals are the intentions of the designer the system was made for, or the purpose of the system. For the modeling of the system's functions, simple functional perceptions are used which relate to the control of linked mass, energy or information flows. The suggested functions are sources, transports, balances, storages, barriers, sinks, and, in addition only for information flows, observers, decisions, and actors. All types are realized by symbols as part of a graphical representation language. They can link together into structures in consideration of different syntax rules. The symbols do not represent physical components as in common representations, but show functional coherences. The physical components realize the flow functions in MFM and, by implementation of the functions, a goal can be fulfilled. According to Larsson [10], MFM shows the functional structure of a system as a number of flow structures relating to each other at different abstraction levels. The abstraction levels again are connected by different relations. Figure 4 shows partly the symbols of the MFM objects and their indications.

3. THE STRUCTURE OF GRAPHICAL USER INTERFACES

Before considering the evolutionary optimization of the graphical interfaces their objects have to be described in a certain structure. This allows the relation of the different objects of the interface to each other. The defined terms are *objects*, *object types*, *object classes* and *levels*.

3.1 Objects and Object Types

Object types are the constituents of the modeling techniques introduced in the sections above. They have different attributes and different graphical representations. For example one object type of the topological view is the heat exchanger. Objects are the instantiations of the object types. So a concrete heat exchanger for e.g. heating a fluid would be an object.

3.2 Object Classes

The object types are combined to object classes to get a superior relation for their evaluation. The heat exchanger e.g. belongs to the component class. Further classes are:

Table 1: Object classes of graphic user interfaces

Object class	Description
Systems	Highest level in the whole part model
Subsystems	Intermediate levels in the whole part model
Components	Equipment of the system, e.g. pumps, heat exchangers, etc.
Mappings	Mass and energy balances of the EID modeling
Goal hierarchy	Goal tree of the MFM modeling
Goal	Single goal of the MFM goal tree
Flow structure	Combination of the MFM flow functions
Flow function	Single Function of the MFM modeling

3.3 Levels

At the *main* level there is the entire system, i.e. in this level the system is shown in a comprising view. TOP and Virt3D show at this level an overview about all the equipment of the system. Additionally for EID we have the mappings, for MFM there is the hierarchy (see also table 2). At the second level (*sub*) there are the next stages in zooming the system. According to the system's complexity different sub levels may exist. At the *interaction call* level there are the objects of section 3.1 like the heat exchangers for heating the fluid. This level is called *interaction call level* because users can call the objects for performing interaction (e.g. choosing the component of a heat exchanger to gets his control buttons) here. At the lowest *interaction perform* level information out of the system will be shown and inputs of the users will be transferred to the system. The table 3 shows the levels of graphical user interfaces.

Table 3: Levels of graphical user interfaces

Level	**Model**				
	Virt3D	**TOP**	**EID**		**MFM**
Main	System	System	System	Mapping	Hierarchy Goals
Sub	Sub-system	Sub-system	Sub-system		Flow-structures
Interaction call	Components				Flow-functions
Interaction performance	Controls, displays				

The evolutionary optimization only deals with the first three levels, because the fourth level is the same for all four used modeling techniques.

4. EVOLUTIONARY OPTIMIZATION OF GRAPHICAL USER INTERFACES

During the second step of the optimization, the information representation adapts to the needs of the individual operator by using an evolutionary algorithm. The natural evolution is the basic model for this kind of optimization [11]. Evolutionary algorithms are stochastic search techniques which work with populations of individuals [12]. The principle of the "*survival of the fittest*" is applied to the individuals to gain better individuals with regard to a goal function.

The developed *EOGUI*-algorithm (*E*volutionary *O*ptimization of *G*raphical *U*ser *I*nterfaces) is applied to combinations of all four different representations. The result of the algorithm is dependent on the user preferences in controlling the process, e.g. engineers may prefer to work with EID, operators with TOP[4]. Evaluations in the optimization are performed by different quality factor, i.e. objective functions Ψ must be calculated. These objective functions refer to the mentioned user preferences according to the process visualization. However, the better the fitness of an object is the higher the survival probability of the object. The fitness of an object depends on the actions which the user performs with it and on his subjective sense of the object. This leads to a maximum problem for this optimization $f(\Psi)$? max.

4.1 The *EOGUI*-Algorithm

To perform an evolutionary optimization the evolutionary algorithm EOGUI was developed by the author. This algorithm concludes the objects of the four used graphical interfaces in one newly generated interface by processing the objective actions and subjective evaluations of the user. The objects (individuals) of the interfaces (populations) "fight" for admission into the new interface. Figure 5 shows the structure of the optimization. The decision which individuals survive and which have to die depends on the interactions and evaluation of the user mentioned above. The user has to accomplish a training phase where he is asked to perform some task scenarios. In the course of this scenarios all user actions are recorded in log-files. At the end of the scenarios the online questionnaire occurs with questions to the user's appreciation concerning e.g. transparency and navigation. This procedure is performed until a canceling criteria (e.g. user message) is achieved. In that case the best individuals are reached and optimization is finished.[13] During normal operation the EOGUI algorithm should not be activated to avoid confusion of the users about changing graphical displays by performing the evolutionary optimization.

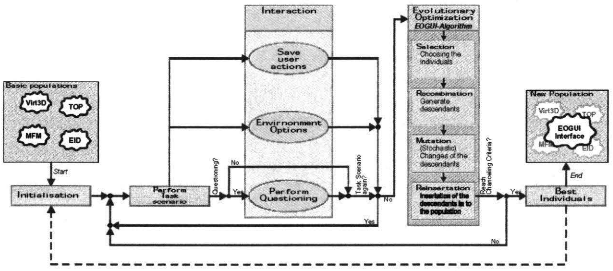


Figure 5: Structure of the optimization

4.2 Selection

The selection complexor is related to the interaction of the user with the interface mentioned above. Which objects are allowed to generate descendants or taken over to the new interface is decided here. The first step is the allocation of the fitness, the second is to select the objects referring to their fitness. Objects owns two parts of fitness: an objective part by measurable data (objective adaptation) and a subjective part which is derived from the evaluations out of the questionnaires (subjective adaptation).

4.2.1 Fitness Allocation

How to allocate the evaluations to the objects is part of this section. From the classification in chapter 3 we can derive the two domains and four relation types for the criteria described below.

The domains *local* and *global* describe the range of values on which the evaluation of a criterion will be related:

Table 4: Domains of Fitness allocation

Global	Relation of the evaluation of an object to the total value of all objects of the same level of *all* interfaces
Local	Relation of the evaluation of an object to the total value of all objects of the same level of *one* interface

By using these domains competition between objects of one interface (local) and of objects of different interfaces (global) are considered. The relation types describe how evaluations are related to other object evaluations. For example topological and Virt3D views of the pump component can be compared. Further standardizations can be seen in table 5.

Table 5: Relation types of Fitness allocation

Absolute	No relation for object evaluation. Examples are the assortment frequency or the assortment time of an object
Object type	Relation to objects evaluation of the same type. This shows the ranking of an object evaluation according to the objects of the same type (e.g. all heat exchangers).
Object class	Relation to object evaluation of the same class (e.g. all components)
Level	Relation to object evaluation of the same level showed in table 3 (e.g. the interaction call level)

There is no coherence between the local domain and the level standardization as well as between the global and absolute one. Both relation would make no sense at all. The value of an objective function Ψ for an Object O_i could be calculated as shown in table 6:

Table 6: Goal functions – relations and domains

Total: $\psi_{crit}(O_i)=\psi_{crit}^{L,A}(O_i)+\psi_{crit}^{L,T}(O_i)+\psi_{crit}^{G,T}(O_i)+\psi_{crit}^{L,C}(O_i)+\psi_{crit}^{G,C}(O_i)+\psi_{crit}^{G,L}(O_i)$		
	Domain	
Relation	*Local*	*Global*
Absolute	$\psi_{crit}^{L,A}(O_i)=crit(O_i)$	-
Object type	$\psi_{crit}^{L,T}(O_i^{T_p})=\frac{crit(O_i^{T_p})}{\sum_{j=1}^{N_{O^{T_p},M_i}} crit(O_j^{T_p})}$	$\psi_{crit}^{G,T}(O_i^{T_p})=\frac{crit(O_i^{T_p})}{\sum_{j=1}^{N_M}\sum_{k=1}^{N_{O^{T_p},M_j}} crit(O_{j,k}^{T_p})}$
Object class	$\psi_{crit}^{L,C}(O_i^{C_q})=\frac{crit(O_i^{C_q})}{\sum_{j=1}^{N_{O^{C_q},M_i}} crit(O_j^{C_q})}$	$\psi_{crit}^{G,C}(O_i^{C_q})=\frac{crit(O_i^{C_q})}{\sum_{j=1}^{N_M}\sum_{k=1}^{N_{O^{C_q},M_j}} crit(O_{j,k}^{C_q})}$
Level	-	$\psi_{crit}^{G,C}(O_i^{E_r})=\frac{crit(O_i^{E_r})}{\sum_{j=1}^{N_M}\sum_{k=1}^{N_{O^{L_r},M_j}} crit(O_{j,k}^{E_r})}$

4.2.2 Objective Adaptation

The criteria or objective functions ε of the objective adaptation are the assortment frequency and the assortment time of objects as well as the efficiency and accuracy of user actions.

4.2.2.1 Assortment Frequency and Time

The assortment frequency is the number of actions η with an object performed by a user of the interface. These are the numbers of object selections, control actions or display calls. The assortment time means the length of time an object is selectable.

$$\varepsilon_{afq}=\sum_{j=1}^{N_\eta}\eta_j(O_i) \qquad \text{Assortment frequency}$$

$$\varepsilon_{atm}=\sum_{k=1}^{N_t}\left[t_{k_{end}}(O_i)-t_{k_{start}}(O_i)\right] \qquad \text{Assortment time}$$

4.2.2.2 Efficiency of action

The efficiency ε_{eff} of action is calculated with the optimal numbers of actions η^{opt} related to the real numbers of action η and the optimal assortment time t^{opt} related to the real assortment time *t*. The optimum value is 1 for both. So the best value will be reached if a failure situation can be handled by just one action in one time unit. Obviously this never happens, but the less action and time is needed to solve a problem, the closer the efficiency evaluation is to the optimum.

$$\varepsilon_{eff}(O_i)=\frac{1}{2}\left[\frac{\eta^{opt}}{\sum_{j=1}^{N_\eta}\eta_j(O_i)}+\frac{t^{opt}}{\sum_{k=1}^{N_t}\left[t_{k_{end}}(O_i)-t_{k_{start}}(O_i)\right]}\right]$$

4.2.2.3 Accuracy of actions

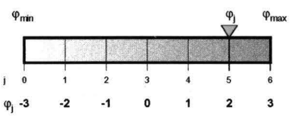
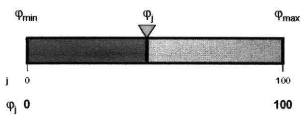
The accuracy of the users action following a preset desired value for a certain time will be acquired by this criterion. The closer a state variable is to his desired value by performing adequate actions the better is the evaluation.

$$\varepsilon_{acc}(O_i)=\frac{1}{t_{end}(O_i)-t_{start}(O_i)}\sum_{i=1}^{t_{end}-t_{start}}\left(1-\frac{\left|x^{soll}(O_i)-x^{ist}(O_i)\right|}{x^{soll}(O_i)}\right)$$

4.2.3 Subjective Adaptation

Via the subjective adaptation not only the objective measurable evaluations are incorporated, but also the subjective impressions of the user. To get this data the user has to answer some questions by using different scales, which occurs directly after getting the objective data. Table 7 shows the used scales and the calculation of the subjective functions φ.

Table 7: Subjective adaptation: goal functions and scales

Seven-point scale	Percent scale
	
$\zeta_{crit}^{7P}(O_i)=\frac{1}{N_{\zeta_{crit}}}\sum_{j=1}^{N_{\zeta_{crit}}}\frac{\varphi_j^{M_r}+\varphi_{j_{\max}}^{M_r}}{\left\vert\varphi_{j_{\min}}^{M_r}\right\vert+\left\vert\varphi_{j_{\max}}^{M_r}\right\vert}$	$\zeta_{crit}^{\%}(O_i)=\begin{cases}\frac{1}{N_{\zeta_{crit}}}\sum_{j=1}^{N_{\zeta_{crit}}}\frac{\varphi_j^{M_r}}{\varphi_{j_{\max}}^{M_r}} & \text{fun, confidence}\\ 1-\frac{1}{N_{\zeta_{crit}}}\sum_{j=1}^{N_{\zeta_{crit}}}\frac{\varphi_j^{M_r}}{\varphi_{j_{\max}}^{M_r}} & \text{strain}\end{cases}$

The following itemization describes the used subjective criteria.

- *Transparency*: The perceptibility of functional coherences between components defining the transparency according to Dutke [14]. The user shall say if the interface informed him at every time clearly about the system's state.
- *Navigation*: Ability of orientation in the technical system. Freiburg [15] identified navigation as finding the way by following logical relations between the objects. The questions is, how easy could the user find information and system components?
- *Error management*: This means according to Johannsen [5] the detection, diagnose, and correction of abnormal situations. What are the problems in finding, categorize and correct respectively compensate an error?
- *Symbolics*: The user is asked about the meaning and the recognizability of the different symbols.
- *Structure*: The user should say if one modeling technique supported him better than the other, as a result of its structure.
- *Strain:* It is questioned to the subjective impression of strain during the scenarios. According to Frieling and Sonntag [16] strain is the subjective reaction of the human organism to an objective acting impact.
- *Confidence:* This term originally comes from psychology, but Muir [17] transformed this term to man-machine communication. One important aspect of confidence is the expectation of the behavior of our environment. The knowledge about the behavior could lead to a capability of making predictions or appreciation future states and occurrences. According to operators of graphical user interfaces the presentation of relevant process information is very important to produce the mentioned predictions.
- *Fun:* This criteria is a very import success indicator. The needs and affinities of user should be considered by designing user interfaces. The less the evaluation of this criterion is the higher is the dislike of the user to the according interface.

4.3 Recombination

Recombination appoints the parents which are selected for generating descendants [18]. Recombination in graphical user interfaces is only applied if the fitness of the objects to be recombined holds comparatively high values, i.e. recombination only would be performed if the fitness values of different objects did not exceed a certain value.

Two types of recombination are designed: the recombination of the symbol form and the recombination by object combination.

