EACE ~ ECCE10

Tenth European Conference on Cognitive Ergonomics

Confronting Reality

Sweden: August 21-23, 2000

European Association of Cognitive Ergonomics

Technical Chair:
Erik Hollnagel

Sponsored by:
Graduate School for Human-Machine Interaction
at Linkoping University
European Office of Aerospace Research and Development

Proceedings Edited by:
P. Wright, S. Dekker, and C.P. Warren

ECCE 10:
Confronting Reality

Proceedings of the Tenth European Conference on Cognitive Ergonomics

Linköping, Sweden, 21-23 August, 2000

Edited by
P. Wright, E. Hollnagel & S. Dekker

ISBN 91-7219-828-1

Table of Contents

Welcome to ECCE-10 - the Tenth European Conference on Cognitive Ergonomics

ABOUT ECCE-10

The theme of ECCE-10 is "Confronting Reality", reflecting the definitive transition from "cognition in the mind" to "cognition in the world". The European Conferences on Cognitive Ergonomics began with the problems of computer-user interaction in 1982 but soon (ECCE-4 and ECCE-5) extended the scope to a concern for how people use and relate to artifacts other than computers, and how this affects the quality of work. The two most recent conferences, ECCE-8 and ECCE-9, broadened the view to consider cognitive ergonomics in relation to the worksystem and to co-operation. In the past decade the growing use of information technology has significantly changed both technical and social work environments. Activities are no longer confined to the local place of work, but may extend over considerable distances and involve collaboration with people who are not physically present. Indeed, we may not always know whether we are interacting with a human or a clever machine, both because machines have become smarter and because the pace of work has constrained the nature of the interaction. Confronting reality therefore has two meanings. First, that cognitive ergonomics must address the use of technology in many different areas of application, from the classical ones to new and challenging services and types of support. Second, that cognitive ergonomics must acknowledge that reality itself is changing from a world of tangible technology to a world where everything can be a representation of something else.

The papers at ECCE-10 are grouped into a number of sessions, with the following titles: Interaction Design, Collective Activities, Controlling Processes, Planning and Doing, Flying Machines, and Safety and Errors. These sessions contain papers that address both theoretical and practical issues, and in the tradition of the ECCE conferences ample time is provided for both presentations and discussions. The best papers, judged by from their contents and presentation, as well as the discussion they generate, will be considered for publication in the journal of Cognition, Technology & Work. This will provide an opportunity to disseminate what happened at the conference to the larger international community of cognitive ergonomics and cognitive engineering.

ECCE-10 differs from the previous conferences by being is arranged in collaboration with the annual conference of the Swedish Centre for Human Factors in Aviation (HFA) – HFA 2000. The idea to do so came about because of the obvious common interests between the two conferences. The collaboration is reflected concretely in the last day of the programme, which includes a special session on cognitive ergonomics applied to aviation. When ECCE-10 ends, the HFA conference will continue with its annual meeting. The HFA centre has also been instrumental in arranging the conference reception at the Air Force Museum in Linköping. The willingness of the HFA centre to try out this collaboration between the two conferences is greatly appreciated.

ABOUT EACE

The ECCE-10 conference is organised by EACE - the European Association for Cognitive Ergonomics. The ECCE series of conferences has been organised every two years since 1982, thereby making it the oldest conference series of its kind internationally. In addition, EACE also organises a conference on Cognitive Science Approaches to Process Control (CSAPC), of which the next will take place in Munich, Germany, in 2001. The CSAPC series of conferences is also biennial, and the first conference was held in 1987.

The purpose of EACE is to bring together European researchers in the domain of cognitive ergonomics, which aims to combine cognitive engineering and technical information processing system developments to improve the design of joint, interactive systems. EACE is mainly composed of members who are confirmed researchers (i.e., who have published research related to Cognitive Ergonomics) regardless of whether they have a doctoral diploma, and probationers who are not yet researchers but registered as doctoral students and demonstrate the capacity to do research in Cognitive Ergonomics. Only members are entitled to vote in the General Assembly. The probationer status is considered as temporary; a probationer should become a member within a five-year period. Because of the European nature of the association, a member or a probationer must be a national of or domiciled in a European country. However, nationals of countries outside Europe may become corresponding members.

Membership of EACE, which currently is 40 Euro per year, is an excellent way of getting in contact with the cognitive ergonomics community in Europe. In addition, EACE members receive the regular EACE Newsletter, enjoy reduced fees to the conferences organised by EACE, as well as a number of other benefits. Anyone interested in joining EACE is encouraged to contact the membership officer at the following address: Elly Lammers, Free University, Amsterdam (email: elly@cs.vu.nl).

ACKNOWLEDGEMENTS

Anyone who has ever been involved in organising a conference will know that it depends on the hard work of a number of volunteers. This conference is no exception to that, and it is only proper that full recognition is given to the following individuals.

Technical Programme Committee

Rene Amalberti; Sebastiano Bagnara; Lisanne Bainbridge; David Benyon; Guy Boy; Jean-Marie Burkhardt; Carlo Cacciabue; Francoise Darses; Francoise Detienne; Martin Helander; Jean-Michel Hoc; Alfred Kobsa; John Long; Neville Moray; David Novick; Reinhard Opperman; Matthias Rauterberg; Janine Rogalski; Gerrit van der Veer; Peter Wright; and Michael Tauber.

International Organising Committee

Liam Bannon, Jose Cañas, and Clive Warren.

Local Organising Committee

Sidney Dekker and Nalini Suparamaniam, with administrative support from Christina Karlsson (IKP) and Birgitta Franzén (IDA).

One of the most important contributions to a successful conference comes from the Editor of the Proceedings. In the good old days, authors would submit the printed manuscripts, and there was little to that could be done if they had not followed the rules – and authors at the ECCE conferences are as unwilling to follow instructions as anyone else! In these modern times, papers are submitted electronically. This makes it possible to produce a set of proceedings that looks more professional, but only at the cost of a significant effort from the editor. This time the task was performed by Peter Wright, who did an excellent job with valuable support from Chris Fairburn and Julia Hill.

Finally, we wish to thank the following for their contribution to the success of this conference:

- European Office of Aerospace Research and Development, Air Force Office of Scientific Research, United States Air Force Research Laboratory. (EOARD)
- Graduate School for Human Machine Interaction, University of Linköping (HMI)
- Swedish Centre for Human Factors in Aviation (HFA)
- European Association for Cognitive Ergonomics (EACE)

Welcome to Linköping and to a stimulating and enjoyable conference.

Erik Hollnagel,
Chairman, EACE

Session #1:
Interaction Design

Interacting with a Personal Wearable Device

Geert de Haan
IPO, Center for User-System Interaction
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven
The Netherlands
G.d.Haan@tue.nl
http://www.ipo.tue.nl/homepages/gdhaan/

ABSTRACT

Comris is a research project that aims to create a wearable assistant, "the parrot", for conference and workshop visitors. A personal interest profile and an active badge system enable agents in a virtual information space to provide context-sensitive information about interesting persons and events to conference and workshop visitors in their own physical information space.

This paper describes a usability experiment for the interaction design of the parrot device as a device that should not distract the users from their regular activities and provide the users with a sense of being in control. We describe the usability evaluation of 4 prototypes for the design of the user-parrot dialogue, each based on a different principle for function-to-button allocation. Because the parrot device was still being developed, the evaluations were performed in simulated settings.

The results of the experiment are related to the importance of conceptual aspects versus perceptual aspects in user interface design methods, and in particular the role of perceptual and dialogue consistency.

Keywords

Auditory I/O, hand-held devices, intelligent systems, mobile computing, consistency, prototyping, dialogue design, speech and voice, usability evaluation.

INTRODUCTION

COMRIS (Cohabited Mixed Reality Information Spaces; Van der Velde, 1997) is a research and development project that seeks to develop a wearable assistant for conference and workshop visitors. On the basis of a personal interest profile and an active badge system, conference visitors receive context-sensitive information about interesting persons and events in their spatio-temporal vicinity.
For example, a person who has indicated a high interest in wearable computers on her home-page or on the web-form that is used to specify one's personal interest profile may receive a message that "a demonstration of the parrot, a wearable conference assistant, is about to start in 5 minutes at the Comris booth".

To the user, the most tangible part of the Comris system is a small, personal and wearable device, "the parrot". The parrot consists of a small box that is worn on the belt, an audio device for speech output that is worn at or near an ear, and an interactive device which probably will be worn on a wrist. The wrist-worn device provides a small LCD screen and a few buttons to fine-tune messaging defaults to the circumstances and to delete, repeat and respond to messages.

To the system, the parrot also functions as an active badge (Want et al., 1992) which, by means of a number of infrared beacons helps the system to keep track of the whereabouts of people. Knowing where people are enables Comris to provide user with advice about meeting other people or to inform users 'just-in-time' about events which are about to take place at distant locations.