4.3.1 Recombination of the Form

This kind of recombination is applied to the inner and outer form of the different symbols. The outer form depends on the object with higher fitness, the inner form is generated by combining the inner forms of both parent objects. Figure 6 shows an example for the recombination of a MFM source and a TOP heat exchanger.

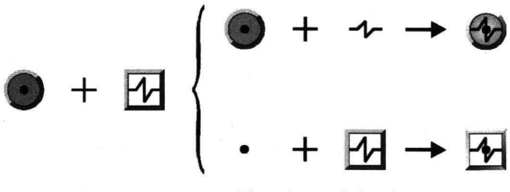


Figure 6: Recombination of the Form – Example

The object orientated recombination refers to single objects (e.g. one source and one heat exchanger), whereas all objects of the according types are taken for the object type orientated recombination (e.g. all sources and all heat exchangers). In the first case the fitness value of the single objects decides about recombination, in the second case the mean value over all object of the according type.

4.3.2 Recombination by Object Combination

With the recombination by object combination two objects will not be "melted" but completely taken by connecting them. This kind of recombination mainly concerns the EID mappings and the MFM goals and flow structures as well as the system object from TOP, Virt3D, and EID.

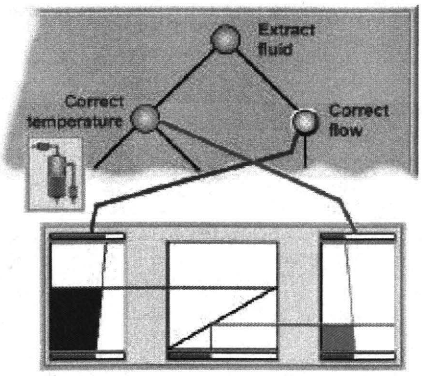


Figure 7: Recombination by combination – Example 1

MFM separates mass and energy flows. EID visualizes system information in graphs, which can be traced back to mass and energy balances. On this basis MFM and EID can be connected. A further possibility of recombination is adding the subsystem objects to the relating goals, i.e. the subsection which supports reaching the goal will be shown in the goal hierarchy. The same recombination is possible for the object classes of flow functions and components. Components realise flow functions, this connection is used for this kind of recombination. Figure 8 shows an example for the object recombination by combination. Here the pump and the valve are realising the transport function *admit fluid*.

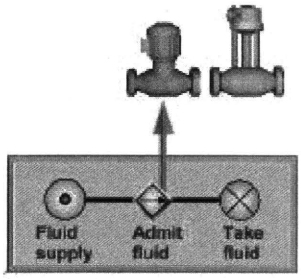


Figure 8: Recombination by combination – Example 2

4.4 Mutation

Mutation changes the descendants stochastically, though with low probability. With the mutation, random gets into the optimization of graphical user interfaces. Selection and recombination are more purposeful, because they are derived form the user's actions. In case of graphical user interfaces, mutation changes the graphical attributes of the objects. Table 7 gives an overview about the mutation possibilities of the EOGUI algorithm.

Table 7: Mutation - Possibilities

Attribute	**Object class**							
	Sys	Sb	Cp	Mp	Gh	G	Fs	Ff
Size	-	-	*obj* *ind*	*obj* *lin*	*lin*	*obj*	-	*obj* *ind* *lin*
Color	*bgr*	*bgr*	*ind*	*lin* *fil*	*bgr* *lin*	-	*bgr*	*lin*
Animation	*obj*	*obj*	*obj*	-	-	*obj*	-	*obj*
3D	-	-	*obj*	*obj*	-	*obj*	-	*obj*

Sys: System
Sb: Subsystem
Cp: Component
Gh: Goal hierarchy
G: Goal
Fs: Flow structure
Ff: Flow function

bgr: Background
fil: Filling EID Mapping
ind: Indication
lin: Line
obj: Object

4.5 Reinsertation

The reinsertation takes the descendants and/or the parents back into the population. For the graphical user interfaces the objects chosen in the selection, generated by recombination, and possibly changed by mutation are reinserted in the new population, i.e. to the new EOGUI interface. So both parents and descendants could be selected for the new interface. However, after the reinsertation the optimization is closed for the moment, but could be started again on demand.

5. AN APPLICATION: MIPS

For testing the new method of the evolutionary optimization of graphic user interfaces, the Mixture Process Simulation (MIPS) was developed. Figure 9 shows the structure of this process. In the left part there is the subsystem SUB01. Here a input of fluid will be delivered to the process. The upper right part contains the subsystem SUB02, where another input conveys a powder to the subsystem SUB03. This subsystem mixes the fluid and the powder. Finally the subsystem SUB04 is the output of the produced mixture.

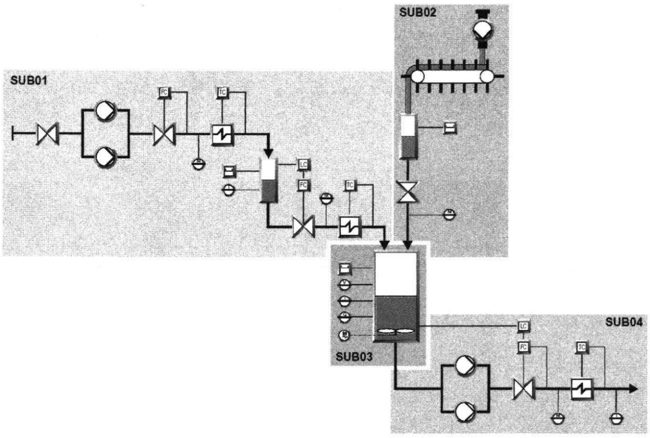


Figure 9: An application for the EOGUI-algorithm: the Mixture Process Simulation (MIPS)

The constraints of this process are to maintain certain mass flows, fill levels, temperatures and the quality of the mixture. For this process four graphical user interfaces were developed and connected to the EOGUI-Algorithm to perform experiments.

6. EXPERIMENTS

In a comparing study the effectiveness of the developed method will be examined in the very near future. An experimental environment shown in figure 10 were build to perform the experiments.

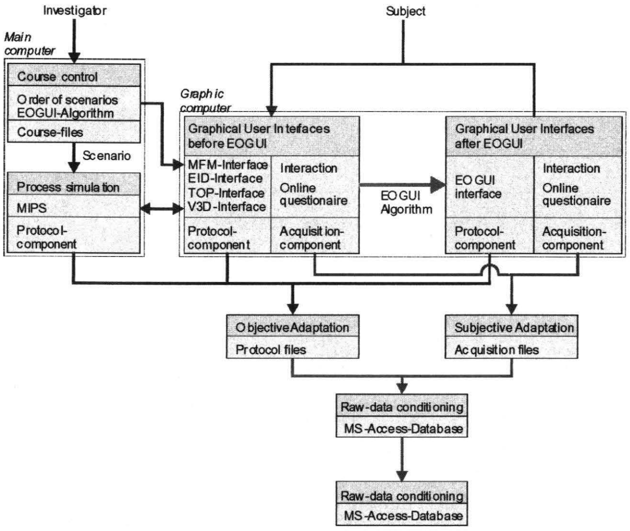


Figure 10: Experimental environment

6.1 Variables of the experiment

The independent variables are the two user interfaces, the objects, the subjects, and the scenarios. The interfaces are the visualisation before the optimization (MFM, EID, TOP, and Virt3D), and after performing the optimization (EOGUI). To make the objects of the interfaces comparable, the levels of section 3.3 are used. The scenarios are the tasks for the subjects in the experiment. First they have to bring the system from zero state to normal state with no time constraint. Second three different failures occur which have to be corrected by the subjects.

The independent variables are the objective and subjective criteria mentioned in section 4.2 and additionally questions to the impressions about the recombination and mutation results.

6.2 Hypothesis

The question that needs to be answered is if there is an approvement for handling technical processes by using evolutionary optimizations for graphical interfaces. It is expected that by using the EOGUI algorithm subjects actions will be more effective and more accurate. Even for subjective evaluation better values are expected. To prove the statements mathematical zero hypothesis H_0 and a alternative hypothesis H_A will be formulated. In case of the zero hypothesis H_0, no differences in the mean values of the according dependent variables μ are expected, but for the alternative hypothesis H_A, significant differences are expected. By using non parametric Mann-Whitney Tests decisions will be made whether hypothesis are accepted or denied.

6.3 Results

Preliminarily experiments with the EOGUI algorithm showed different results for different people as expected. The figures 11 and 12 show parts of some of the preliminarily results.

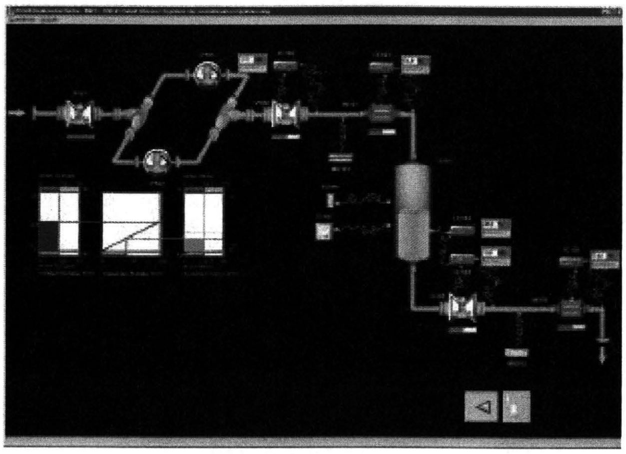


Figure 11: Result of the EOGUI-algorithm – Example 1

Figure 11 shows the result of the EOGUI-algorithm for the subsystem SUB01 of MIPS. This user prefers mainly the Virt3D-Model for the visualizations of pipelines, buffers, etc. But his values for the TOP visualizations of the valves and pumps were high enough to execute an object type based recombination.

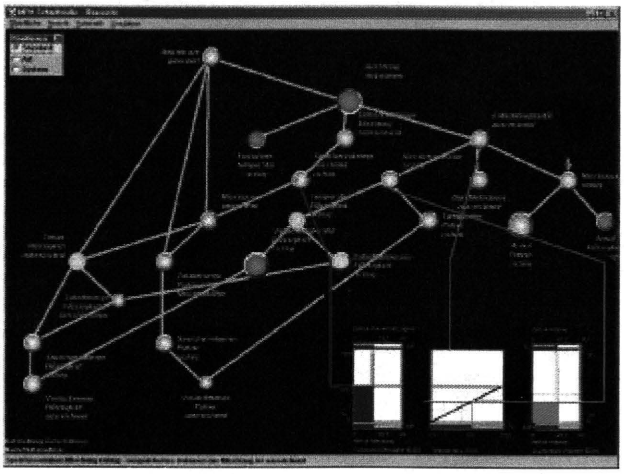


Figure 12: Result of the EOGUI-algorithm – Example 2

The second example shows the result for an better evaluation of the MFM Model. Figure 12 shows the main view with the goal hierarchy of the MFM Interface with mutation of the object size.

This generated graphical interface contains additionally to the MFM goals some EID mapping objects. These objects are the result of a recombination orientated on object combination. The size and colours of the MFM goal symbols and indications change due to mutation.

7. ACKNOWLEDGMENTS

My thanks to the DFG (German Research Foundation) for promoting this project.

8. REFERENCES

[1] Thierfelder, H.G.: Mit neuester Technik bestehende Kraftwerke modernisieren. ABB Technik 2 (1997), S.15-24.

[2] Herbst, L., Rieger, W.: Modernste Prozessvisualisierung im Kraftwerk Schkopau. ABB Technik 1 (1997), S. 13-18.

[3] Bainbridge, L.: Ironies of Automation, Automatica, 19 (1983) H. 6, S. 775-779.

[4] Völkel, A.: Experimentelle Untersuchung kognitionsbezogener Prozessvisualisierungen in Ziel-Mittel-orientierten Mensch-Maschine-Schnittstellen - Ergebnisse. Bericht IMAT-MMS18 ISSN 0940-094X, Fachgebiet Systemtechnik und Mensch-Maschine-Systeme, Institut für Mess- und Automatisierungstechnik (IMAT), Universität Kassel 2001.

[5] Johannsen, G.: Mensch-Maschine-Systeme. Springer-Verlag, Berlin u.a. 1993.

[6] Wittenberg, C.: Virtuelle Prozessvisualisierung, Dissertation, Universität Kassel, Institut für Mess- und Automatisierungstechnik (IMAT), Fachgebiet Systemtechnik und Mensch-Maschine-Systeme (2000).

[7] Vicente, K. J., Rasmussen, J.: The ecology of human-machine systems II: mediating direct perception in complex work domains, Ecological Psychology, Vol. 2, Nr. 3 (1990), S. 207-249.

[8] Rasmussen, J.: Skills, rules and knowledge: signals, signs, and symbols, and other distinctions in human performance models. IEEE Transactions on Systems, Mans, and Cybernetics, 13 (1983), S. 257-266.

[9] Lind, M.: Representing goals and functions of complex systems: An introduction to Multilevel-Flow-Modeling, Technischer Bericht 90-D-381 ISBN 87-87950-52-9, Institute of Automatic Control Systems, Technical University of Denmark (1990).

[10] Larsson, J. E.: Knowledge-Based Methods for Control Systems. PhD-Thesis. Department of Automatic Control, Lund Institute of Technology (1992).

[11] Schwefel, H.-P.: Evolutionäres Rechnen in der Computational Intelligence. In: Computational Intelligence: Neuronale Netze, evolutionäre Algorithmen, Fuzzy Control im industriellen Einsatz, VDI-Berichte (S. Hafner, H. Kiendl und H.-P. Schwefel, Hrsg.). VDI-Verlag, Düsseldorf, VDI-VDE-GMA-Fachtagung 3.-4. März 1998, Vol. 1381, S. 19-23, 1998.

[12] Bäck, T., Hammel, U., Schwefel, H.-P.: Modelloptimierung mit evolutionären Algorithmen. In: Fortschritte in der Simulationstechnik. In: 8. Symposium Simulationstechnik der Arbeitsgemeinschaft Simulation (ASIM) Band 6 (A. Sydow, Hrsg.). Vieweg-Verlag, Braunschweig, S. 49-57, 1993.

[13] Völkel, A.: Cognition Related Process Visualization with MFM and EID: Improvement or not?. In: 8th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kassel, Germany, 18-20 September, S. 657-662, 2001.

[14] Dutke, S.: Mentale Modelle: Konstrukte des Wissens und Verstehens. Verlag für angewandte Psychologie 1994.

[15] D. Freiburg, Ergonomie in Dokumenten Retrieval Systemen. Walter de Gruyter, Berlin 1987.

[16] Frieling, E., Sonntag, K.: Arbeitspsychologie. Verlag Hans Huber, Bern, Göttingen, Toronto, Seattle 1999.

[17] Muir, B. M.: Trusts between humans and machines, and the design of decision aids. Int. Journal of Man-Machine Studies, 27 (1987), S. 527-539.

[18] Pohlheim, H.: Entwicklung und systemtechnische Anwendung Evolutionärer Algorithmen. Dissertation, Technische Universität Ilmenau 1998.

Using micro-worlds in research on distributed cognition in complex dynamic worlds: a ten year retrospective on the cabin air management system (CAMS)

David Wastell
Information Systems Institute
University of Salford
Salford M5 4WT, UK
d.wastell@salford.ac.uk

Juergen Sauer
Institute of Psychology
Darmstadt University of Technology,
Darmstadt, Germany
sauer@psychologie.tu-darmstadt.de

Robert Hockey
School of Psychology
University of Leeds
Leeds, LS2 9JT, UK
bob@bobh.karoo.co.uk

ABSTRACT

Human machine design in complex cognitive systems involves addressing the respective roles to be played by human and non-human agents within the overall system. This paper describes the main themes emerging from an extensive programme of research into issues of interface design, stress and function allocation in complex systems. All the studies have made use of a simulated task environment known as CAMS, which is generically representative of a complex system. The design of CAMS is based on a theory of cognition that emphasizes the adaptive nature of cognitive processes and the need to assess performance decrements using secondary task components. The studies have investigated a number of factors relating to interface design, the impact of work stress, and the effectiveness of different training regimes.

Several broad themes emerge from the work which have significant implications for the design of complex distributed cognitive systems. First, recognition of the goal-directed, energetical features of cognitive work underlines the need to consider the functional state of the operator in designing optimal systems and the need to make more use of methodologies for assessing mental workload (in addition to conventional perfomance metrics). Second, the issue of function allocation is shown to be a complex one in which the naïve Tayloristic approach of automating as much as possible of the human role is less effective than a human-centred orientation which aims at achieving the optimum symbiosis. Thirdly, the value of simple computer-based simulations (microworlds) as heuristic tools for generating design knowledge is attested by our work. A fourth theme is the need to be aware that decrements in human performance can often be subtle and covert, and that secondary task methods can be necessary to detect their presence. Finally, the need to investigate human-machine performance in adverse as well as ideal working conditions is underscored by our work.

Keywords

Complex systems, human factors, function allocation, operator functional state, training methods, stress, fatigue, computer simulations, microworlds.

1. INTRODUCTION

The "problem" of function allocation in human-machine systems has long exercised researchers and engineers [18]. What tasks to give the human operator, what to the machine, and under what conditions might this be varied? The problem is a complex one. The Tayloristic solution of automating as much as possible of the human role has been shown to be inadequate. It has long been recognised that automating a process can often degrade the overall performance of the human machine system. Human operators, being displaced from the control loop, become alienated, deskilled and out-of-touch; their performance, as supervisors of increasingly automated systems, can be severely compromised when they are required to deal with malfunctions or in other exceptional situations where manual control is necessary [1,8,18].

Modern human-machine systems are becoming increasingly pervasive, complex and knowledge intensive. The function allocation problem is central to the design of distributed cognitive systems that involve the collaborative endeavour of multiple human and non-human agents, just it has been key to the design of more traditional systems. How best to divide the cognitive labour across such cooperating actors is a critical design question, especially in dynamic environments where a fixed allocation of roles may not be optimal. As human operators become fatigued or new work priorities emerge, a re-alocation of tasks may become necessary. The general question of function allocation underlies much of the research to be reported in this paper. In particular, we will be concerned with decisions regarding which tasks are best handled by the human operator, how human and non-human elements should inter-relate and collaborate, and what external factors (e.g. stress) influence the optimal balance.

In this paper, we report a programme of experiments, begun over ten years ago, using a so-called "micro-world" simulation known as the Cabin Air Management System (CAMS). Following other advocates of micro-world research [e.g. 3], we believe such simulations to have considerable potential for exploring theoretical and design issues of fundamental importance to cognitive engineering in controlled but realistic experimental conditions [13,17]. CAMS has been extensively used in our laboratories as the platform for a long term research programme [5, 9-17] and has also been adopted by other research groups [e.g. 7]. The paper will provide a brief description of CAMS and of the various versions that have been developed and deployed in our experimental programme. We will then present a retrospective overview of the main theoretical themes that have emerged from this work, concluding with some brief comments on the general implications for the cognitive engineering of complex systems.

2. A BRIEF HISTORY OF CAMS

The original research involving CAMS was commissioned by the European Space Agency as part of their long term research on manned space flight. Its scientific rationale was to enable the dynamics of human-machine interaction to be examined in a controlled environment using a task that was canonically representative of complex human machine systems. Reflecting the interests of its sponsors, CAMS purports to simulate the life support system on a hypothetical spaceship. It was designed to embody two generic features of complex human-machine systems: a cognitively challenging task involving the management of a remote dynamic process, and the need for close cooperation between the human operator and the automatic components to achieve this.

The design of CAMS was founded on a model of the human operator [4] known as the VSAT (Variable State Activation Theory). This model emphasizes the adaptive nature of cognitive processes, i.e. that performance is regulated by goals and that human agents typically respond in a complex way to external and intrinsic pressures, often through compensatory adjustments of various kinds in order to maintain performance on high priorty activities. Mental workload is a key concern in our work and CAMS incorporated a broad range of primary (high priority) and secondary (low priority) tasks to provide both realism, and a methodology for measuring mental effort. Secondary tasks afford a more sensitive method for measuring workload than primary tasks, performance (in accord with VSAT) being "protected" on the latter under conditions of high mental demands. A range of secondary tasks were naturalistically embedded within the CAMS environment, e.g. a simple reaction time task (annunciator cancellation time) and a task of prospective memory (see below). A more detailed description of CAMS may be found in several of our publications, in particular [5, 13].

The CAMS software was developed in Visual Basic and can be run on any IBM-compatible personal computer under all versions of Windows, from Windows 95 onwards. Figure 1 shows the main display of version 2.0. CAMS comprises a number of automatic controllers that maintain five main system parameters (O_2, CO_2, cabin pressure, temperature, humidity) within specified ranges. During the operation of the automatic subsystems, the primary job of the human operator is to monitor the safe functioning of the automatic controllers and to assume manual control when a system fault occurs (i.e. they operated as supervisory controllers). In the case of a system disturbance, the operator needs to engage in a process of *fault diagnosis* to identify the nature of the problem and then to take measures to rectify it. In addition to these two primary tasks, two secondary tasks also had to be completed by the operator. *Alarm acknowledgement* is a reaction time task requiring the operator to react to system alarms as soon as they occur. In the other secondary task (*prospective memory*), the operator has to remember to carry out an action at a specified time in the future, namely to make a record of O_2 tank levels at 3-minute intervals.

Since the first experiments, CAMS has undergone a number of metamorphoses in response to the evolving demands of the research programme; four distinct versions have emerged, all built around the original functionality. Version 1 was designed to explore the interaction between levels of control and sleep deprivation in relation to fault management performance in supervisory control [5]. Faults of varying levels of difficulty could be programmed by the experimenter, and two interface modes were incorporated: a human-centred interface where the division of cognitive labour was dynamic and at the discretion of the operator, and a machine-centred interface, where the division was largely fixed with the automatic subsystems embedded in CAMS taking responsibility for routine management of the life support systems.

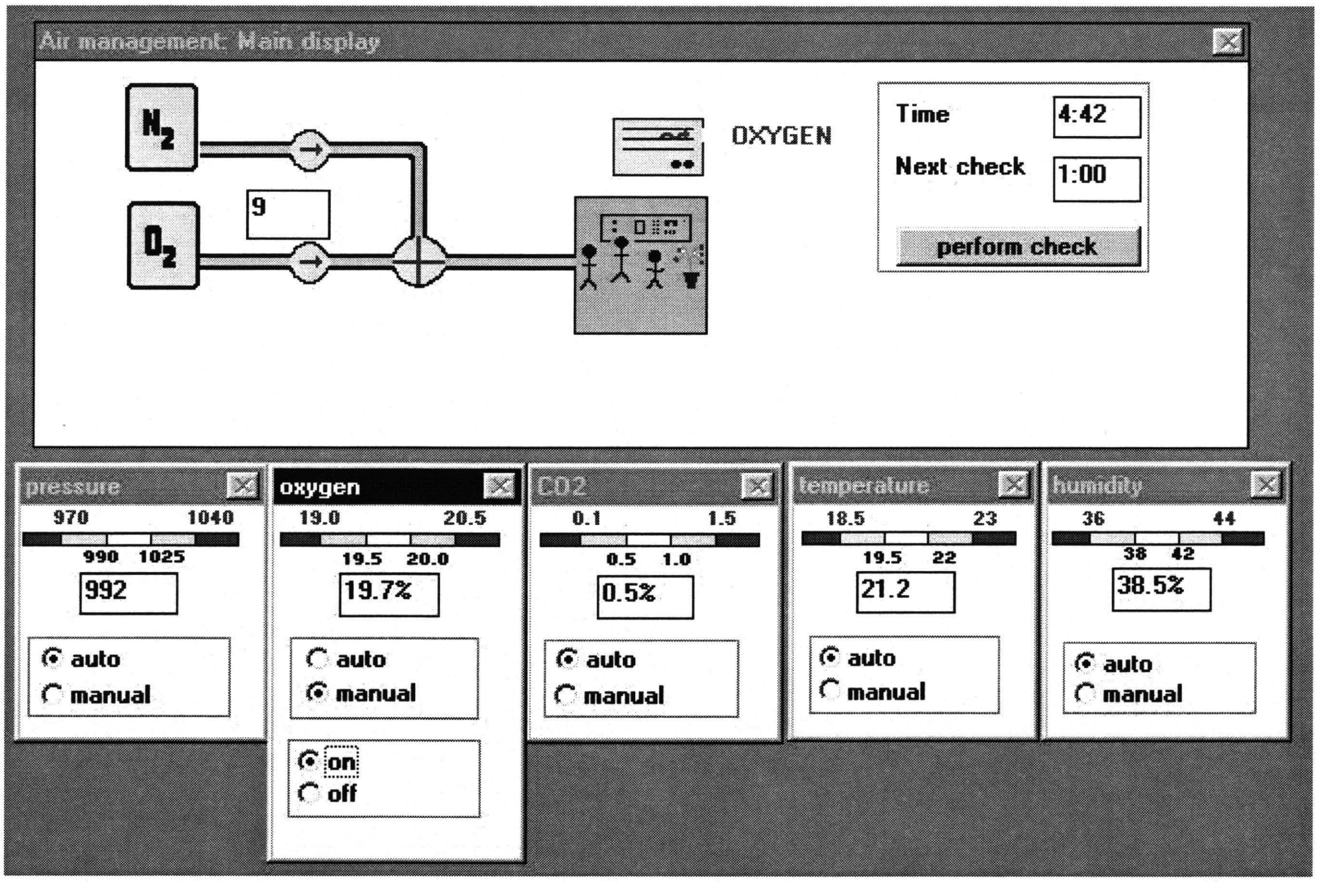
The first experiment involving CAMS [5] is typical of much of the rest of the programme and will be briefly described to give a general flavour of our work. 16 subjects were recruited for the experiment and rigorously trained in the operation of CAMS. Given the complexity of the task environment (by deliberate design to optimize its realism) up to 6 training sessions were required, of around one hour each. These sessions began by instructing subjects in the theory of CAMS; they were then given practice at controlling the system in fault-free mode before moving on to training in fault identification and management.

The experiment itself involved two independent variables: human versus machine-centred dialogue control and fatigue induced by sleep deprivation (normal sleep versus one night's sleep loss). In the human-centred version, operators could intervene whenever they wished, whereas the machine centred interface only allowed intervention when there was a problem situation. Subjects were tested under all four combinations of these two variables in a counter-balanced sequence of four 2 hour simulated work sessions. Each work session involved two faults, which could be either simple (e.g. a set-point failure) or complex (a leak) requiring the coordinated manipulation of several subsystems. The subjects' main task was that of system management, which entailed constant monitoring for correct system operation, with the rapid identification of faults and subsequent manual intervention and control.

Version 2 of CAMS was a highly simplified version, involving a set of routine faults (all set point failures) requiring manual control. It was designed to study the impact of isolation on human performance and has been used in simulated space missions [9,11] as well as an Antarctic over-wintering study [10]. This version is particularly useful in settings where training time is at a premium.

Version 3 constituted a return to the original paradigm, with the inclusion of a richer set of faults and a broader set of support aids (e.g. graphical tools for representing time trends). It has been used in studies of training methods [12] and adverse working conditions involving noise [14] and occasional night work [15]. The training study involved a comparison of procedure-based versus theory-based training. The study of night work addressed the increasingly common phenomenon of individuals being required to carry out work sporadically at times outside their normal working day, e.g. on-call maintenance work. This generic situation was simulated by having subjects carry out a one-off work session in the middle of the night, and comparing their performance with that during the day. The latest application of version 3 has been in a study of cognitive diversity in teamwork [16].

The most recent version of CAMS is known as CAMS-AUTO [7]; it has been developed by another research group in order to explore the impact of different levels of automation on human-machine performance. CAMS-AUTO incorporates an expert system that can provide three levels of operator assistance, ranging from a simple fault-finding guide to automatic diagnostic support including recommendations for operator intervention.



Time 4:42

Next check 1:00

pressure: 992 ● auto ○ manual

oxygen: 19.7% ○ auto ● manual ● on ○ off

CO2: 0.5% ● auto ○ manual

temperature: 21.2 ● auto ○ manual

humidity: 38.5% ● auto ○ manual

Figure 1. The main display of CAMS, version 2. The upper part of the screen shows a "mimic" representation of the underlying life support system. The nitrogen and oxygen tanks are shown to the left; clicking on these icons reveals the current levels remaining in the tanks. Flow meters are also available to inspect flow rates, with the operator here checking the oxygen flow. Above the cabin icon is the air purifier, which incorporates a CO2 scrubber, together with the temperature and humidity elements. Visual annunciators appear to the right of the cabin icon (the oxygen alert is on). For each subsystem, there is a control panel. These are shown below the mimic. In version two, the operator is restricted to switching the subsystem on and off when in manual as opposed to automatic mode. In the figure, the oxygen sub-system has been taken under manual control, and the supply has been switched on.