Conference visitors also have a number of information kiosks at their disposal to enter and adjust their personal interest profiles, as well as to consult the conference schedule and their personal agenda.

It is only indirectly that the user may notice the virtual agent space which forms the software side of the 'cohabited information space' that Comris is concerned with. In this space, agents represent the interests in events and persons from the interest profiles of the users. Every one in a while, the user's personal agent collects the matching interests and selects the most competent match on the basis of interest-value and contextual-fit. The selected interest match is passes on to the user's parrot for language and speech generation, and for presentation as text on the LCD screen.

USABILITY ISSUES

The primary motivation for the Comris project is to investigate the opportunities for using advanced information technology, such as wireless networking and intelligent agent systems, in everyday life. The practical aim of the project is the development of a wearable device. Comris is also a means to study the application of the technology, including the social and technical usability issues of interacting with intelligent assistants.

Usability evaluation in Comris focuses on the wearable parrot since this device is the most frequently used as well as the most direct and important way to influence the behaviour of the system.

Three usability experiments were planned, each with a similar set-up. The first experiment concerns the user preferences regarding message presentation i.e. dealing with the results of information push. The second experiment investigates ways that allow users to set interruption levels and to adapt the number of messages to the circumstances. The third experiment investigates miscellaneous functionality and interaction issues with the Comris parrot.

The results of the first experiment (de Haan, 1999) indicate that, in general, users dislike information push, especially when it concerns spoken information; they prefer to determine themselves when and where to listen to messages rather then having messages presented or announced under the control of the system. In short: in conference settings, particularly when engaged in social activities, users have a strong preference for email-wise information over telephone-wise information.

To the Comris project two implications follow from the results of the first experiment. First, in providing advice information, more concern should be given to the user's context, for example, by having the parrot eavesdrop to determine whether or not the user is engaged in a conversation.

Secondly, more weight should be given to the design of the user facilities to interact with the system: the parrot should no longer be seen primarily as a speech output device but more as a facility to interact with and control the system. Given that users should thus be able to control interruption levels and messaging, the second and third experiment were combined into one experiment to investigate all dialogue aspects at once.

When performing the experiment the parrot device was not yet available. Consequently, rather than the wearable device itself, the experiment used a simulated parrot device on a PC screen and a mouse.

INTERACTION WITH THE PARROT

At the start of the project a number of hardware and functional requirements had been specified. The parrot should be equipped with an audio output device like a headphone and with a small display screen providing room for 2 lines of 16 characters (or icons), which should be suitable for the synchronised presentation of messages in spoken and visual/textual form. The parrot should have a volume control and the display should provide labels to indicate the meaning of each of the five buttons to allow the users to manipulate message presentation and other settings.
Furthermore, the parrot should retain a list of messages for the user to walk through. This requirement came under scrutiny by the information-push idea but the results of the first experiment re-established its relevance. In the second experiment, the following requirements on feedback and parrot controls were used:

- adjust the volume (via a separate slider);
- repeat a message;
- interrupt or postpone a message;
- suppress or delete a message;
- raise or lower the context threshold (how interesting messages must at least be);
- move 'up' to the next message;
- move 'down' to the previous message;
- give a positive or "yes" response;
- give a negative or "no" response;
- elaborate a message (present additional information about the content of a message).

It was not certain if the agent system could provide for the last function. Instead, it was decided to add an option to show a message in abbreviated form or to have the complete message scrolled by on the display and/or spoken out via the audio channel.
Finally, since it is clear from the list that there are more functions than there are buttons available, alternative function-to-button mappings are required for which a mode function is added.

The complete list of functions is as follows: to change mode, to show a message, to play and stop a message (audibly) playing, to start and stop a message (visually) scrolling, to respond with yes, to respond with no, to select the next message, to select the previous message, to delete a message, and to set the context threshold higher (+) or lower (-).

Four different interfaces were designed, each according to a general principle to deal with user interfaces in which there are more functions to accommodate than there is room for buttons available. For each user interface, the principle of mapping functions to functions is described, followed by a description of the behaviour of the interface, and a scheme of the dialogue.

Mode interface

The mode interface uses a dedicated key to switch (toggle) between two modes in which buttons represents different commands. After selecting a command, the system reacts, for example, by playing a message or by showing the current threshold as a row of #'s, and the system returns to the starting point of the applicable mode.

Mode 1 display:

X	delete message
>	play and scroll message (and stop)
<<	previous message
>>	next message

Mode 2 display:

-	decrease threshold
+	increase threshold
N	respond NO
Y	respond YES

Dialogue interface

In the dialogue interface, the user selects a general function which is further specified through a series of interaction or dialogue steps. After selecting a general function, the user can select a specific command, for example, to show the current message or to move to the next or the previous one. After selecting a specific command, the system reacts but remains in the general dialogue until the user selects a command to move up.

Main display:

>	play and scroll message (and stop)		
X	delete message		
	UP	goto main dialogue	
	Y	delete message	
	N	dont delete message	
yn	respond		
	UP	goto main dialogue	
	Y	respond YES	
	N	respond NO	
+-	set threshold		
	UP	goto main dialogue	
	-	decrease threshold	
	+	increase threshold	
<>	move to		
	UP	goto main dialogue	
	<<	previous message	
	>>	next message	
	=	show short message	

Menu interface

The menu interface is a combination of the mode interface and the dialogue interface with one button used to walk through a menu of the general functions, each presenting a yes/no/cancel choice and a choice to move to the next or to the previous message. After selecting a particular choice, for example, to "yes" delete the current message, the system reacts but it remains in the general dialogue until the user selects the command to move to the next general function. The general functions are arranged in a linked-list, and, as such, there is no permanent starting point.

Play msg display:

M1	next menu entry
Y	play and scroll message
N	dont play and scroll (stop)
<<	previous message
>>	next message

Show msg display:

M2	next menu entry

Y	scroll message
N	dont scroll (stop)
<<	previous message
>>	next message

Answer display:

M3	next menu entry
Ok	respond YES
No	respond NO
<<	previous message
>>	next message

Delete msg display

M4	next menu entry
Y	delete message
N	dont delete message
<<	previous message
>>	next message

Number of msgs display

M5	next menu entry
-	decrease threshold
+	increase threshold
<<	previous message
>>	next message

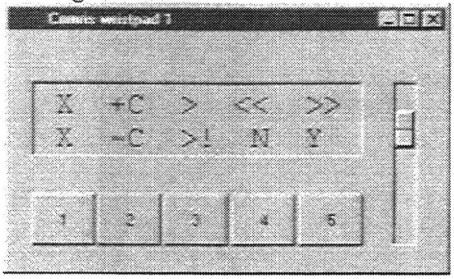
Mbutton interface

In the mode button interface, a particular mode is activated by pressing a special mode key simultaneously with the function key. This interface was included only for comparison purposes since it would not be possible to implement it on the parrot because this would require an additional button. With respect to usability, the Mbutton interface is not desirable because, as a walk-up-and-use interface (Lewis et al. 1990), the interface should not require users to learn artificial key-bindings or double-clicking.

In the experiment, the different modes of the Mbutton interface were assigned to the left and the right mouse-button, such that the user could select the more general and frequent commands, such as, moving to the next or the previous message, with the left-button, and the less general and infrequent commands, such as responding to a message, with the right-button. After selecting a command, the system reacts, for example, by scrolling or by playing-and-scrolling a message, and the system returns to the starting point.

Figure 1 shows the Mbutton interface as an example of the wrist-worn device. The slider is used to control the volume.

Figure 1: The Mbutton interface



Main display:

X	X	delete message (left / right button)
+C	-C	increase / decrease threshold
>	>!	scroll / play and scroll msg (and stop)
<<	N	previous message / respond NO
>>	Y	next message / respond YES

The interfaces were designed such that only the last action was processed every 250 milliseconds that the interface was not busy (selecting "play a message" for three times only plays it once). After 2 seconds of inactivity after presenting a message, the Mbutton interface, the dialogue interface, and the menu interface showed the labels of each button; for the mode interface it was assumed that pressing the mode button would be a more convenient way to determine one's whereabouts.

Expectations

The different user interfaces were created with the sole purpose of selecting the best one with respect to performance characteristics and user preferences. Apart from the Mbutton interface, which was created to serve in intuitive ground as the base level, no hypotheses were created concerning performance and preference scores.

There are however in methods for user interface design, several differences between the user interface prototypes which are of theoretical interest with regard to the relevance of the specification of conceptual and task elements versus the relevance of the perceptual or graphical elements. We advocate the use of formal methods in user interface design because they offer opportunities to create engineering approaches to design which makes user interface design less dependent on iteration (or trial-and-error) and individual expertise (de Haan, 2000).