3. PRINCIPAL FINDINGS

In this section, we will provide an overview of the main findings that have emerged from our work, in terms of cross-cutting themes rather than a study-by-study chronology. We will concentrate on the laboratory studies using versions 1 and 3 of CAMS (as well as CAMS-AUTO) rather than the field studies using version 2 which were concerned more with evaluating the impact of isolation than generating theory regarding human-machine system design. The main theoretical themes are as follows:

3.1 Protection of primary task performance.

The conventional information processing model of cognition is essentially an architectural model (based on a crude computer metaphor) which describes the various components of the cognitive apparatus (e.g. memory, sensory buffers etc.) and how they fit together, with certain key parameters largely relating to the capacity and throughput of these elements. Such static models do not address the dynamic, regulatory processes that govern the deployment of these components in the work situation. By illustrating the durability of performance under stress [5,14,15], our work repeatedly brings out the key role of task goals, supported by self-monitoring negative feedback loops, in regulating and hence sustaining high levels of performance when decrements might intuitively have been predicted. Both studies involving sleep loss showed performance on high priority tasks (primary tasks) to be largely unaffected by fatigue [5, 15]. This clearly demonstrates that cognition is not a linear, mechanistic process but is a dynamic cybernetic system in which performance monitoring entails constant internal adjustments in order to optimise the achievement of task goals.

3.2 Compensatory adjustments and the hierarchy of control

Given the cybernetical nature of the cognitive system, performance decrements will not necessarily occur when operators are put under pressure due to extrinsic or intrinsic stress. The desire to maintain target levels of performance will trigger a range of possible compensatory adjustments, such as working harder (mobilising more "cognitive effort"), goal rationalisation (e.g. focusing resources on top priority tasks) and shifting to more efficient and less effortful control modes. Regarding the latter, our work suggests two basic forms of control: reactive, closed loop control and proactive open loop control [5,17]. The former is less cognitively demanding, though it is often associated with higher levels of intervention [17] with operators simply responding to events rather than formulating forward-looking, problem-solving strategies. Under adverse working conditions, there is a tendency to regress to this less effortful mode, which is often manifest in more control activity but reduced information sampling [5,17].

3.3 Mental workload, energetical costs and secondary task degradation

Different cognitive strategies differ in the degree of cognitive effort (and hence mental workload) they require. It is important to be able to measure such "energetical costs" in designing and evaluating human-machine systems. This is strongly emphasised in our work; the optimal system is one which affords maximum performance at minimum cost in terms of mental effort. Our work stresses the need to measure the costs underlying performance.

Various methods are available for measuring workload, such as the use of psychophysiological techniques [17,19]. As noted above, CAMS makes use of the well-established dual task methodology to measure the mental effort associated with task performance. All our studies confirm the expected heightened sensitivity of secondary tasks to adverse working conditions and the value of such measures for assessing the mental workload associated with different aspects of system design and operation. The test of prospective memory has proved itself to be a particularly sensitive measure.

3.4 Ideographic analysis of operator coping strategies

An important feature of CAMS is the production of comprehensive interaction protocols whereby a detailed journal is kept recording every operator interaction with the system. This has provided revealing insights into inter- and intra-individual variation in operator coping strategies. For instance, two distinct patterns of "withdrawal" have been noted which occur when the operator is increasingly overwhelmed by the cognitive demands of the task relative to his/her knowledge and energetical resources. Either operators focus narrowly on one aspect of system management to the neglect of all other issues, or they flit randomly from one aspect to another [17]. These maladaptive strategies are reminiscent of the pathological behaviour patterns observed in the work of other micro-world researchers such as Dorner [3] who identified two rather similar dysfunctional coping patterns which he exotically denominated encystment and thematic vagabonding.

3.5 Ironies of automation.

As has been found in much previous research, in the laboratory and in the field, the hoped for benefits of automation are often not realised in practice [18]. Bainbridge [1] has dubbed these deficits the "ironies of automation". In the first CAMS study, no benefits of machine-centred control on human-machine performance were found in relation to primary performance targets. Looking at "psychological costs", high levels of automation were actually associated with higher cognitive workload during fault management, especially under sleep deprivation. This reflects the lack of expertise and low confidence of operators used to a dependent relationship on the automatic systems. In other studies, although automation can bring benefits in terms of reduced workload and enhanced fault diagnosis, the relationship between automation, aiding and performance is not straightforward. Intermediate levels of automation were found in [7] to be associated with a greater degree of performance breakdown than either low or high levels of automation. Our findings confirm Norman's truism [8] that there are no absolute principles regarding the merits of automatic versus manual control; the issue is one of appropriate design.

3.6 Training methods, skill decrements and management strategies

Methods of training clearly have a major impact on patterns of operator interaction and effectiveness within the overall human-machine system. The benefits of different training regimes have been extensively studied in the research literature [12,18]. Our work has shown some benefits for training methods that incorporate a greater theoretical content. Theory-based training, as opposed to procedural training, has been shown to engender richer mental models and different management strategies, although the expected benefit of such training when dealing with novel problems has been difficult to demonstrate [12]. Procedural training can also be advantageous under certain circumstances; e.g. it was found to be less susceptible to environmental stress [14]. This confirms the general finding that training which emphasizes the use of rules and procedures for system management is typically more resilient than more theoretically oriented training [14]. The remarkable durability of human skill is shown in [12] where after an 8 month lay-off very little impact on human-machine performance was found.

4. DISCUSSION

In this final section we will tease out some broad implications of our work for the theory and practice of cognitive engineering. We are conscious that the theoretical themes adumbrated in the last section resonate with much established work in the area, but in the confines of this short paper we have been unable to embed our work fully within the established literature beyond making one or two obvious correspondences. The aim here was simply to bring our work to the attention of those working in the field of distributed cognition and complex system design, and to accentuate what we feel to be its principal implications. We would highlight five general messages.

4.1 The need to consider the functional state of human agents in complex system design

First, although cognition is distributed across human and non-human actors in complex systems, there are fundamental differences between these two types of actor which system designers should take into account, especially in relation to dynamic task allocation. Our work strongly emphasizes the importance of what may be termed the "psychological fitness" of the human operator and provides some simple tools for assessing aspects of this condition. The "energy budget" of the operator is limited, and different operational strategies and interface components can be more or less costly in terms of this budget. The incorporation of methodological features to assess subjective state and mental workload is a distinctive feature of CAMS. We have shown how the operator's state of fatigue, for instance, can influence his effectiveness within the overall human-machine system. In general, the importance of measuring mental workload is key in assessing the quality of a given human-machine design. It is vital that the costs in terms of mental energy are measured as well as overt performance in order to gain a true picture of the efficiency of the design.

The concept of *operator functional state* (OFS) captures this idea and was the subject of a recent NATO workshop [19]. If the performance of complex systems is to be optimised (especially in safety critical applications in domains such as aviation and industrial process control) consideration should be given to assessing OFS, particularly when task allocation decisions are made. If the functional state of the operator is judged to be inadequate for the performance of a task, then the allocation mechanism could either prevent it from being performed by the human actor or reassign it to an automatic component. Non-invasive psychophysiological methods have considerable promise for assessing OFS, and their potential deployment in actual working environments is explored in [19]. Building intelligence within the automated elements to detect dysfunctional operator behaviour is another possible line of attack [15].

4.2 The need for human-centred design

Regarding the benefits of automation and the more general issue of cognitive task allocation, our work has confirmed the general consensus that there is nothing inherently good or bad about automation; the issue, as we commented above echoing Norman's dictum [8], is one of *appropriate design*.

Automation does not always enhance the overall performance of the human machine system; nor is it quite as malignant as it is sometimes portrayed. Achieving a mutually supportive relationship between human operators and automation is key. Tayloristic design, in which the drive is to replace the human at all costs, can often be counter-productive. A human-centred approach is to be preferred in which the emphasis is on the support and empowerment of the human operator [2,19]. The recent work using CAMS-AUTO elegantly illustrates the point that appropriately designed machine support need not displace the operator from the loop and prejudice performance during automation failures [7].

4.3 The value of microworlds for generating design knowledge

Methodologically, our work has shown the value of micro-worlds as tools for generating and testing theoretical ideas and design knowledge in realistic, controlled conditions. Although there are obvious issues of external validity, nonetheless this approach has some clear benefits. Such medium fidelity simulations are inexpensive to produce and can be applied to generate theory and test ideas at the stage before expensive investments are made in detailed design and implementation. Micro-worlds can be used to develop ideas, create scenarios and explore human performance issues when access to field situations is impossible or highly problematic.

In [13] we present a detailed methodology for designing microworlds in relation to specific work domains and design problems. Our future work is focused on the development of a multi-user version of CAMS in order to enable team dynamics to be studied, to examine different approaches for supporting the human operator, and to develop methods and tools for dynamic task allocation in distributed cognitive systems.

4.4 The covert nature of performance decrements

In addition to these specific implications, one general cross-cutting point must be re-emphasised [15]. This is the finding, not only throughout the present work but in a growing number of studies elsewhere [4,18] that decrements in human performance may be effectively hidden in the management of complex systems under stress and high workload. The imperative to protect performance means that execution of main task can appear to be at normal levels. As we have seen, this occurs through a compensatory control process serving to protect the main activity from disruption. However, the degradation in secondary tasks, and reduction of safety margins associated with the use of more risky task management strategies, means that overall efficiency is reduced, and the overall integrity of the system could be threatened. A sudden surge of demand or loss of attention on the part of the operator may be enough to push the system into a danger state.

One of our early studies [5] found that such effects could be minimized when a user-centered interface was adopted. Such an approach has attracted considerable support in automation design [2] but has not been widely tested under demanding conditions such as night work and sleep deprivation where it is most likely to be manifest. In order to be able to assess the current operator state effectively, monitoring systems are required that track performance of not simply the top-level operational goals, but all task behaviors relevant to integrated system functioning, and relate these to baseline operator-specific characteristics. Failures in these peripheral activities may provide a more useful marker of potential breakdown than a lack of change in overt system performance[15].

4.5 The need to investigate adverse working conditions

Finally, our general concern with the impact of adverse working conditions underlines the need for empirical studies of complex system design to investigate performance under sub-optimal as well as ideal working conditions. This is necessary in order to produce a full account of the interaction between system design features, performance and the functional state of the operator. This attention to adverse conditions is a leitmotif of our work and of others [6]. This area is often neglected in much mainstream research on human-machine interaction, yet the deleterious impact of poor design may only be revealed in extreme conditions, with potentially catastrophic impact in the operational situation.

5. ACKNOWLEDGEMENTS

Much of this research was funded by the European Space Agency's Long Term Project Office (ESA-LTPO), Paris, France. We are very grateful for the financial and practical support provided by ESA, and for the encouragement provided by the directors of LTPO, A. Bichi and J. Collet.

6. REFERENCES

[1] Bainbridge, L. Ironies of automation. *Automatica*, 19 (1983), 775-779.

[2] Billings, C. Towards a Human-centered Approach to Automation. Erlbaum, Englewood Cliffs, NJ, 1996.

[3] Dorner, D. One the difficulties people have in dealing with complexity. *Games and Simulation*, 11 (1980), 87-106.

[4] Hockey, G. R. J. Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology, 45 (1997),* 73-93.

[5] Hockey, G. Wastell, D. and Sauer, J. Effects of sleep deprivation and dialogue control on the management of a complex system: a test of the compensatory regulation hypothesis. *Human Factors*, 40 (1998), 233-253.

[6] Hollnagel, E. *Human reliability analysis and control*. Academic Press, London, 1993.

[7] Lorenz, B., Di Nocera, F. and Parasuraman, R. Automated fault management in a simulated space-flight micro-world.

Aviation, Space, and Environmental Medicine, 73 (2002), 886-987.

[8] Norman, D.A. The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'. In D. Broadbent et al., Eds., *Human factors in hazardous situations*, Clarendon Press, Oxford, 1990, 585-593.

[9] Sauer J., Wastell D. and Hockey G.R.J. Multiple-task performance on a computer-simulated life support system during a space mission simulation. *Acta Astronautica*, 44 (1999), 43-52.

[10] Sauer J., Hockey G.R.J. and Wastell D. Performance evaluation in analogue space environments: adaptation during an 8-month Antarctic over-wintering expedition. *Aviation, Space, and Environmental Medicine*, 70 (1999), 230-235.

[11] Sauer, J., Hockey, G.R.J. and Wastell, D. Maintenance of complex performance during a 135-day spaceflight simulation. *Aviation, Space, and Environmental Medicine*, 70 (1999), 236-244.

[12] Sauer J., Hockey G.R.J. and Wastell D. Effects of training on short and long term skill retention in a complex multiple-task environment. *Ergonomics*, 43 (2000), 2043-2064.

[13] Sauer, J., Wastell, D. and Hockey, G.R.J. A conceptual framework for designing micro-worlds for complex work domains: a case study of the Cabin Air Management System. *Computers and Human Behaviour*, 16 (2000), 45-58.

[14] Sauer J., Hockey G.R.J. and Wastell, D. Effectiveness of training methods under different levels of noise, MS in preparation, 2003.

[15] Sauer, J. Wastell, D. Hockey, G. Earle, F. (2002). Performance in a complex multiple-task environment: effects of occasional night work and extended task involvement. *Human Factors*, In Press.

[16] Sauer, J. and Felsing, T. Effects of cognitive team diversity on performance in a complex process control task. *European Work and Organisation Psychology Conference* (Lisbon, 2003).

[17] Wastell, D.G. Sauer, J. and Hockey, G. The AMT environment: a tool for research on human behaviour in complex and highly automated systems. In A. Ozok and G. Salvendy, Eds., *Advances in Applied Ergonomics* USA Publishing, 1996, 823-26.

[18] Wickens, C. and Hollands, J.G. *Engineering psychology and human performance*. Prentice-Hall, New Jersey, 2000.

[19] Hockey, G. and Gaillard, A. (2002). *Operator functional state and impaired performance in complex work environments*. IOS Press, Amsterdam, 2003.

Session 5 Process Control Requirements and Collaborative Work: Theory and Method

Requirements Engineering for Groupware to Manage Business Process Dynamics and Stakeholders' Mindset

Johan F. Hoorn
Vrije Universiteit
De Boelelaan 1081a
1081 HV Amsterdam
jfhoorn@cs.vu.nl

Gerrit C. van der Veer
Vrije Universiteit
De Boelelaan 1081a
1081 HV Amsterdam
gerrit@cs.vu.nl

ABSTRACT

Process control in dynamic environments is most vulnerable to changes in business models. If under time critical conditions, a group decision support system causes air traffic collision or nuclear meltdown, the call for speeded decision processes may increase. However, higher negotiation speed among operators and decision makers may inherit decision inaccuracy as well, rendering the control process error prone. After a serious accident, the choices that are made for an improved software application depend on the business model en vogue. Time is money in an era of market oriented organizations (public and commercial), a business goal that directs the speed aspect of control system development. If saving lives at all costs is the core business of an organization, the accuracy aspect may be emphasized. Additionally, business goals change as a function of the mindset of workforce and managers. A big event such as a financial crisis or an accident changes the importance and attitude that stakeholders of a system attach to, for example, air traffic or nuclear power. We want to provide an overview of the state of the art that is exhibited in CSAPC 2003 and connect it to a general model of business goals and processes that direct human-machine system requirements. We do so by arguing that requirements change is a function of personal and business goals and related business processes of the worker/manager and the team. That is, requirements change of process control systems is a function of the dynamics in business models, which is regulated by the stakeholders' mindset in terms of relevance and valence toward internal and external events (cf. [2]).

Keywords

Requirements engineering, process control, business models.

1. INTRODUCTION

One of the obvious answers to the importance of process control is 'to avoid accidents.' However, organizations are not always that humane and avoiding accidents may be neglected in view of higher profits. Hence, what is required from a process control system is liable to the mission of an organization. If the mission of an organization (e.g., a hospital) is to guarantee high quality health care, one of the business subgoals may be to optimize usability of monitoring systems in intensive care units for nurses. Currently such systems are devised mainly from the perspective of the physicians. Therefore, new applications may be developed to improve the success of the control system as a whole; that is, the technical process in interaction with the workfloor operators (here nurses). In other words, a business mission (e.g., become market leader, guarantee high quality) can be decomposed into several subgoals (e.g., process optimization) to which the software applications only are a means. The CSAPC work recedes in research into the complex processes on the physical or technical level (e.g., what is required to monitor nuclear fusion) in interaction with the complex processes of the mind of the individual operator (HCI) in collaboration with other operators and managers (distributed cognition). As such, the means of developing groupware for process control to the CSAPC community is a goal to reach by conducting cognitive research in its broadest sense.

Regarding certain hot topics in process control, then, this paper offers a tabulation of the CSAPC '03 contributions. We wish to connect this overview to a model of requirements change to reach conclusions about general requirements of process control systems that (hopefully) can adapt to the stakeholders' at times turbulent mindset.

2. THE CSAPC '03 CONTRIBUTIONS

Table 1 lists the contributions to CSAPC '03 in relation to running issues: Cognitive science, distributed cognition, process control, and complex processes. The application domains covered this year are foremostly mobility/traffic ([1], [3], [4], [7], [9], [12], [14], [17], [18], [20], [21]), (nuclear) industry ([5], [21], [19]), and health care ([6], [10], [13]). Smaller fields are telephone ([2], [17]) and IT ([15]).

CSAPC explores the possibilities to add "intelligence" and "knowledge" to process control systems. As a multi-disciplinary approach of mind, knowledge, and intelligence, cognitive science ? offers CSAPC researchers various lines of investigation, from ? ethnography to psychobiology with unconscious cognition somewhere in between. Yet, the conference shows quite a small subset of approaches, mainly focussed on cognitive psychology, cognitive ergonomics, HCI, and Human Factors. There are exceptions, of course, among others, organizational psychology ! [6].

This has some serious consequences for studying distributed cognition in teams. Stakeholders of a process control system such as human operators, clients, and business managers increasingly work with groupware such as tools for group decision making or the CCT real-time animated Gantt chart method for sequencing and controlling critical operations (http://www.cctcorp.com/leader.html). The said approaches, however, have a narrow scope (perhaps even tunnel vision [18]) in cracking problems from the perspective of an individual user. High-frequency topics of this year are (mutual) situation awareness ([3], [7], [11], [16]), cooperation, coordination, and communication ([1], [4], [7], [11], [13], [16], [20]) of standard regulations as opposed to

teams!

Table 1. Contributions as related to CSAPC '03 issues

Contribution	Application domain	Cognitive science	Distributed cognition	Process control	Complex processes
[1]	Seafaring	Social interactions and rules, decision making	Collision Avoidance Rules vs. actual behavior (best practice)	Ship-to-ship communication, traffic control center	Interaction management and communication
[2]	Telephone	Cognitive ergonomics	Problem solving strategies, databases	Expert system for troubleshooting diagnosis	Comprehension, decision making
[3]	Naval	Problem solving, decision making	Situation awareness of operator and agent	Knowledge-based user assistant	Situation awareness in dynamic environments
[4]	Aviation	Collective mind theory, coordination theory	Coordination and collaboration	Air traffic control	Coordination and collaboration during collision avoidance
[5]	Nuclear power plant	Normal Accident Theory, ethnography	Standard procedures vs. human improvisation	Operational Readiness Verification, Surprise management (planning, predicting)	Coordinating systems after maintenance, Efficiency-thoroughness trade-off, adjusting to dynamic situations
[6]	Health care	Organizational psychology	Safety culture	Error management and communication (i.e. reducing *power distance*)	Human error, attitudes
[7]	Car	HCI, ergonomics	Situation awareness, interference, team cooperation and activity	In-car automation, driving assistance	Navigation, driver-car cooperation
[9]	Aviation	User-centered design	Context of use	Validation of prototype systems	Unexpected and complex events, dynamic environments
[10]	Health care	Ecological safety	Collective management	Error management	Human error assessment and human reliability assessment in dynamic situations
[11]	General	HCI, task analysis	Situation awareness, communication, coordination	Error assessment by scenario description, goal decomposition, error analysis	Collaborative errors
[12]	Naval, car	Human Factors	Information transfer, training	Identification of potentially critical situations	High-demand situations
[13]	Health care	Cognitive ergonomics	Collaborative work	Model activity tool	Group decision making, patient management of multiple traumas
[14]	Car	HCI, cognitive psychology	Metacognitive knowledge, situation awareness	Adaptive cruise control	Trust and intention of use, attitude ≠ behavior, complex and uncertain environments
[15]	Search engines	Cognitive psychology	-	Strategies to relief working memory	Human memory processes (deliberate forgetting), problem solving
[16]	Plant simulator	Human Factors	Situation and mutual awareness, cooperative activity	Inference algorithm for Team Situation Awareness	Collaboration in dynamic environments
[17]	Telephone, cars	Cognitive psychology	-	-	Driving behavior in critical situations
[18]	Aviation	Autonomous living systems	Sense of influence of crew, group decisions	Detecting contexts of fixation (tunnel vision)	Tunnel vision in complex dynamic environments, decision errors
[19]	Industry	Ecological Interface Design	Standardized technical symbols	Evolutionary algorithm for optimization of GUI	Human-machine communication
[20]	Aviation	Cognitive engineering	Collaboration with multiple agents	Micro-world simulations	Function allocation in dynamic environments
[21]	Railways	Human Factors	Common cause hypotheses	Accident cause analysis	Confounded common cause hypotheses

Note: [8] is not part of CSAPC '03

work practice ([1], [5], less so [19]). Yet, disciplines like social psychology and communication science are missing out, while organizational psychology and CSCW are underrepresented. Thus, valuable theories and methods remain unemployed.

If cognition is distributed among human stakeholders and among stakeholders and information technology (e.g., [4]), all this leads to complexity at the level of technology and of human cognition. Complex processes, therefore, have two angles, one related to the structure and function of the process control system (e.g., [2], [16],

[19], [20]) and one related to the human understanding of that system in its dynamic environment and subsequent behaviors (e.g., [2]). Table 1 shows preferences to study human aspects rather than technology, a central theme being (group) decision making. Emerging questions in the CSAPC community are about the risks of distributed decision making (e.g., [6], [11]), distributed support for decisions (e.g., [2], [3], [5]), and the application of multiple types of representations (e.g., [19]), particularly in time- and safety-critical situations. Decision making, however, is only the end phase of a far more important process, that is, (group) problem solving. Although certain studies mention this line of research (e.g., [2], [3], [15]), as yet we have no systematic insight in problem solving strategies (difference reduction, means-end analysis, forward/backward reasoning) or decision biases (e.g., aspiration levels, framing) in teams that need to safeguard consistency, mutual understanding, and knowledge management. A complicating factor is that the processes we study happen in *dynamic* environments. Therefore, interactions between the technical and human processes are per definition higher-order interactions because the environment constantly adds to the obtained variance.

Figure 1 provides an overview of CSAPC '03 in terms of the general problem solving strategy we as a group employ for developing process control systems. Note this is a rough summary and individual researchers may take different angles. The mission of every organization that supports control system development is to make the (work) environment safer. In our case, CSAPC '03 mainly focuses on mobility, industry, and health. Goal decomposition is established by setting business goals, one of which is to control the work processes that go on in the application domains. We wish to reach this subgoal by trying to understand the complex processes involved in, for example, nuclear power, telephony, or database searches. These may be business processes (e.g., distribution of waiting time in routers, nuclear fusion) or cognitive processes (problem solving, decision making, error handling). Yet, all of them happen in situations that change, with critical events, which in any case cause third-order interactions and sometimes even higher ones. Problems are solved best when approached from different views. The views we take are quite limited because often they are restricted to the individual operator (other stakeholders are considered less).

Overviewing the questions and needs of the CSAPC community of this year, then, as they appear from Table 1 and Figure 1, the following research agenda could be proposed. We are in need of more CSCW, social and organizational psychology approaches as well as communication science to study the effects of situation awareness on problem solving strategies (i.e. standard rules vs. work practice) and decision biases (e.g., induced by (lack of) trust, reliability, function allocation, speed or accuracy fixation) while cooperating and coordinating work in teams that have to cope with critical events.

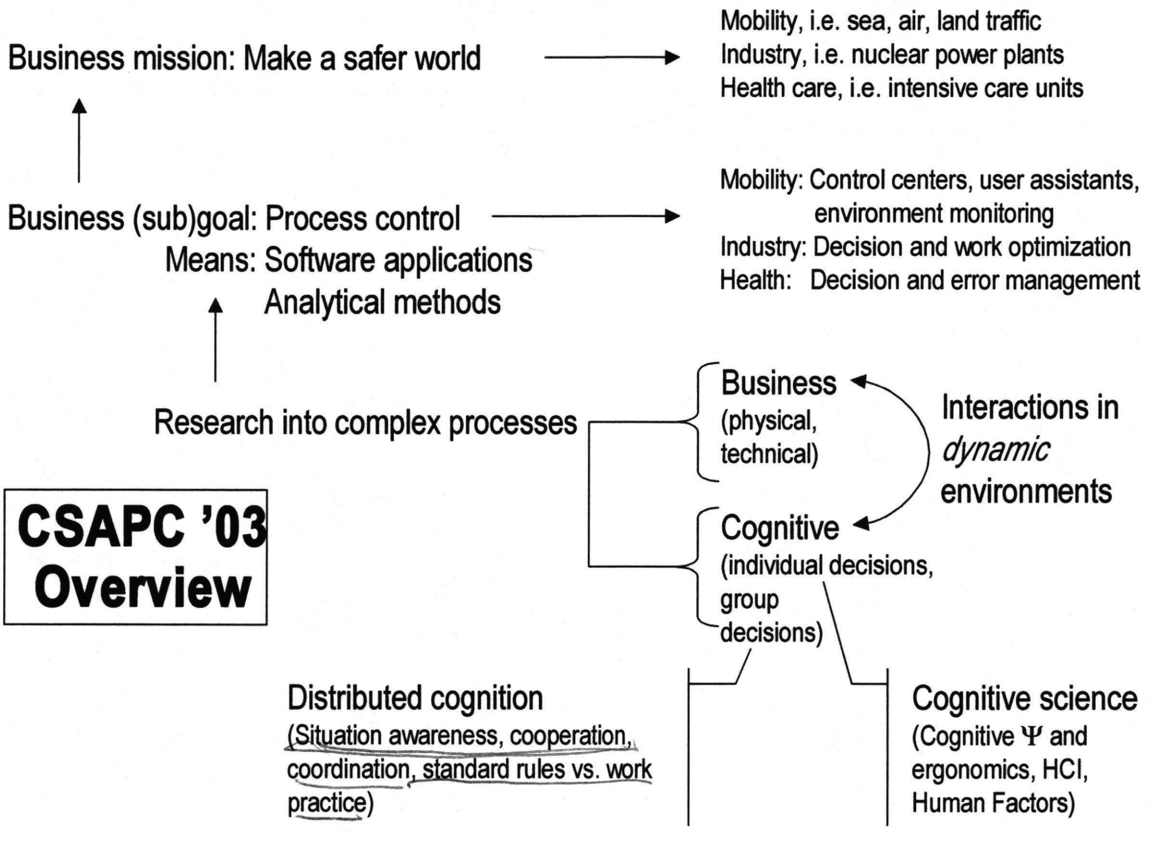


Figure 1. CSAPC '03 viewed as a problem solving organization

3. REQUIREMENTS CHANGE IN PROCESS CONTROL GROUPWARE

There is a need for more sociotechnical approaches in process control because most systems used are groupware (e.g., communication systems, plant control systems, patient monitoring systems, and control center systems). In addition, almost every study in the CSAPC '03 acknowledges the complicating effects of dynamic environments in which critical events occur (Table 1). Accordingly, risk assessment or the perceived threat to goals may differ for each team member. Yet, team members should cooperate to avoid accidents so that conflicts can occur in handling the groupware for coordinating and communicating the necessary work. In other words, knowing dynamic environments is to know also the group dynamics in taking decisions under safety critical conditions.

Critical events may lead to formulation of new requirements of a control system (better feedback, stability, transparency, etc.). However, what is required does not need to be the same for everyone. Whereas the one operator judges that a system should tolerate exceeding the critical mass limit of a uranium solution, the other would want automatic shutdown when that limit is reached. Additionally, requirements may not only differ within operator teams; management, client, government, and other stakeholders also have their say (e.g., in time-efficiency and cost-effectiveness of a control system). When the CSAPC contributors suggest new features for a process control system, the formulated requirements should be viewed from different perspectives. What serves the goals of an operator team (e.g., longer time delays between flights) may interfere with the business goals of the management (e.g., to serve as many airliners as fast as possible).