A problem in methods for user interface design which employ formal specifications is that the design of the perceptual interface is either left to external methods and tools (cf. ETAG-based design; de Haan, 2000; ConcurTaskTrees; Paternò, 2000) or it is treated as an aspect of design which is only meant to explain and convey the conceptual interface (cf. MUSE; Lim and Long, 1994; ADEPT; Wilson et al., 1993).

Evidence from e.g. Cockayne et al. (1999) and Holst et al. (1997) indicates that the perceptual design rather than the conceptual design is essential to the user's understanding. This supports design methods like de Greef and Neerincx (1995) which put less emphasise on the conceptual interface and more on the perceptual interface, as a source of information about task performance.
The main differences between the user interface prototypes concern the functionality, the amount of guidance, the number of dialogue steps, the consistency, and the amount of information they provide about the available functions. Of these, methods which focus on the conceptual interface

tend to emphasise the functionality, the number of dialogue steps and the consistency whereas methods which focus on the perceptual interface will emphasise on information support.

There are only small differences in functionality between the interfaces. All the user interfaces offer 9 or 10 functions. These differences are due to the need for a navigation function to change the mode, for example, and whether or not functions are provided to show a message in abbreviated form and to present a message as scrolling text without having it spoken out. On the basis of functionality, few if any differences performance should be expected.

There are some differences in the way the interfaces provide guidance through the dialogue steps and, possibly, prevent the user from making errors. Both the Mode interface and the Mbutton interface merely execute the function that is selected, and the Mode interface even allows that functions are selected without that the associated icon is visible on the LCD screen.
The Dialogue interface provides direct access to 2 functions, to play and to delete a message, respectively. This is similar to the Menu interface, except that only the 'harmless' functions to move to the previous or next message may be invoked without answering a Yes/No question.

The interfaces differ in the (average) number of dialogue steps that are required to invoke a function. Note that there is a trade-off between guidance and the number of dialogue steps. The Mbutton interface requires only 1 step or button-press to invoke a function. The Mode interface requires between 1 and 2 steps because it may be necessary to change modes first.

The Menu and the Dialogue interface each require between 2 and 3 steps to invoke a function. The exact number depends on set of tasks to perform: on the one hand, in the Menu interface the functions are arranged in a one directional linked list and locating a function may require up to 4 steps, but on the other hand, within each function 'mode', the Menu interface also provides the frequently-used functions to move to the previous or next message. With respect to the task assignments of the experiment, the Menu interface is slightly more efficient.

In addition, the user interfaces differ in the consistency of the human-computer dialogue. In terms of the smallest number of exception rules, the Menu interface is best, followed by the Mbutton interface which has a visible but not completely consistent distinction between the functions associated with the left and with the right mouse button. The Dialogue interface is slightly less consistent because the distinction between direct and indirect functions is not semantically motivated but by the number of available buttons. The Mode interface with its possibly invisible modes is the least consistent interface.

Finally, the interfaces differ with respect to how much information they provide about the available functions. Since the interactive part of the parrot is worn on the wrist, information about the current state of the dialogue is hard to miss. As such, information about the non-visible parts of the interface and overview information may become more important to the usability of the interfaces.

Not considering the navigation functions, at any given dialogue step, the Menu interface only shows 3 out of 9 functions. The Mode interface shows 4 functions and the Dialogue interface shows 5 functions or function groups. The Mbutton interface, however, presents for all of its 9 functions in iconic form.

THE EXPERIMENT

Method

Instructions and procedure

At the start of the experiment, participants read a short introductory text which briefly explained the Comris system and described the purpose and the procedure of the experiment.

At the start of each trial, participants were provided with a one-page explanation which showed pictures of the main modes of the user interface with a description of the functions and the symbols used to represent the functions. It asked participants to carefully read the instruction but not learn it by heart. Once read, the instruction was turned upside down and replaced by a task assignment with a number of tasks and the session questionnaire. Participants were asked to work at a steady pace while trying to avoid mistakes.

The session questionnaire asking the subjects to rate the user interface on ease of use, pleasantness, learnability, how much the he or she felt in control over the interface. The questionnaire also asked for remarks about the interface.

At the start of the first trial one or two example tasks were demonstrated and the system was restarted.
At the end of each trial, participants were provided with the next instruction, task assignment and user interface prototype. At the end of the last trial participants were provided with the post-session questionnaire. The post-session questionnaire asked the subjects for a comparative rating of the user interfaces on, respectively: ease of use, pleasantness, learnability, control over the interface, efficiency, de degree of confusion, complexity, and how cumbersome each interface was

Tasks

Each task assignment consisted of 10 tasks, starting with an easy task. Each task asked participants to do something to a message such as reading or removing it, after locating it by content such as by the name of a person or event. Task assignments contained the same types of tasks. Except for the first task, the order of tasks in each assignment was independently randomized. Examples of the tasks are as follows:

- Where is the demonstration of the Comris system taking place?
- A meeting starts, ensure that you are only disturbed by urgent messages;
- Reject the meeting proposal about object orientation.

Design

Four different user interfaces were used with 4 different task assignments in 4 trials. In four consecutive sessions, each subject used a different interface to perform the tasks. The presentation of interfaces over trials and subjects was balanced according to a latin-greek design to control for order effects.

The presentation of task assignments was balanced with user interface presentation as well as over trials and subjects (i.e. one half of the subjects received 1234 and the other half 4321).

Material and equipment

Each interface used the same list of regular Comris messages. For speech presentation the messages were presented by a high-quality female voice using AT&T's (1998) text-to-speech synthesizer for

US-English. The abbreviated messages, used for browsing through the list of messages, were created by leaving out any unnecessary words from the complete messages in order to make the messages fit the LCD screen.

The user interfaces were implemented in Tcl/Tk running on Ms Windows98. Each interface was presented on a standard PC display screen and subjects only used the mouse to interact with it. The experiment took place in a regular office workplace.
In the case with the Mbutton interface the left and right mousebuttons were used to distinguish between the two functions of each button (note: it was hypothesised that the Mbutton interface would be worst; but in these conditions it came out differently).

Subjects

Sixteen colleague researchers at IPO/Center for User-System Interaction participated in the experiment on a voluntary basis. All subjects, 10 males and 6 females, were experienced computer users who spoke English as their second language. The subjects were rewarded for their participation in the experiment with plums and apricots.

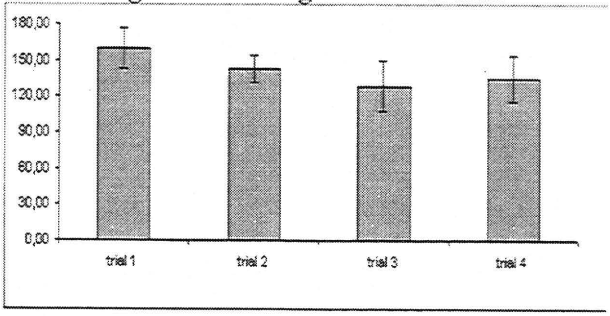
Results

Results consist of the performance data as the number of steps required to complete the tasks, and the rating scale data of the usability questionnaires which were presented at the end of each session and at the end of the experiment. The open questions of the questionnaires did not produce sufficient data to allow for a systematic analysis.

Performance Data

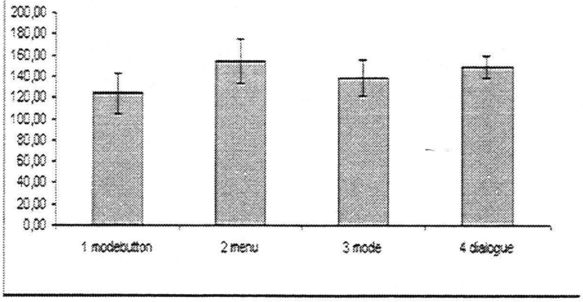
Three analyses were performed on the performance data, one analysing the results between instruction materials to investigate instruction effects, one analysing the results between trials to investigate learning effects, and one analysing the results between interfaces to investigate the effects of using a different user interface.
An analysis of the results over instructions shows that there is no effect of instructions : $F(3,45) = .908$; $p =$ ns. An analysis of the results over trials shows a significant learning effect: $F(3,24) = 4.476$; $p = .012$. The test also shows a significant between-subjects effect of the ordering of user interfaces: $F(7,8) = 5.214$; $p = .017$ but according to an analysis of the interaction effects, there are no interactions between learning and the ordering of user interfaces $F(21,24) = 1.461$; $p =$ ns. or between learning and the ordering of instructions: $F(3,42) = .439$; $p =$ ns. Figure 2 shows performance over trials with an indication of the 5% confidence intervals.

Figure 2: Learning effects over trials



An analysis of the results over interfaces shows a significant effect of using the different user interfaces: $F(3,45) = 3.688$; $p = .019$. Figure 3 shows performance with the different user interfaces with the 5% confidence intervals.