Figure 2 gives an outline of the dynamics of conflict that can take place while a (process control) system is developed or redesigned [8]. On top are the requirements as they are agreed upon in a requirements document. Such documents can be compiled from, for example, the existing system, stakeholder elicitation, actual work practice, and standard (safety) regulations. However, these lists can be contradictory or actually not agreed upon at all but forced upon the operators by the management or the client. Although requirements should be analyzed and negotiated by the different stakeholder groups (including designers and software architects), the results may still contain quite an amount of tension.

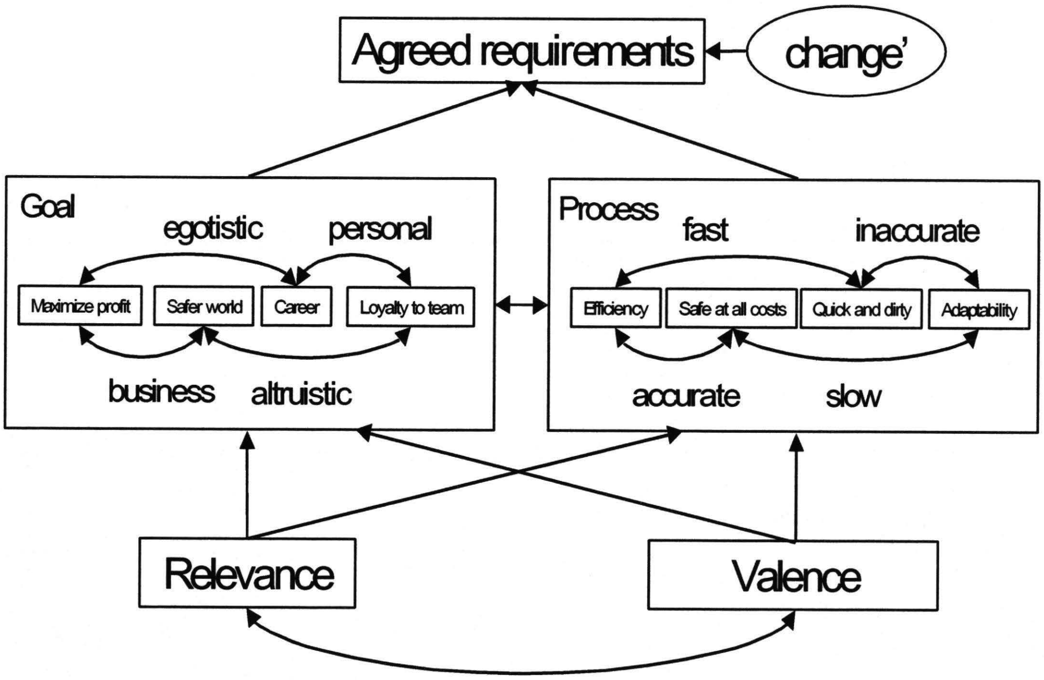


Figure 2. Model of Requirements Change (MoRC)

The Model of Requirements Change (MoRC) claims that the goals people have (whether as a business or as an individual) and the type of work processes they employ to reach these goals determine the level of (dis)agreement about the requirements of the IT. Regarding goals, the business may have a mission (e.g., fire-fighting), but also business goals to reach (e.g., to break even or to maximize profit). As an individual, employees also have certain goals and concerns. They have to make a living, want to make a career, but also want to help people, cure diseases, and be a partner to their colleagues. In other words, goals can be *business* as well *personal* and they can be *egotistic* as well as *altruistic*. Everyday work practice shows that tensions among these types of goals may arise within and between stakeholders (cf. "cognitive compromise" in [2]).

Regarding processes, a frequently mentioned aspect in the CSAPC '03 studies is that a (control) process should be efficient (e.g., [5], [12]). Efficiency has two dimensions, one being time and one being place (cf. setting speed and distance in an adaptive cruise control [14]). The most efficient process is the one that runs fastest. However, it should not make errors or lead the operator into making mistakes. The most efficient process, then, also is the one that is most accurate (cf. "Efficiency-Thoroughness Trade-Off" in [5]). That is, each processed element is arranged in the right place, every button is correctly pushed, and all modules call on each other appropriately. If efficiency is combining high speeds with high accuracy, as a result, three other process types can be deducted: Low speeds with high accuracy, high speeds with low accuracy, and low speeds with low accuracy. This grid of processes being *fast* or *slow* and *accurate* or *inaccurate* is not judgmental about what combination is right in a given situation. An efficient process may be too fast to follow for a human operator. Moreover, such a process needs every element at the right place in the right time, which makes it vulnerable to dynamic environments because there is no time to adapt (cf. [20]). A slow and sloppy process may be annoying and cost-ineffective, but on the other hand, because there is plenty of time to think up other solutions and system failure can be corrected manually, operators can exert their creativity and adapt easily to changing situations. Between the extremes of being either fast and accurate or being slow and inaccurate are the *safe-at-all-costs* and *quick-and-dirty* types of processes. Organizations that value a safe environment (cf. "safety culture" in [6]) want to work with accurate control processes, but often, this is at the cost of the technique becoming slow. Requirements change, in that case, may stem from cost-effectiveness, for example, and are directed towards efficiency in terms of increasing the speed of control systems while maintaining accuracy. Other organizations, however, value speeded treatment over accurately dealing with a problem (cf. troubleshooting by helpdesks, rapid prototyping by design teams). These organizations want to work with fast control processes, although they are at the expense of becoming inaccurate. For instance, [7] states that "in-car safety equipments are seldom sold as pure safety devices, but above all as devices capable of improving the performance (e.g., speed) of the driver-car system." Requirements change, then, may be instigated by end-user dissatisfaction and is directed towards efficiency in terms of increasing the accuracy of control systems while maintaining quickness.

What the goals and work processes of a stakeholder or group of stakeholders are determines what is requested from the control process. If making the world safer really is top priority, an accurate process is required that is allowed to be slow as long as that is the way to avoid (human) errors. In contrast, how much electricity a power plant produces is determined also by making profit. Accordingly, process speed may be pushed to the limit, making the process become more error sensitive than is strictly necessary (the *quick-and-dirty* approach). This relation also holds the other way round (Figure 2, double-headed straight arrow). If it is technically impossible to optimize speed and accuracy of a control process up to the level of efficiency, goals have to be reformulated in less ambitious terms (e.g., tolerating 2% more errors than targeted, making the world less safe). The tension between goals and/or between processes emerges most prominently when a critical event happens and business and personal goals are at stake.

Here we get to the deepest level of human behavior (Figure 2), namely the assessment of risk in terms of the *relevance* of the critical event to the goals and concerns of the stakeholder(s), and the evolving action-tendency or *valence* towards the event: Fight, flight, or sit still (see [8] for theoretical backgrounds; see also “withdrawal” in [20] and error reporting in [6]). When the event is life threatening (high relevance, negative valence), the call for new requirements will be strong. If the critical event has a positive outcome because the operator team handled the problem well (high relevance, positive valence), the high risk is balanced by high satisfaction, creating a lesser need for new requirements. When the relevance of an event to goals and concerns is low (cf. “near misses” in [21]), the intensity of the need for new features of the system will be low. At the level of the management, a critical event can evoke the need to change the business model, for example, from decentralized traffic control to a central command structure (cf. [1]). Here also, the new business goals are checked for their relevance and valence towards the goals and concerns of the (individual) stakeholders. Increasing the level of automation may cause “a loss of manual skills in guiding a process while the responsibility for the operators increases” [19]. Because operators cannot interfere with what they are held responsible for (high relevance, negative valence), anger and frustration may occur, inhibiting the acceptance of the new technology. As [17] states: “The drivers must remain as the main car controllers. They cannot be transformed into car automation supervisors” (also [20]). In contrast, when automation relieves a user from certain responsibilities (e.g., anti-collision with fixed or slow obstacles [7]), positive emotions may occur (high relevance, positive valence), facilitating the acceptance of the new technology. Hence, requirements change is a matter of cognitive-emotional assessment of a control system’s features in terms of relevance and valence to goals and concerns (e.g., ‘reliability’), i.e. of the individual stakeholder in relation to his/her peers (e.g., colleagues). When valence is positive and relevance is high, trust in the machine and intention of use (cf. [14]) may evolve from this constellation.

When empirical values are added to the relations depicted by the arrows in Figure 2, we can speak of a “stakeholder’s mindset.” Over time (i.e. after a collision), these values may change. The fixed aspect lies in the relations or functions among the constructs of the MoRC. The change is accounted for by the changing values that can be filled in for the factor loads or regression weights (single-headed arrows) or covariances and correlations (double-headed arrows). Put plainly, people do not quickly change their goals of wanting to make a living or work in a safe environment. Given the circumstances, they just put a different weight to them. Within this complex of trade-offs, people are willing to take (big) chances when they judge, for instance, that production speed is more important than always being on the safe side, particularly in view of a big reward. The MoRC, then, is an instrument for requirements validation, for instance, when employed in a structured questionnaire.

4. REQUIREMENTS IN CSAPC ’03

Based on Table 1, Table 2 is an inventory of the general requirements of a control system that CSAPC ’03 agrees upon, enriched with insights derived from the MoRC (Figure 2). The main goal our community has is to create a control system (consisting of humans and machines) that is safe as well as efficient. On the work floor, safety, of course, is a highly relevant egotistic personal goal, whereas for the management, it is an egotistic business subgoal, for instance, to secure a smooth production line. The altruistic aspect may be present as well, for the personnel on a personal level (brothers in arms, team players) and for the management perhaps to a lesser extent (cf. power distances in [6]). An efficient control process is fast while accurate, but the hardest to achieve, vide the 20 contributions discussed in this paper. Moreover, efficiency is not always in agreement with the needs of the operators ([19], [20]).

Table 2. General requirements of the human-machine system in process control

Goal:	Safe while efficient
Subgoal	**Means/processes**
1. Increase domain knowledge	Analysis of dynamic environments, i.e. signaling irregularities (e.g., errors)
2. Group problem solving, i.e. in task coordination and function allocation, power distances, communication	Support tools that suggest alternative strategies, possible solutions, and that warn against confounding factors (cf. the common cause hypothesis in [21])
3. Group decision making	Support tools that calculate the risks of options in terms of relevance and valence to goals and concerns, and that warn against decision biases
4. Error recovery	If error, go to 1→2→3, else 5
5. Adaptation to dynamic environments	Given the output of 3, change behavior (human and machine) towards cooperation

Subgoal 1 in Table 2 is derived from the CSAPC contributions that mention the importance of situational awareness and the effects of dynamic environments. The underlying purpose of monitoring the situation of the system in its surroundings is to increase domain knowledge, which is needed to generalize events, extract mechanisms, and to solve problems with more precision (accuracy). The analysis of errors suggests that the CSAPC contributors expect relatively fixed or regulated patterns of system behavior and relatively predictable evolvement of the surroundings. Thus, analysis of the environment should be, first of all, to signal irregularities in control system progression (humans and machines), which can be done, for example, by employing certain AI applications.

Subgoal 2 pertains to problem solving in the case an accident happens. A complicating factor is that the cognitive psychological literature on problem solving is oriented to the individual and does not often involve the social psychology of group dynamics, such as the effects of power distances on the points 1-5 (Table 2). The CSAPC ’03 group signals that the biggest problems occur in the synchronization of work and tasks, function allocation (i.e. between humans and machines), and communication of the points 1-5

(Table 2). As is, there are problem solving and decision support tools, such as expert systems and case-based reasoning systems, but they are restricted to one approach only, whereas the power of human problem solving is the use of multiple approaches. Moreover, conventional software does not suggest alternative explanations of the same phenomena (cf. [21]). Another helpful feature would be to have multiple representations of the same problem space (cf. [19]).

Subgoal 3 is to make decisions about the options derived from 2 and to find consensus among decision makers. Descriptive (how things are) as well as normative (what you should do) decision support systems assess risks in terms of (commercial or military) profit and loss, without considering the emotional aspects of decision making, such as the harm or benefits to other (e.g., personal) goals. Group decision support systems do not account for or correct for the different power distances in establishing consensus among group members. Moreover, it would be helpful if tools did not generate one type of output (e.g., 60% lost or 40% saved), but would counterbalance different decision frames (60% lost *and* 40% saved) and warn against positive or negative biases based on such frames.

Error management and error recovery (subgoal 4) is repeating the loop from 1 up to 3. With the error, an irregularity in the standard safety procedures or best practice behaviors has occurred, which is one of the factors that make environments dynamic. If no errors occur (any more), the human-machine control system is ready to adapt to changes.

Subgoal 5 is to have an intelligent and knowledgeable control system that can adapt to change. In relation to all five subgoals, intelligence should come from the humans while the computer could add some knowledge. System adaptation should be based upon the recommendations done under 3, and as understood from the CSAPC '03 contributions, the points 1-5 should be directed towards cooperation to eventually reach the highest goal, that is, establishing a safe while efficient process control system.

However, the MoRC argues that an adaptive process probably is not efficient and vice versa. The more options the system can contemplate, and thus, the more situations it can adjust to, the slower process time becomes. Moreover, an adaptive system adapts to errors too and allows inaccuracy in how the work is done to open up new possibilities and views, which can lead to new solutions. For instance, according to rule it is a safety violation to create short circuit but if this is a way to shut down a system that is in a critical state, the error merely is an unconventional solution to reaching a higher goal. Thus, errors are only mistakes in the context of certain preset goals and their respective ranking. Error making, then, may not only indicate lack of skill, but also a different mindset in which different goals are considered important.

Making the means and processes in Table 2 fast will be a hard thing to do. Offering alternative solutions, contemplating confounding factors, negotiating decisions, evaluating different decision frames, etc. may add to the accuracy of controlling a process but not to process speed. For the time being, then, while working without commercial aspirations, the CSAPC community probably has to settle for control processes that are safe at all costs, that is, accurate but slow. Otherwise, process control will be a quick and dirty job.