Figure 3: User interface performance differences



Session Rating Scales

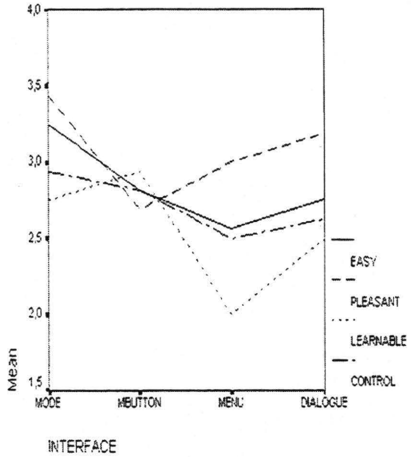
The session questions were as follows:

How easy to use do you evaluate this interface?
How pleasant do you evaluate this interface?
How easy to learn do you evaluate this interface?
How much do you feel in control of the interface?

Questions were answered on a 5 point scale:
[Very much] 1 -- 2 -- 3 -- 4 -- 5 [Not at all]

An analysis of the results indicated that the user interfaces differences are not significant: F(3,45) = 1.761, p = .168. The differences between the questions (read: sessions or learning) was significant: F(3,45) = 3.198, p = .005, and there was a significant interaction between questions (read: sessions) and interfaces, F(9,135) = 2.054, p = .038. In short: from earlier sessions, subjects learn to do the task during later sessions, and there are no overall better or worse interfaces. Figure 4 shows the results of the session rating scales (lower is better).

Figure 4: Session rating scale results.



The results indicate that, although the difference between the interfaces was statistically not significant, the Menu interface should be considered the best (M = 2.51), the Dialogue interface second (M = 2.77), the Mbutton interface third (M = 2.81), and the Mode interface fourth and worst (M = 3.09).

Post-session Rating Scales

The post-session questions were as follows:

- How easy to use do you evaluate each interface?
- How efficient do you evaluate each interface?
- How pleasant to work with do you evaluate each interface?
- How easy to learn do you evaluate each interface?
- How confusing do you evaluate each interface? *
- How cumbersome do you evaluate each interface? *
- How complex do you evaluate each interface? *
- How much do you feel in control of each interface?

Questions were answered by marking all the interfaces on one 5 point scale:
[Most] 1 -- 2 -- 3 -- 4 -- 5 [Least]
The results for the post-session questionnaires were first encoded into a uniform positive-negative scale e.g. "How confusing do you evaluate each interface?" was encoded as "How non-confusing do you evaluate each interface?". The thus encoded questions are marked with a "*" above. The results were analysed as a within subjects design with repeated measures with 16 subjects x 8 questions (repeated) x 4 interfaces design. There were 7 missing values which were replaced by the question/interface mean.

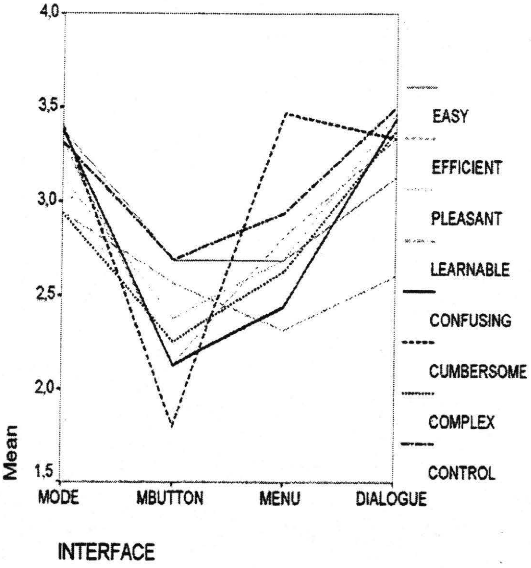
An analysis of the overall post-session results indicated that there was a significant effect of the interface type, F(3,45) = 3.362, p = .027, but no effect of the questions, F(7,105) = 1,208, p = .305, nor was there an interaction effect, F(21,315) = 1.278, p = .187. In short: there are better and worse interfaces. Figure 5 shows the post-session results (lower is better).

The results indicate that the Mbutton interface should be considered the best (M = 2.33), the Menu interface second (M = 2.75), and the Mode interface third (M = 3.22) and about equal to the Dialogue interface (m = 3.27).

DISCUSSION

There are 2 questions to answer. First, what is the best user interface for the Comris parrot. Secondly, what the results imply concerning the importance of the conceptual versus the perceptual aspects of user interfaces in the design process.

Figure 5: Post-session rating scale results



The best user interface is the Mbutton interface, which is surprising because it was created for comparison purposes and serve as the worst-case bottom-line. Our expectations were based on the need to learn to distinguish and avoid confusion between two sets of functions. The results indicate that with a consistent simple mapping between icons and motor actions little if any learning is necessary. As such, it serves to remind of the dangers of intuition.

In terms of the number of steps required to perform the tasks, including those to recover from errors, it is by far the fastest interface. The Mbutton interface also scores best on the post-session rating scales, except for ease-of-use, where it is second, and for learnability.

The Mbutton interface does not score very well on the session rating scales, where it scores third, after the Menu and the Dialogue interfaces. The session scores should not receive too much weight because the differences between the interfaces are too small to reach significance. The post-session rating scales are more important because they ask the subjects to make a comparative judgement and a forced choice.

The results leave the Comris project with the problem of implementing an interface with 2 functions for each button. It may be an option to use buttons with three states or, from the idea that there are regular and special functions (i.e. moving to the next message versus responding to a message), to distinguish between single and double clicks. If not feasible, it is possible to chose for the Menu interface, which, albeit less fast, scores better on learnability and ease of use on both the session and post-session rating scales.

With respect to the implications for user interface design it is surprising that the 2 best user interfaces are very dissimilar.

Had speed been the most important criterion, either in terms of the objective number of steps required or in terms of the subjective judgement about cumbersomeness then the Mode interface should have scored better than the Menu interface.

When information about the available functionality had been the most important criterion then the Menu interface should have scored worse than the Dialogue interface.

When we focus on the main differences between the Mbutton and the Menu interfaces on the one hand, and the Mode and Dialogue interfaces on the other, the main differences are in the consistency and in the information about both, the available functions and the user's whereabouts within the dialogue.

Regarding information about where one is, the Mode and the Dialogue interface provide some information about the users current position in the dialogue but the information is different from one state to another. In the Mbutton interface there is only one dialogue state and no information about the current state is required whereas in the Menu interface there are multiple states but in each state, users receive consistent information about where they are.

Regarding information about what may be done next, the Menu and the Dialogue interface provide some information about the possibilities but not always and not about all the possibilities. The Mbutton interface provides all the information and since there is only one dialogue state, it does so all the time. The Menu interfaces provides little information about what to do next but because there are only a few choices provided, and provided in a consistent way, the information may be sufficient.

The results of the experiment are is no way conclusive about this line of reasoning. It may be that the concept of consistency plays a very different role in the perceptual structure of information about dialogue-states and opportunities for interaction than it does in the structure of actions in human-computer dialogues. The results of the experiment allow us to tentatively conclude that the consistency of action and information in combination is important but this has to be further investigated.

REFERENCES

AT&T (1998). The Next-Generation Text-To-Speech Synthesiser. Available from http://www.research.att.com/projects/tts/

Cockayne, A., Wright, P.C. and Fields, B. (1999). Supporting Interaction Strategies Through the Externalization of Strategy Concepts. In: Sasse, M.A and Johnson, C. (eds.). Proceedings Interact '99, IFIP TC13, IOP Press, Amsterdam, pp. 582-588.

de Greef, H.P. and Neerincx, M.A. (1995). Cognitive Support: designing aiding to support human knowledge. International Journal for Human-Computer Studies 42, 531-571.

de Haan, G. (1999). The Usability of Interacting with the Virtual and the Real in COMRIS. Proceedings of the Twente Workshop on Language Technology 15 - Interactions in Virtual Worlds. Enschede, The Netherlands, May 19-21, 1999.

de Haan, G. (2000). ETAG, A Formal Model of competence Knowledge for User Interface Design. PhD. thesis, Dept. of Computer Science, Vrije Universiteit, Amsterdam.

Holst, S.J., Churchill, E.F. and Gilmore, D.J. (1997). Balloons, Boats and Ponies: Interface Manipulation Style and Learning in a Constraint-based Planning Task. In: Howard, S., Hammond, J., and Lindgaard, G.K. (eds.). Proceedings Interact '97, Chapman & Hall, pp. 180-187.

Lewis, C, Polson, P., Wharton, C. and Rieman, J. (1990). Testing a Walkthrough Methodology for Theory-Based Design of Walk-Up-and-Use Interfaces. In: Proceedings CHI'90, ACM, New York, 235-242.

Lim, K.Y. and Long, J.B. (1994). The MUSE Method for Usability Engineering. Cambridge University Press, Cambridge, UK.

Paternò, F. (2000). Model-Based Design and Evaluation of Interactive Applications. Springer Verlag, London.

van der Velde (1997) Co-Habited Mixed Reality. Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence. Aichi, Japan, August 23-29, 1997.

Want, R., Hopper, A., Falcão, V and Gibbons, J. (1992). The Active Badge Location System. ACM Transactions on Information Systems 10, no. 1, 91-102.

Wilson, S., Johnson, P., Kelly, C., Cunningham, J. and Markopoulos, P. (1993). Beyond Hacking: A Model Based Approach to User Interface Design. In: Alty, J.L., Diaper, D. and Guest, S. (eds.) Proceedings People and Computers VIII. Cambridge University Press, Cambridge, UK., 217-232.

Structured Method for Research and Design of an In-Car Information System

Adam Stork, Becky Hill, Andrew Life, and John Long

Ergonomics & HCI Unit
University College London
26 Bedford Way
London, WC1H 0AP, UK
{a.stork, b.hill, a.life, j.long}@ucl.ac.uk
http://www.ergohci.ucl.ac.uk/

ABSTRACT

We were recently commissioned to design an 'in-car information system' in conjunction with a major automobile in-car-entertainment manufacturer. The aim of the system was to enable car commuters to 'read' a newspaper while driving.

To design the system, we used a structured method and a model of how users plan and control multiple tasks (PCMT). The aim of the design was two-fold: to produce a prototype in-car information system; and to support development of the PCMT model and structured method. This last aim is a research aim, and we used an early version of a structured method configured for research (MUSE/R).

This paper describes the MUSE/R method, the PCMT model, the design, and the research to support development of the PCMT model and MUSE/R method.

Keywords

Reliable HCI design, structured methods, planning and control, in-car information.

INTRODUCTION

Many business people need to read a newspaper for their work. If they commute by train, they can read the newspaper on the train. However, if they commute by car, they can listen to the radio, but they will not necessarily hear the news that they are interested in, and would subsequently need to read the newspaper.

The advent of Digital Audio Broadcasting (DAB) gives the possibility of transmitting a large quantity of digital data to a receiver in a car. We were commissioned to design an 'in-car information system' to enable car commuters to 'read' a newspaper while driving.

To design the system, we used a structured method and a model of how users plan and control multiple tasks (PCMT). The aim of the design was two-fold:

- Produce a prototype in-car information system;
- Support development of the PCMT model and the structured method.

This last aim is a research aim, involving initial validation of the PCMT model and the structured method. We used an early version of a structured method configured for research, called MUSE/R—Method for Usability Engineering for Research.

This paper describes the MUSE/R method, the PCMT model, the design, and the research to support development of the PCMT model and MUSE/R method.

STRUCTURED METHOD FOR RESEARCH

Supporting Research

Stork and Long (1997) identified a potential role for structured analysis and design methods to support research as well as design. They saw that structured methods support design in the following ways, all of which are essential to research:

- Products that describe specific design cases;
- generalisations over specific design cases;
- applications of general knowledge.

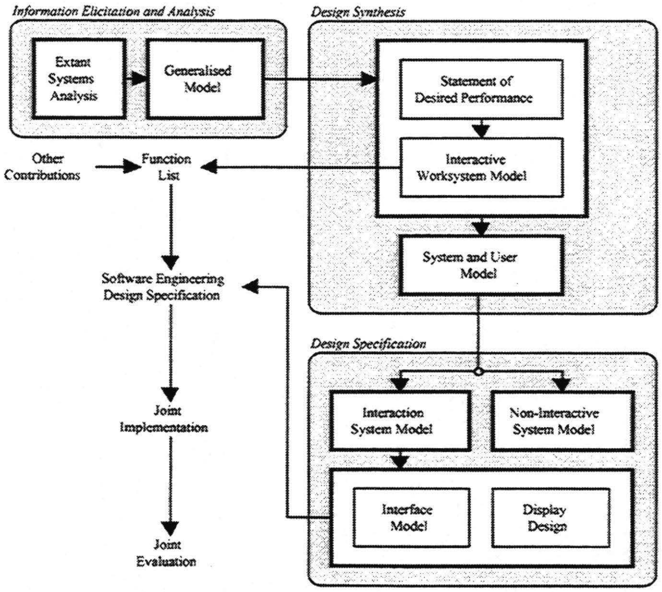
For example, the first could be generalised to give design knowledge, the second might be generic design knowledge, and the third would enable the assessment and validation of design knowledge.

MUSE/R

Stork and Long gave an example of a structured method for research, based on MUSE—Method for USability Engineering (Lim and Long, 1994)—called MUSE/R. Similar to MUSE, MUSE/R is a structured analysis and design method for human factors engineers. The product of MUSE/R is the specification of an interaction artefact.

MUSE/R used the concepts from Cognitive Engineering (CE) research (Dowell and Long, 1998), supported by a research strategy, which included applying and assessing design knowledge, based on those concepts (MUSE itself does not use these concepts).

Figure 1: Schematic of MUSE/R



The CE concepts include 'performance', the quality of the work done by an 'interactive worksystem' with the costs of achieving that work. The work is achieved in a domain of application, and consists of transformations of the attribute values of the objects in that domain of application.

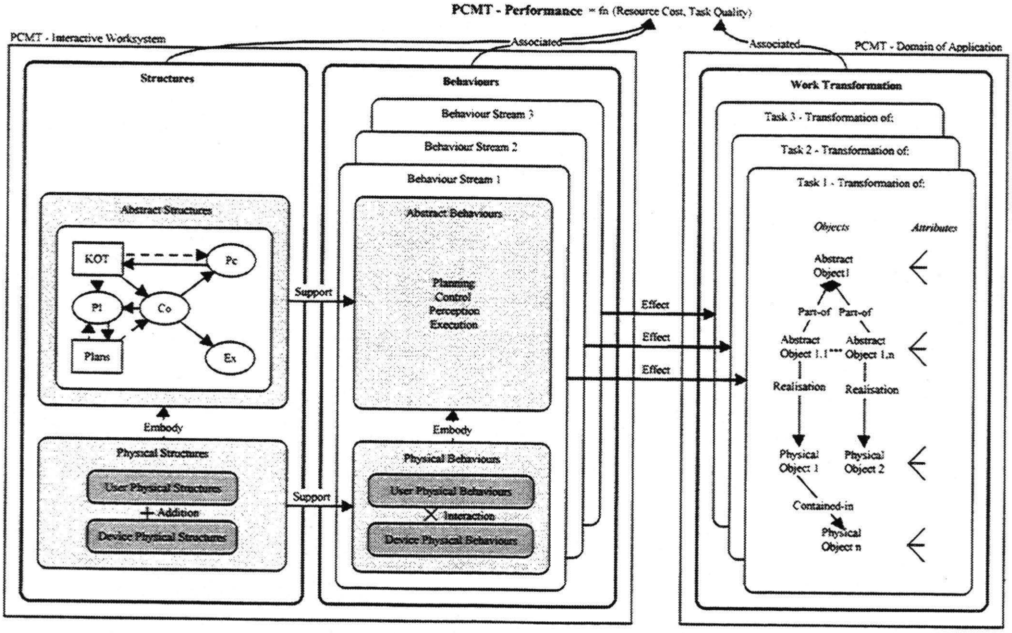
The interactive worksystem is a system of human and computer structures that exhibit behaviours which interact to achieve the work. The costs (e.g. workload) of achieving the work arise from this worksystem, for example, from structural changes, like learning, and behaviours, like thinking or moving. The products of MUSE/R represent these concepts for existing designs, the design problem, and the proposed design solution.

MUSE/R approaches design in a 'top-down' manner, based on information derived 'bottom-up'. It involves specification of the design products as defined by the method, employing specific notations. Design progresses from the specification of general features of the tasks and behaviours to be performed by the user, derived from analysis of the user requirements and from existing systems, to the specification of the details of the interaction artefact. The application of MUSE/R is considered to be an iterative process, both overall and internally, supporting the production of the best 'first-attempt' artefact, following the initial complete application.

Figure 1 shows a schematic diagram of the MUSE/R method. MUSE/R has three phases: the Extant Systems Analysis phase, the Conceptual Design phase, and the Detailed Design phase. These phases are described later in further detail, along with examples of the products of those phases.

MUSE/R is still under development. The scope of the phases and products is well-defined, but the procedures for generating the products are poorly defined as yet. Further, MUSE/R has not been used for design until this project. We were interested in applying MUSE/R to provide an initial validation of it and so determine whether to continue its development.