5. CONCLUSIONS

In this paper, we made an inventory of the contributions to CSAPC '03 with respect to this year's issues. The application domains are mainly mobility/traffic (sea, land, air), (nuclear) industry, and (intensive) health care. The cognitive science that is most frequently employed concentrates on cognitive psychology, cognitive ergonomics, HCI, and Human Factors. This is a bit at odds with the importance of distributed cognition, from which one would expect more contributions from social psychology, communication science, organizational psychology, and CSCW. Although almost all studies acknowledge the importance of teamwork, many focus on the behavior of the single operator. It should be added that other stakeholders than those on the work floor are important to process control as well. After all, the clients and their managers or the government decide what safety measures will be taken (or not). Another peculiar thing is that problem solving and decision making seem to be central to process control, but the psychological literature on these matters is hardly consulted. Moreover, this year's contributions all mention that process control has to deal with the ongoing changes in environments. Because humans are one component in process control, machines the second, and the environment is the third, empirical investigations will always have to consider multifactorial models and higher-order interactions that explicitly take 'change' into account.

To make a contribution into that direction, this paper connected the inventory of CSAPC '03 issues to a model of requirements change (MoRC) to deduce general requirements of process control systems that should be capable to adapt to environmental changes. Our thesis is that stakeholders evaluate a critical event, or a control system's feature that should help cope with that event, for relevance and valence towards personal and business goals, which may be egotistic as well as altruistic. Stakeholders evaluate also the relevance and valence of the speed and accuracy of the related business processes. The trade-off outcome space is used to assess the need for a control system feature, which is reflected by the degree of agreement to that feature. The changes that occur in the requirements of a system are based on the weights a stakeholder attaches to a feature or to requesting different features. However, the structure of the stakeholder's mindset as laid down in the MoRC will remain relatively fixed. Applied to the needs as expressed by the CSAPC '03 contributions, general requirements of a process control system should guarantee that the (human-machine) system is *safe while efficient.* As analyzed from the overview of the CSAPC '03 issues and extended with the views of the MoRC, the following requirements of a human-machine process control system have top priority:

1) Systems should be capable of increasing the domain knowledge about dynamic environments, including, humans, machines, and surroundings. 2) Systems should facilitate problem solving in *groups*, i.e. coordination and communication of tasks and work, thereby accounting for social dynamics as well as offering multiple views and representations. 3) Systems should support group decision making not only in the conventional way of risk assessment and negotiating for consensus, but also in terms of relevance and valence of events towards stakeholder goals and concerns. Here also, correction for social dynamics (power structure of decision making) and warning feedback against decision biases should be incorporated. 4) Tools for error recovery should not only correct for irregular behavior, but check whether or not the error occurred as the result of an inference based on premises (i.e. goals to reach) different from those in standard regulations and best practice behaviors. Results should be fed back to 1. 5) Based on the group decisions, the human-machine system should change its behavior in accordance with the changes in the environment in order to be adaptive. As derived from the inventory in Table 1, this change should mainly concern the improvement of human to human and human-machine cooperation.

Yet, the MoRC predicts that an adaptive system is the opposite of an efficient one. Being efficient is not the same as being effective. Efficiency is process or means oriented (i.e. working fast and accurate), whereas effectiveness is goal related, meaning 'reaching the preset goal.' It may be possible for a machine to be fast as well as accurate, but humans will not be able to keep up with extreme levels of pace and precision. In contrast, humans may be capable to adapt to change, but the machinery will have great difficulty in handling inaccuracy (e.g., effective errors). It depends on the urgency of a situation which aspect should be improved. Under time-critical conditions, increasing the speed of the process will receive the most attention. If time is less of a constraint, accuracy (i.e. in handling inaccuracy) may be improved.

With the MoRC, instruments, such as structured questionnaires, can be developed for requirements validation. The kind of statement requirements engineers would like to tease out from a (control) system's stakeholder are "I agree to this feature of the system because it supports my goal of getting some decision support. I agree to this feature very much because certainty about my decisions is very important to me." Therefore, statements in such a questionnaire should follow a systematic framework (Table 3).

Table 3. Conceptual framework for requirements validation of a process control system

Goal type	**Type 1: Business egotistic**	**Type 2: Business altruistic**	**Type 3: Personal egotistic**	**Type 4: Personal altruistic**
System feature	E.g., cost-effective	E.g., safe conduct	E.g., decision support	E.g., be a team player
x_1	Relevant? Irrelevant? Support? Obstruct?	"	"	"
x_n	"	"	"	"
(Sub) process	**Sub 1: Fast/accurate**	**Sub 2: Slow/accurate**	**Sub 3: Fast/inaccurate**	**Sub 4: Slow/inaccurate**
System feature	E.g., start-up	E.g., maintenance	E.g., trouble-shooting	E.g., manual data input
x_1	Relevant? Irrelevant? Support? Obstruct?	"	"	"
x_n	"	"	"	"

Requirements validation should start with compiling a negotiated requirements document, administering the level of agreement to all required features in several stakeholder groups (e.g., operators vs. management). Additionally, a questionnaire that is structured according to Table 3 has a part on goal related items and a part on process related items. Items should consist of the features (Table 3, gray area) in the negotiated requirements document extended with questions on its relevance (relevant-irrelevant) and valence (supportive-obstructive) to certain goals. In the first part, goals follow the grid of the MoRC: Business or personal, egotistic or altruistic. What goals should be filled in is derived from (ethnographical) field studies and task analysis. In the second part, goals are the desired states of the business processes, whether they should be fast or slow, accurate or inaccurate. What the relevant (phases of) business processes are can be derived from workflow analysis. The obtained degrees of relevance and valence towards goals and processes, finally, should explain the level of agreement to a process control system's feature. Changes in the degree of relevance and valence of a feature, then, explain why people want something different from a process control system after a critical event has taken place. It is the authors' future work to pursue empirical investigation of this here framework in the hope to distribute our new cognitions at CSAPC '05.

6. ACKNOWLEDGMENTS

This paper is part of the Innovation Oriented research Program (IOP) for Human-Machine Interaction entitled *Integrating Design of Business Processes and Task Analysis,* granted by the Senter Agency of the Dutch Ministry of Economic Affairs in The Hague, grant Mmi99009.

7. REFERENCES

[1] Andro, M., Chauvin, C., and Le Bouar, G. Interaction management in collision avoidance at sea, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 61-66.

[2] Cegarra, J., and Hoc. J.-M. Individual differences and distributed cognition: The case of troubleshooting diagnosis, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 113-118.

[3] Döring, B., Dörfel, G., and Distelmaier, H. Object-oriented modeling of situational awareness information for knowledge-based user interfaces, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 3-10.

[4] Fenema, van, P. C., and Simon, S. J. Failure of distributed coordination in high reliability organizations: The midair collision over Germany on July 1, 2002, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 23-28.

[5] Gauthereau, V., and Hollnagel, E. Building expectations and the management of surprises: Studies of safety-train outages at a Swedish nuclear power plant, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 97-103.

[6] Itoh, K., Abe, T., and Andersen, H. B. Healthcare staff attitudes towards management, job, teamwork and leadership in Japanese Hospitals, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 67-74.

[7] Hoc, J.-M., and Blosseville, J.-M. Cooperation between drivers and in-car automatic driving assistance, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 17-22.

[8] Hoorn, J. F., and Van der Veer, G. C. Requirements analysis and task design in a dynamic environment, in D. Harris, V. Duffy, M. Smith, and C. Stephanidis (Eds.), Human-centred computing: Cognitive, social, and ergonomic aspects. Volume 3 of the proceedings of HCI International 2003 (Hersonissos Crete, Greece, June 2003), Mahwah, NJ, Erlbaum, 472-476.

[9] Lanzi, P., Marti, P., Battino, P., and Pozzi, S. Pre-operational validation in safety critical environments: The challenge of advanced shadow mode trials, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 129-136.

[10] Marc, J., and Rogalski, J. Individual contribution to safe collective management in dynamic situations. The case of a medical emergency center (SAMU), in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 75-81.

[11] Miguel, A., and Wright, P. CHLOE: A technique for analyzing collaborative systems, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 53-60.

[12] Neerincx, M. A., Rypkema, J., and Passenier, P. O. Cognitive and functional (COLFUN) framework for envisioning and assessing high-demand situations, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 11-16.

[13] Pellegrin, L., Bonnardel, N., deVictor, N., Bastien, C., and Chaudet, H. Modeling a cooperative medical decision making in an intensive care unit, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 105-111.

[14] Rajaonah, B., Anceaux, F., Espié, S., and Hoc, J.-M. A study of the link between trust and use of adaptive cruise control, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 29-35.

[15] Rizzo, F., Parlangeli, O., and Bagnara, S. Remembering and forgetting effects in analyzing search engine list results, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 83-88.

[16] Shu, Y., Furuta, K., and Nakata, K. A method for defining and inferring team situation awareness in cooperative activity based on mutual awareness, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 121-128.

[17] Svenson, O., and Patten, C. Safety of a human-technology-organization system: Mobile phones and driver cognition disturbance, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 37-43.

[18] Vlugt, van der, M., and Wieringa, P. A. Searching for ways to recover from fixation: Proposal for a different viewpoint, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 47-52.

[19] Völkel, A. Evolutionary optimization of graphical human-machine interfaces for controlling technical processes, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 137-143.

[20] Wastell, D. G., Sauer, J., and Hockey, G. R. J. Using micro-worlds in research on distributed cognition in complex dynamic worlds: A ten year retrospective on the Cabin Air Management System (CAMS), in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 145-150.

[21] Wright, L., and Van der Schaaf, T. Causation patterns of accidents versus near misses: A critical review of the literature, and an empirical test in the railway domain, in *Proceedings of CSAPC '03* (Amsterdam NL, September 2003), ACM Press, 91-95.

Cognition and Collaboration: Trends, problems and challenges from academia and design practice

Cristina Chisalita,
Vrije Universiteit, Amsterdam
De Boelelaan 1081a
1081 HV, Amsterdam
cristina@cs.vu.nl

Peter H. Jones,
Redesign Research, USA
4098 Wagner Road,
Dayton, Ohio
peter@redesignresearch.com

ABSTRACT

In this paper, we present the results of the discussions from two workshops held by the authors with the main subject: the analysis of the collaborative activity. Several trends, problems and challenges related with the theory and practice of the analysis of collaborative activity are identified and discussed.

Keywords

Collaborative activity, distributed cognition, theory, research methods, workshop

1. INTRODUCTION

The current workshop, Cognition and Collaboration - Distributed Cognition in Complex Processes, is the second workshop in a series of research discussions. The first workshop, Analyzing Collaborative Activity, was held at CSCW 2002 and focused on analyzing collaborative work and representation of collaboration in the context of Computer-Supported Cooperative Work technologies. Both workshops relate the use of multi-method and multi-theory approaches from field research and ethnographic studies of collaborative process. Contributions range from empirical questions of field research to theoretical integration. The "middle ground" of inquiry has been developed, with field researchers inquiring into design issues, designers considering theory, and theoretical investigators looking at applications. See papers and information: www.redesignresearch.com/cscw/.

Following the more general findings of the first workshop, the second workshop aimed at a more specific target, developing the conceptual framework of Distributed Cognition for analyzing and understanding collaborative work.

2. ANALYZING COLLABORATIVE ACTIVITY – GENERAL ISSUES

The discussions of the first workshop elicited several broad perspectives or themes relevant for analyzing collaborative work:

CSCW? As we continue to explore this domain of inquiry, some find it artificially limiting to articulate research problems specifically about cooperative *work*. A trend toward broadening interest in intellectual cooperative activity extends the to Learning (CSCL), Play (CSCP?), and more generally, "practice." Although the roots of CSCW draw from the original interest in group technologies and enhancing cooperation in organizational settings, many researchers have pushed beyond these bounds. Since this topic has been circulating across other discussions, this topic should gain momentum in future workshops. As cooperative technology extends beyond the conventional descriptions of computing, we might also call for an extension of the field into *cooperative technology for collaborative practices.*

CSCW for: Design, Products, Organization, Community, Work Practice. Participants disclosed various target domains for the application of collaborative technology. Opening up the space for cooperative activity and opening up the technology also opens up the possible applications space. Some articulated this "designing for" as *intervention.* Any design activity oriented toward collaboration support might be understood as an intervention in the social space, and treated as such by researchers and managers.

Adoption of interpretive frameworks. Many of us reported from experience with interpreting field research through theory or integrating theories to generate broader understanding or applicability from findings. Our common ground was the theoretical orientation of distributed cognition, with several reporting findings interpreted from the lens of distributed cognition or activity theory. We further articulated the use of theoretical frameworks as such as "Lens," a way of viewing the domain that focuses on some aspects clearly, while necessarily relegating other aspects to the periphery. This choice of lens specifies explanatory power; disclosing this feature of theorizing allows us to work across theories by choice, agnostically.

Beyond Contextual Design. Over half of the projects reported on work integrating at least some Contextual Design (CD) representations (Holtzblatt & Beyers, 1998), with several utilizing much of the complete CD toolkit. To some extent, the use and effectiveness of the models were taken for granted, as though CD now afforded a canonical approach. One distinction of the workshop was to not take representations for granted, but to explore their uptake in our research and design projects. While this was an intention of the workshop, due to our focus on issues and not specific models per se, little discourse emerged on the CD toolkit.

An interesting sidebar should be noted that it seems CD models are now widely used across both ethnographically informed research and design-oriented research. Perhaps researchers may find little to share on the specifics of its use due to its acceptance; we found little sharing on any downsides of reliance on this popular toolkit. Due to the wide uptake of CD, as responsible researchers we might further explore, critique, and evolve its representations.

Multidisciplinary researchers and practitioners. An obvious and pervasive theme was the disciplinary variance among workshop participants. Ranging from anthropologists to social scientists to design researchers, and using methods ranging from ethnography through organizational modeling to prototyping, we found differences in method and interpretation. Although all participants might be considered interdisciplinary to some extent, we still found need to smooth out variations of understanding, which takes time. We might attempt to build more bridges of background in advance of future workshops.

In collaborative work analysis, both designers and academics reported significant **issues and challenges**. The workshop identified the following salient issues, among others.

2.1 Ethnography and Design Research - How do we integrate ethnography into design and research contexts?

With an emphasis on field data analysis, most projects reported use of ethnographic methodology. Ethnographic methods are often used as part of a multi-method research approach, with more or less applicability and validity based on the experience and articulation of the individual researcher. Rather than invalidating the strength of ethnography, the multiple methods keep teams honest by providing multiple interpretations of field data.