Figure 2: Schematic of PCMT



Initial Validation of MUSE/R

Previous validation of HCI knowledge (Stork et al., 1995) suggests that validation of the MUSE/R method requires three stages:

- Evaluation of the in-car information system designed by applying the method. This stage ensures that the method delivers a design that satisfies the user requirements.
- Assessment of the correct application of the MUSE/R method. This stage ensures that the method was applied;
- Initial validation of the PCMT model (see below). This stage ensures that the method also supports a research aim.

The accomplishment of these stages is described later in this paper, after the description of the design.

PLANNING AND CONTROL OF MULTIPLE TASKS MODEL

Hill et al. (1995) developed a model for representing how users, with their devices, including computers, plan and control the tasks that they need to perform, particularly when they have more than one task at a time, i.e. multiple tasks.

The model was grounded in scientific models of planning and was then developed by analysing how different groups of people planned and controlled multiple task work. The different groups included secretarial administration, medical reception work, and legal service provision. They employed the concepts from CE research, so MUSE/R was the method of choice for applying the PCMT model to a design, since it also employed the CE concepts (unlike MUSE).

The interactive worksystem is identified as having two cognitive structures, plans and knowledge-of-tasks (KoT), and four processes supporting behaviours of planning, perception, execution, and control, the first three of which can be temporally interleaved, i.e. 'having no necessarily fixed order in which to be performed'.

The plans are defined as 'specifications of required transformations of domain objects and/or of required behaviours'. Planning behaviours specify these transformations and behaviours.

The KoT is a representation of information about domain objects. Perception behaviours acquire information about these domain objects.

Control behaviours 'entail deciding which behaviour to carry out next, but may involve more than reading off the next behaviour from a complete and fully-elaborated plan'. Execution behaviours transform the attributes of the domain objects.

Figure 2 shows a schematic diagram of the PCMT model. The PCMT model has not been used for design until this project, so were were interested in applying the PCMT model to provide an initial validation for it, and to support the initial validation of the MUSE/R method.

Initial Validation of the PCMT Model

Following Stork et al. (1995) again, the following stages for validating the PCMT model are suggested:

- Evaluation of the in-car information service. This stage ensures that the model delivers a design that satisfies the user requirements.

- Assessment of the correct application of the PCMT model. This stage ensures that the model was applied.

The first of these stages is the same as the first stage
in the initial validation of MUSE/R. These two stages together are the third stage in the initial validation of MUSE/R. The accomplishment of the second stage is described later in this paper, after the description of the design.

DESIGN OF IN-CAR INFORMATION SYSTEM

Requirements

The requirements for designing the in-car information system were two-fold:

- Car commuters could gain the business-relevant information that they would normally gain from a newspaper, while driving to work;
- safety was maintained while driving and operating the in-car information device.

Extant Systems Analysis

The first phase of the method involved analysing existing systems that would be relevant to the design. Although seven alternative systems were considered, only three systems were selected for analysis:

- A DAB receiver interface;
- a train commuter who read the newspaper for his business, during his train journey to work;
- a car-driver listening to and operating a state-of-the-art car radio, during his commute to work.

The first system ensured potential DAB features were considered for the design. The second made certain that reading style and content were considered for the design. The third system identified which features of current radio systems could be considered safe or not.

The second and third systems were modelled using the PCMT model. They both involved planning and control: the second with only one task, and the third with two tasks, driving and using the car radio. The multiple task capability of the model enabled both tasks of the third to be modelled, and the second to be combined in the following design phases.

Figure 3: Domain Model for Newspaper Extant System

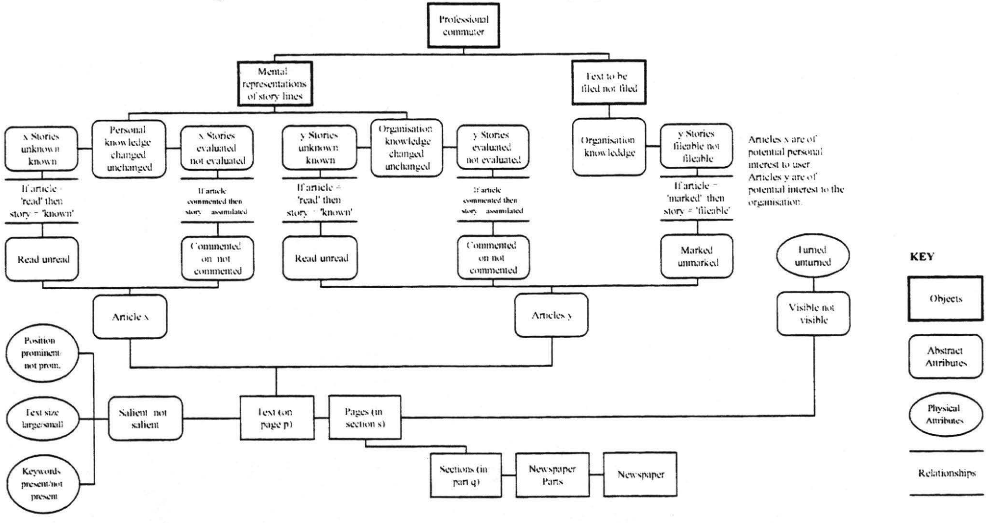
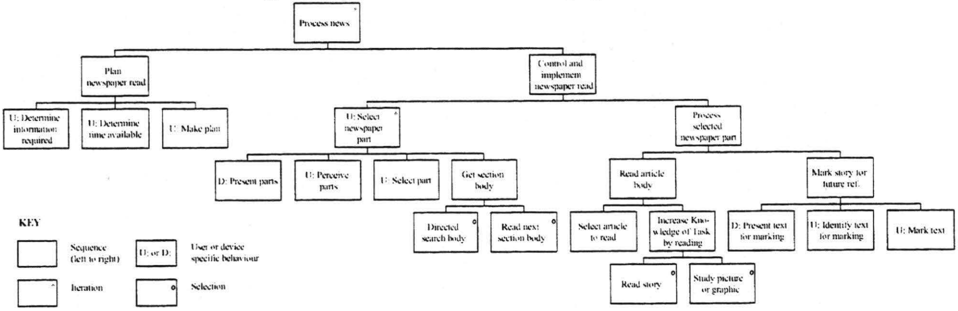


Figure 4: Behaviours of Newspaper Reader



For each of these systems, a model of the domain, structures, behaviours, and performance was generated from observations of individuals carrying out appropriate scenarios. Alongside these models, we identified possible solutions to poor performance.

For example, Figure 3 shows the domain model from the newspaper study and Figure 4 shows the generalised behaviours of the newspaper reader. The attributes of the domain are transformed by the behaviours 'Read Story' and 'Study Picture or Graphic'. However, the effectiveness of these transformations depends on whether the article is of interest to the user or to their organisation, a distinction captured by the domain model.

The domain instantiates the PCMT model's domain, in that it has abstract objects (for example, 'mental representations of story lines') and physical objects (for example, 'text' and 'pages'), each with attributes (for example, 'read/unread' and 'position') that are transformed. Similarly, the structures and behaviours instantiate those of the PCMT model (for example, 'Plan newspaper read', 'U: Perceive parts' and 'Control and implement newspaper read').

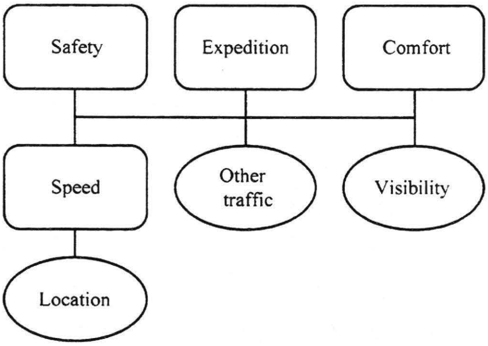
An example of a performance issue that was identified showed that the reader's costs (workload) were minimised by having a consistent ordering to the paper, which was modelled in the reader's

Knowledge of Tasks structure. The suggested possible solution was to maintain a consistent ordering of the paper.

Conceptual Design

The second phase of the method developed a model, in the form and at the level of a PCMT model, of how the in-car information service would deliver the information and safely. This model relied heavily on the previous models and the potential solutions, identified during the first phase.

Figure 5: Additional Domain Model Objects and Attributes for the In-Car Information Service

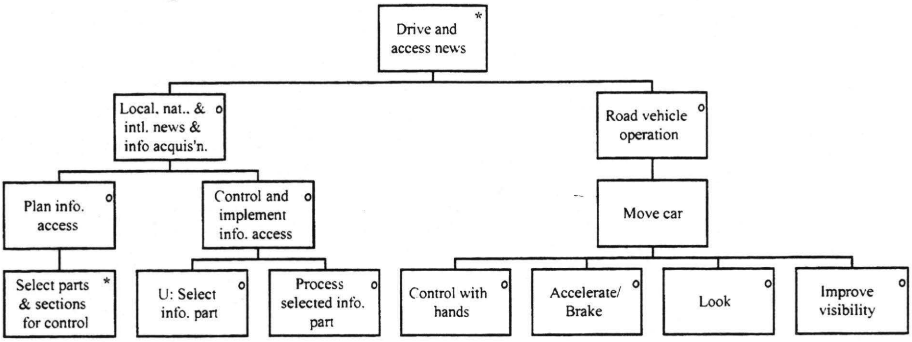


Issues such as the ordering of the information were finalised in this phase. A particular concern was the requirement for a driver to concentrate on driving and not on the information at certain times. The extant system model of the car radio suggested that the system should allow the driver to repeat sections of the information easily. A prototype assessment of this repetition confirmed its inclusion in the design.

For example, the proposed domain of the in-car information service brought together the domain model of the newspaper extant system and that of the driving radio extant system. The result is similar to that in Figure 3 with the additions of the objects and attributes in Figure 5. These domain objects and attributes ensure that the transformations to achieve good driving can be expressed.

Figure 6 shows the proposed general behaviours for the in-car information service, with planning behaviours that matched those of the PCMT model. As is clear in the figure, some of these behaviours are ported from the newspaper extant system study and others from the driver extant system study. However, the accompanying table starts to identify how the two types will fit together. For example, the design, at this level, included a menu-like display of the newspaper section titles, selection of which would give a choice of article titles and graphics, and further selection of these would give the contents of the article.

Figure 6: In-Car Information Service General Behaviours



Detailed Design

Once the PCMT-level model of the in-car information service appeared to offer a solution, the third phase of the method moved to the detailed design. The third phase considered issues such as:

- Textual display of the information (part of the original DAB concept), which was rejected as being unsafe for driving;
- voice-command of the interface, which was rejected as an unnecessary complication in favour of a joystick control, to keep the driver's hands close to the steering wheel;
- the replaying of information to give repetition as identified in the previous phase.

In the final design, the driver had four navigational buttons on a joystick control placed within reach of the steering wheel. The four buttons were in the shape of a cursor keypad, and signified up, down, forward, and back.

When turned on, a recorded voice started reading the section titles of the newspaper. A press of the down button gave the article and graphic titles within that section. A further press of the down button gave the content of the article or a description of the graphic. The titles, content, and descriptions were modified to be more suitable for reading to a driver. For example, the article would become more detailed as the reading continued, allowing the driver to hear an overview of the content without having to hear the rest of the article, like 'skimming'. A driver wishing to hear the detail could skip the overview.

The up button moved back to the previous level, for example, from article content it would move to article titles. The forward and back button moved forward and back in the current reading. A quick press would jump the amount of the current item, or a ten second jump if within the longer readings, like the article content. The back button thus ensured that the driver could be distracted and quickly press the back button to recover their position in the reading. A longer jump, to the end of the titles, for example, would occur for a longer press of the forward and back buttons.

INITIAL VALIDATION OF THE METHOD AND MODEL

The stages identified above to initially validate the method and model were conducted by an independent assessor (Malhi, 1999). He had not previously worked on the research or on the design of the in-car information system. The following sections describe the outcome of each stage.

Evaluation of the In-Car Information Service

The user, who was analysed reading the newspaper during the design, was evaluated during a commute to work using the prototype of the in-car information service. He managed to 'read' almost the same amount of business-relevant information as he had using the newspaper[1].

His driving safety was assessed using the UK Driver Standards Agency assessment criteria. He made minor and infrequent driving errors, except in one category, where several errors occurred. However, the assessor judged that the errors in this category might be due to the assessment of a prototype, rather than the use of a finished product.

Therefore, the evaluation of the in-car information service was positive, with two possible provisos identified by the assessor:

- One category of driving errors occurred;
- the assessment was not very broad, involving only a single user.

Assessment of MUSE/R Correct Application

The assessor found substantial evidence of correct MUSE/R application. For example, all of the MUSE/R products were present and current HCI knowledge had been applied.

However, the assessor raised the difficulty with fully assessing the development version of the method. For example, he claimed that the poor definition of the procedures prevented proper assessment of the procedures.

Assessment of PCMT Correct Application

The assessor found substantial evidence of correct PCMT application. For example, all of the PCMT concepts were present in the products of the method and planning, control, and multiple tasks had been considered during the design.

CONCLUSIONS

The two aims of the design were successful. First, a prototype in-car information system was developed, to enable car commuters to 'read' a newspaper while driving.

Second, initial validation of the PCMT model and the MUSE/R structured method was achieved. This initial validation was 'successful' (Stork et al., 1995), albeit with two possible provisos identified by the independent assessor for the evaluation.

The two possible provisos suggest that further initial validation could be conducted to:

Develop a more faithful prototype to re-assess the category of driving errors of the user involved in the evaluation;
evaluate a larger number of business users.

However, we consider that these possible provisos are onerous relative to the further validation of the PCMT model and the MUSE/R method. Therefore, we propose to develop the PCMT model and method by:

---

[1] 38% for the in-car information service against 39% for the newspaper. It should be noted that non-business relevant information was significantly lower with the in-car information service.

- Applying the PCMT model to other designs, and validating it for these designs where possible;
- developing the MUSE/R method further, particularly the procedures, so that it will better support both design and research.

ACKNOWLEDGEMENTS

This research was carried out for the European Commission ACTS MEMO programme.

REFERENCES

Dowell, J. and Long, J. (1998) Conception of the Cognitive Engineering Design Problem. Ergonomics, 41, 2.

Hill, B., Long J., Smith, W., and Whitefield, A. (1995) A Model of Medical Reception — the Planning and Control of Multiple Task Work. Applied Cognitive Psychology, 9.

Lim, K. and Long, J. (1994) The MUSE Method for Usability Engineering. Cambridge University Press, London.

Malhi, H. (1999) Assessing Research Applied to the Design of an In-Car Information Service. MSc Ergonomics, University of London.

Stork, A. and Long, J. (1997) Structured Methods for Human Factors Reseach and Development. Proceedings of HCI International '97, San Francis.

Stork, A., Middlemass, J., and Long, J. (1995) Applying a Structured Method for Usability Engineering to Domestic Energy Management User Requirements: A Successful Case-Study. People and Computers X, Proceedings of HCI '95.

Using Cognitive Dimensions to Analyse Graphical Notations

Julia Hill

Human-Interaction Group,
Department of Computer Science,
University of York
Heslington, York, YO10 5DD

Peter Wright

Human-Interaction Group,
Department of Computer Science,
University of York
Heslington, York, YO10 5DD
pcw@cs.york.ac.uk

ABSTRACT

Within many engineering fields there is a growing popularity for the use of graphical and pictorial notations. Traditional ways of assessing the usability of these notations come in the form of empirical investigation, such as experiments and case studies. Unfortunately these can be very time-consuming. It is proposed that an analytical method based on the cognitive dimensions approach can be used to provide an appraisal of notations, and support findings using more traditional methods. An illustration is given of adapting the cognitive dimensions for selecting graphical notations to aid air accident analysis. Finally, a discussion is given reflecting the findings in comparison to those achieved using more traditional methods.

Keywords

Cognitive dimensions, usability, graphical notations, air accident reports.

INTRODUCTION

In interface and other design disciplines there are many notations to choose from for the different stages of design. However, it can be difficult to know which are the most suitable for a given task. Assessing the usability of these notations is therefore important. The term usability in this paper is used to refer to a broad set of things with respect to notations being easy to use, appropriate and effective for the task at hand. The traditional approach to analyse the usability of notations has been to carry out experiments or empirical case studies to explore these issues. For example, Gilmore and Smith (1984) investigated the utility of flowcharts to aid computer program debugging. Cunniff and Taylor (1987) carried out an empirical study comparing novices' program comprehension using either textual or graphical representations. Such studies can be very time-consuming, especially if the usability, usefulness, and appropriateness of a wide range of notations are being explored. This has been the case with current research in the investigation of using graphical notations for modeling accident sequences.

In comparison to traditional empirical methods, Green (1989) introduced the cognitive dimensions approach, with a focus on programming notations and environments (see also Green and Petre

1996). It is described as a broad-brush evaluation technique appraising the usability of notations quickly and easily, exploring issues such as: Does the notation use secondary notation (e.g. layout, colour) to portray information and how well does it model the domain? The cognitive dimensions have also been adapted to investigate what properties of notations make software specification languages easy for novice users to understand (Britton and Jones 1999).