One of the attractions of toolkits such as Contextual Design is their support for filling in the gaps of researcher method: ethnographers unfamiliar with design process may use CD and its models as a communicative bridge to create value for the intended designers, even though they may lack in design experience. Likewise, design professionals with limited field research experience may bridge their own way into deeper understanding of user practice through the scaffold of models, as part of a multidisciplinary design team.

A continuum of issues (similar to that of disciplinary focus) ranged from ethnography as a social research process to using ethnography as a method for understanding users in design projects. These extreme ends of the spectrum may require different representations, levels of skill, and degrees of engagement.

We also noted how ethnographic studies often represent a *"situational slice,"* leaving us with the issues of authentic representation of practice when our studies necessarily encompass a specific time slice. One issue that warrants much further discussion is the notion of studying activity and collaboration *over time*. Temporality remains a major factor not effectively addressed by CD or other models; observations of any unstructured (non workflow) activity show patterns of activity that are only partially revealed through the situational slice of the current study. While longer studies (as advocated by experienced ethnomethodologists) may reveal much more of the structure or relationship in activity over time, we find a paucity of representations for interpreting the temporal dimensions of collaboration.

The workshop discussed the *evaluation of research method effectiveness*. How do researchers assess the effectiveness of specific methods and multi-method hybrid approaches? How do we know which approaches offer the sensitivity to specific research questions? How do we improve skill and practice to better validates our own field research? While experimental and quantitative research methodologies enjoy well-established standards of effectiveness, field research remains a researcher-centric skill, developed by practice, feedback, and self-reflection. While experiments are designed to be replicable, each field study remains its own unique case, and is subject to influences of the researcher, the organizational unit of analysis, the type of intervention, the specific time chosen, and many other factors. Traditional controls include analysis and disclosure of possible effects, multiple brief studies, using different researchers and informants, and so on. However, when ethnography becomes widely used for design research, and Contextual Design becomes used as part of a team design process, some types of simple and practiced research controls might be considered valuable to share and document.

Ethnography was discussed *as a "method" and also as a way of approaching research.* Relating this to the uses of ethnography discussed above, we find ethnography integrated into larger research projects as one of several methods for engaging the field. This does not in and of itself "reduce" ethnography to a method – effective use of ethnography requires a mindset and understanding of ethnomethodology. As a way of approaching a research effort, the ethnographer starts with good tools for ethnography and expands the toolkit from there to include various interpretive approaches, models (such as CD), and process-specific representations.

2.2 Theoretical Issues, and Uses of Theory

A distinct difference was found between theoretical approaches to field research in collaborative work and empirical studies focusing on findings in a defined field setting. Although these approaches can also be described as deductive (top-down) and inductive (bottom-up) in contrast to each other, theoretical frameworks are by design deductive approaches. In one issue session, theoretical papers raised questions of appropriate theories and frameworks. We acknowledge a field where one set of theories may be useful for understanding social systems, another set may be appropriate for relating practices to design, and yet another for selecting and implementing interventions in organizational practices.

Theories are not being left to themselves – many of us are integrating theory to build a better framework for explanation and interpretation. One paper (Andriessen) brings together four major theoretical frameworks to propose a Dynamic Group Interaction Model (Andriessen, 2002). Chisalita adds organizational culture theory to the well-documented design process models (DUTCH design – Van der Veer, 2002). Activity Theory is extended with temporal lifecycle models to describe collaborative information behavior. Value analysis is brought into an urban planning simulation environment. And so we find bricolage of theory taking place in deliberate attempts to improve the explanatory power of an adopted framework, or to extend that which has already been used in prior work.

In the final analysis, design goals and research goals differ, theories are not necessary for successful field research of collaboration. In many cases, the object may be to evaluate the effectiveness of a design intervention, through introducing and evaluating prototypes or early product designs. In these cases, the designer can be theory agnostic and effective. In other cases, we might find a successful design or social system having reached a peak or stasis. We turn to theory to explain the situation and to articulate innovations that offer possible breakthroughs to the community of use. We see we should "try on" theory, that theory should serve the research or design problem, not the other way around.

2.3 Values in Design

Finally, we found a trend toward developing means of understanding values in the domains of study. There were raised some specific research questions for studying values in the context of field research for design. Values of interest can be defined across a range of individual and social contexts – values of the users and designers, universal and particular human values, organizational and cultural values. Discourse revealed the requirement for developing models for values in different research settings. While values studies may be more typical in sociological studies, product design, and even business research, they are still an emerging issue in CSCW and field research in collaborative work. It remains to be tried whether traditional values measures and models (e.g., Rokeach, 1973; Cameron and Quinn, 1998; Maslow, 1970) may even be applicable to the complex social and work domains of the research projects covered within the workshop. What frameworks afford sensitivity to both individual and group values? At this point we remain aware of the trend and open to sharing from experience and effective approaches used across the wider community of researchers.

3. ANALYZING COLLABORATIVE WORK – A DISTRIBUTED COGNITION FRAMEWORK

Although several theoretical frameworks were raised in the first workshop, we considered it useful to explore the contribution of Distributed Cognition (Hutchins, 1995) collaborative analysis issues. Distributed Cognition has emerged as a significant theoretical framework for analyzing and understanding collaboration, and sheds light on the issues of analysis of holistic team interaction, multiple actors and artifacts, temporality and the organizational context of individual activity. We invited discussions about how this framework is used in practice of research and design, what problems or challenges it raises.

Distributed Cognition (DC) offers a theoretical and methodological framework with a broadened cognitive perspective, emphasizing the integrated cognitive system formed by people and their artifacts. DC analysis analyzes two relevant functions in particular:

- representations of internal and external knowledge
- propagation of knowledge between the individuals, artifacts and the environment as well as the transformations of information and knowledge resources during the activities.

A distributed cognition analysis identifies the shared cognitive components and artifacts, and describes explicit and implicit coordination among people in collective practices. DC analysis also specifies breakdowns and critical events occurring within processes and cognitive systems distributed among multiple participants. The outcome of analysis formulates proposals (recommendations) for functions to be preserved and identifies design and redesign of tools and practices in order to improve collaborative work. Distributed cognition approaches are considered useful for describing interactions among conscious actors and interactive systems in complex processes, and accounting for the effects of distributed action, awareness, and decision-making.

3.1 Theoretical issues

In the workshop contributions we find DC accommodating related theories, conceptual frameworks, and the extension of previous models which intend to explain collaborative work. Among these are Activity Theory (Engestroom, 1990), Collective Mind theory (Weick, 1993), and conceptual frameworks of extended cognition and cognitive economy.

As a developing framework, Distributed Cognition shares much in its approach with Activity Theory both supporting the research and design of distributed systems. The primary differences can be considered minimal – DC analyzes artifacts and tools as integral to a cognitive system shared by all participants in the analysis. Activity Theory separates mediating artifacts as specific types of functional instruments, but accommodates the distribution of cognition across common activities. The DC unit of analysis embraces the cognitive system, which includes activity. Activity Theory focuses on a specific domain of activity, which includes distributed internal and external cognition. Both frameworks have support in the workshop's research agenda. The analysis of the workshop contributions did not revealed so much problems or challenges (related either with the theory or the methods) as trends and applications.

Distributed Cognition accommodates Weick's collective mind theory in one paper, integrating principles of collective action and communication from organizational studies. The concept of "collaborative elasticity" drawn from this integration of DC and organizational systems presents a proposal for understanding the effective organizational resilience desired in high-reliability and critical missions. Collaborative elasticity draws from six dimensions of individual and team/organizational behavior that support flexible coordination among multiple stakeholders. Failures to support these dimensions explain breakdowns in collective awareness and action, as analyses from procedural collapse or mission disasters have shown.

By combining DC with the "joint activity" concept, another proposal extends the cognitive engineering (CE) model to support the redesign of social behavior, with a special focus on coordination mechanisms for shared values and goals. Drawing on Speech Act Theory, this approach analyzes the coordination mechanisms used in the process of driving and managing traffic systems, proposing coordination of individuals in traffic as a complex social activity. A game theory approach was also considered as a model for promoting common goals among individuals engaged in the traffic system, as a means of coordinating desired social behavior, minimizing conflict, and avoiding accidents.

Another contribution introduces *extended cognition*, analyzing the social and individual spatial relationships of the socio-technical cognitive system. Extended cognition considers properties of intelligent behavior and information distributed over physical space and over time, elaborating beyond and supporting the individual, and extending the social. Four social science research traditions have contributed to extended cognition, a concept that has reappeared with significant support after its submergence during the individual cognitive and behavioral traditions of the 20th century.

Extended cognition is presented in this case as enabling collaboration in physical and virtual space, and analyzes artifacts as referents for understanding coordinated activity. For example, some theoretical contributions place importance on the artifact or "external representations" of the socio-technical system defining three types of collaborative work spaces: physically centered space, virtually maintained space and a locally distributed space. As such, we find in these workshop contributions both naturalistic observation and theoretical analyses of physical and cognitive space, and of time-based analyses of activity.

3.2 Applications

The workshop contributions demonstrate a variety of domains for applications of distributed cognition in collaborative or coordinated activity. While these papers to some extent share a similar unit of analysis - the socio-technical system – the compelling applicability and flexibility of the DC conceptual framework shows its range from small systems (team) to large (a whole organization) and even continuous systems (traffic). The papers suggested widely differing applications for DC analysis:

- Computer product design and design team organization
- Scientific research, and intellectual collaboration for knowledge production
- Driving in high-volume traffic and intelligent highway systems (ITS)
- High reliability systems (aircraft crews), missions (firefighting), and organizations

We should note the significant application challenges ensuing from these analyses and from the broadened understanding gained from discussion. The primary challenge may be *adapting appropriate design approaches and methods* that can be shown to afford the identified performance, usability, and safety benefits realized in the analysis of distributed cognition. A major challenge includes learning from DC to extend application to the *analysis and design of social systems*, from high-performance teams to elastic organizations, and perhaps to large-scale and continuous social systems such as government and travel systems.

We might address the trend toward designing applications and tools (such as mobile technologies) that enhance awareness of social interactions with distributed communities, artifacts, and communication systems. Think of Smart Mobs in large urban spaces, for example, a phenomenon that emerged from the adoption of mobile phone and text tools for coordinating the spontaneous emergence of crowds at a designated target location. One contributor identifies Common Information Space (CIS), a CSCW concept addressing the orientation of multiple participants toward a common view of multiple and interacting channels of information. The use of status monitors in instant messaging services also addresses the ability to organize multiple channels and multiple participants in a distributed information environment.

However, we encounter significant gaps in applying knowledge of collaborative activity to complex systems and technology design. The dynamic complexity of collaboration remains inaccessible to traditional structures of task and individual analysis. Analytical representations (such as task and process analysis) fail to capture the possibilities of intent, contingency, and action among multiple participants interacting in shared information space. Even the commonly accepted toolkits of Contextual Design have been noted (in the first workshop) as useful for some analyses, but ineffective for others, not having been intended as a methodology for designing tools for collaboration. Numerous CSCW projects demonstrate this gap by reporting ethnographic studies that make a conceptual leap to prototypes or tool adoption without offering an analytical method for identifying the intentions and functions within the collaborative domain. Therefore, we claim the requirement to continue testing and reporting new analysis and design methods specifically for these types of ill-structured socio-technical systems.

4. CONCLUSIONS

We should recognize that the theories and frameworks drawing on distributed cognition represent a significant break from past models of individual cognition and information processing. The field can be seen as in transition, where in the last decade CSCW has developed from interdisciplinary niche to a source of influential research and theoretical directions (e.g., the rise and acceptance of DC and Activity Theory). CSCW research has even been successful in predicting many of the key trends in the post-Internet information ecology (always-on coordination, work ecologies integrating IM and email, integrated video and computer communication, spam). While many CSCW technologies have not proven successful, we should appreciate the wide adoption and influence of the theories and research approaches fostered by CSCW.

Research methods continue to show development, as we find continuing advancement of multi-method approaches integrating ethnography, field research and field experiments, design research, Contextual Design, and lead user research models. Where multi-method studies were uncommon in practice and the literature even 5 years ago, they now appear to be the norm in collaboration research. We expect this trend to continue, and will continue to share learning and advances in the current workshop format.

However, the methodological toolkit requires continual reflection and improvement. The reliance on triangulating multiple methods, while getting the job done also highlights the weakness of current methods to study the complex social phenomena of collaborative work. The analysis of distributed cognition in collaborative work is inherently complicated with multiple participants and their motivations, contingencies of situation, and technology use. We cannot yet consider the analysis and design toolkit sufficient to our domains and the quest to understand.

Finally, we find advantages as well as disadvantages in the inter- and multi-disciplinarity of the field. Learning and adopting methods from multiple disciplines contributes greatly to the breadth of the research agenda, and offers broad support for the creative research approaches required investigating collaborative activity and technologies. However, such continued interdisciplinarity shows up in some lack of influence in academic research agendas, traditional scholarly journals, and related funding. Promising projects investigating collaboration may be squeezed into a narrower technology focus, or find publication by narrowing the scope of findings. Given that this workshop has found hosts in both CSCW and CSAPC conferences, and similar workshops are sponsored at other international conferences, we are encouraged by the broadening appeal of the DC research stream into domains beyond CSCW.

5. REFERENCES

[1] Andriessen, J., H., E. (2002). Working with Groupware. Understanding and Evaluating Collaboration Technology. London: Springer Verlag.

[2] Cameron, K., S. and Quinn, E., R. (1998). Diagnosing and changing organizational culture. Addison-Wesley Publishing Company.

[3] Emgestroom, Y (1990). Learning, Working and Imagining. Twelve studies in activity theory. Orienta-konsultit. Helsinki.

[4] Hutchins, E (1995). Cognition in the wild. MIT Press. Massachutes, 1995b. 2nd ed.

[5] Maslow, A., H. (1970). Motivation and personality (2nd). NN. Y.: Harper & Row.

[6] Rokeach, M. (1973). The nature of human values. New York: Free Press.

[7] Van der Veer, & Van Welie, M. (2000). Task based GroupWare design: Putting theory into practice. Proceedings of DIS 2000, N. Y.

[8] Weick, E., K. (1993). Collective Mind in Organization: heedful Interrelating on Flight Decks. Administrative Science Quarterly, vol. 38, issue 3. Pp. 357-381.

9th European Conference on Cognitive Science Approaches to Process Control

Cognition and Collaboration - Distributed Cognition in Complex Processes

Gerrit C. van der Veer

Johan F. Hoorn