In this paper we explore using cognitive dimensions to aid our choice of notations to support accident analysis in the representation of causal sequences. It is a widely held view that textual descriptions in accident reports are inconsistent, difficult to use and understand (Johnson, McCarthy and Wright 1995; Gerdsmeier, Ladkin and Loer 1997). To aid the comprehension of accident descriptions, the use of notations to model graphically the sequence of events building up to an air accident has been investigated. For example, Hill and Wright (1999) used experiments to find out whether supplementary graphical notations aid the comprehension of air accident descriptions, although these were limited to exploring two notations, petri nets and why-because graphs. Hill and Wright (1997) also carried out empirical case studies researching the issues of using graphical notations (petri nets and why-because graphs) to produce representations of accident sequences. Authoring studies have also been carried out with BAE SYSTEMS investigating for different tasks, the usability of petri nets, why-because graphs, and other notations. As an alternative to these traditional techniques, we propose that the cognitive dimensions approach might allow us to explore the usability of these notations quickly and easily for modeling accident scenarios. These findings are compared with empirical findings in the section entitled, Discussion of Notations.

We start by providing some informal definitions of terms used to describe the problem space. In the next section, Cognitive Dimensions, an account of how the cognitive dimensions is used to explore the suitability of graphical notations for modeling accident scenarios is given. The section, Air Accident Reports describing the structure and content of air accident reports follows this. Fourth, Evaluating the Notations gives an analysis of two notations, petri-nets and why-because graphs. This is in conjunction with a summary of further evaluations of the notations, state diagrams, fault trees and event trees. Fifth, in Discussion of the Notations a reflection of the findings is given to those from the empirical research referred to above.

Some informal definitions

In order to use the cognitive dimensions to evaluate graphical notations for the modeling of accident scenario descriptions, a high level understanding of these relationships in a more generic sense is required. Figure 1 shows these relationships in terms of graphical notation, represented domain, and graphical model. It is these terms that are informally defined for use in this paper.

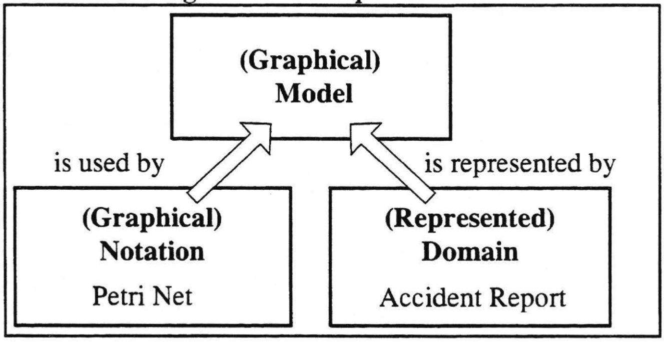
A graphical notation (also referred to as notation) is a language whose lexicon (symbols) and syntactic systems are constructed using pictorial representations. Natural language text is often used to describe instances of lexical use. For example, graphical petri nets, which describe the flow of information in a system, have annotated lexical items. A lexical item of a notation is a single lexical unit that comprises of one symbol and sometimes describing particular instances.

A represented domain (or domain for short) is something in the world, an object (e.g. aircraft) or occurrence (e.g. air accident) that is being represented in some way for the purpose of analysis.

A graphical model is a representation of the domain that uses a notation's pictorial lexical and syntactical system to convey only certain aspects to the reader. For example, a fault tree uses its leaves to show symbolically root causes of failure described in the accident scenario.

We can see therefore that the notation provides the structure and form of the model, while the domain provides the content. Another way to view this is that the notation provides the body in which to carry the character of the domain.

Figure 1: The Graphical Model



In this paper we show how the cognitive dimensions can be used to help choose suitable notations in the modeling of aspects from the domain, the accident scenario.

COGNITIVE DIMENSIONS

There are presently approximately twelve cognitive dimensions, for example visibility, and consistency2. These have been applied mainly to the evaluation of programming languages and their software environments (Green, Petre and Bellamy 1991; Modugno, Green and Myers. 1994; Green and Petre 1996). However, Modugno, Green et al.. (1994) point out that the cognitive dimensions could be applied in the evaluation of any cognitive artifact(s).

For the study on using graphical notations for accident modeling, a subset of these has proved to be useful. These are: role expressiveness, closeness of mapping, diffuseness/terseness, viscosity, and secondary structure. Each is described in turn in the context of our task.

Role expressiveness of lexicon and syntax

Green (Green and Petre 1996) used the dimension, role expressiveness, to evaluate whether the role of a component in a programming language was easily inferred asking, "What is this bit for?" The role expressiveness of a notation is enhanced by what are called meaningful identifiers, for example, using secondary structures such as typography and spacing.

In addition, role expressiveness is extended in order to explore whether the lexicon and syntax of a notation express their role clearly to a reader, for example. The number of lexical items can also be used to indicate how difficult a notation is to use. Britton and Jones (1999) used this lexical count to explore the idea that a notation with fewer symbols is easier to use than one with many they point out that this is not always the case.

Closeness of mapping

This dimension is typically used to help evaluate the mapping relationship between a program world (notation) and its problem world (represented domain) annotated in figure 1. The closer the mapping a programming language has to its problem world, the easier the task of programming is

---

2 . For a full account of the dimensions, see Green, T. R. G. and M. Petre (1996).

(Green and Petre 1996). This is especially the case with novice users because it is assumed they know the problem domain (Modugno, Green et al. 1994).

With respect to modeling accident descriptions, this dimension can be used to evaluate the closeness of mapping between the graphical notation and the accident domain. The closer the mapping, the more suitable it is (Zhang 1996). In order to evaluate closeness of mapping, an understanding of the contents of an accident report description is required (see section, Air accident reports), and knowledge of both a notation's lexicon and syntax. Hill and Wright (1997) showed that if they do not adequately reflect information structures within the text, mapping difficulties might be encountered.

A consequence of a 'non-close' mapping is that the user may have to force-fit aspects of the accident domain onto the notation's structure. This can lead to misrepresentations in the model (Hill and Wright 1997). Since aircraft are safety-critical systems, force-fitting needs to be minimised to help prevent incorrect interpretation of the accident. However, it is important to be aware that mapping errors can also occur by incorrect interpretation of the domain.

Conversely this does not mean that the model and the domain should be identical. Research shows that the use of graphical notations makes the relationships between events more explicit using symbolic graphical structures than when embedded within accident scenario texts (Johnson, McCarthy et al. 1995; Hill and Wright 1997).

Diffuseness/terseness

Similar to closeness of mapping, this dimension is used to refer to the succinctness of the notation: Does it use too many or too few lexical items for the task it is required for? A diffuse notation can be difficult to read due to the increased memory load of a user. Likewise a terse notation may be difficult to read because it does not model the domain in enough detail. Green and Petre (1996) point out the difficulty of evaluating this dimension independently of closeness of mapping. The mapping from a domain to a representation may be terse and correct because of the domain's simplicity. It also may be terse if the notation is unsuitable for modeling the domain, for the task required. Similarly, the mapping from a domain to representation may be diffuse because of the domain's complexity. It also may be diffuse if the notation is unsuitable for modeling the domain. For example, a terse model may be required to remove the complexity embedded in the domain.

A way we can use this dimension is to compare the same domain modeled using two or more different notations. One of the notations may use a small number of lexical items to model the same information compared to another notation using a large number. The first notation could be described as terse, while the second could be described as diffuse.

Viscosity

This dimension is used to evaluate how easy or difficult it is to make a single amendment to a model when using a particular notation. An example (Green and Petre 1996) is the task of making global changes to a program by hand if no global update tool is available in the software environment. This indicates a tool is required to support this task.

For the purpose of our discussion, a non-viscous notation is one where a user finds it easy to make a change to the model built using the notation. A viscous notation can make this task difficult (Hill and Wright 1997). For example, if a user revises an instance (from the domain) that is conveyed in the model, this may cause semantic and syntactic difficulties to both the surrounding areas, and to the model as a whole. It may cause semantic difficulties if the surrounding lexical items no longer semantically fit. It may cause syntactic difficulties if the syntactical structure in which these lexical items sit no longer is correct. Where there are difficulties in making revisions, it may be possible to design a tool to support these changes, thus, making the task of the user easier.

Secondary structure

Also known as secondary notation, this dimension refers to aspects of a notation that are used to convey extra information but are not part of the formal lexicon and syntax. This information can come in the form of spacing of lexicon (symbols) on the page, colour, typography, grouping of related items and so on. These provide intuitive and salient features of a graphical notation that readers may seek to exploit.

AIR ACCIDENT REPORTS

Knowledge of the structure, form and content of accident reports is required to use this adapted taxonomy in order to know and understand what is being represented using a selected notation. The reports used in this research are aircraft accident reports, civil and military. These reports are written in natural language text and vary between approximately fifteen to one hundred pages in length. The accident description itself is divided into sections. They consist of an account of the events building up to the accident (and after) followed by an analysis into what caused the accident, and concluding with a list of findings, causes and contributory factors to the accident and the subsequent safety recommendations. These descriptions tend to be complex and inter-related. For our case study we are interested in the modeling of temporal and causal event sequences building up to the accident event itself, and how the cognitive dimensions can help select appropriate graphical notations.

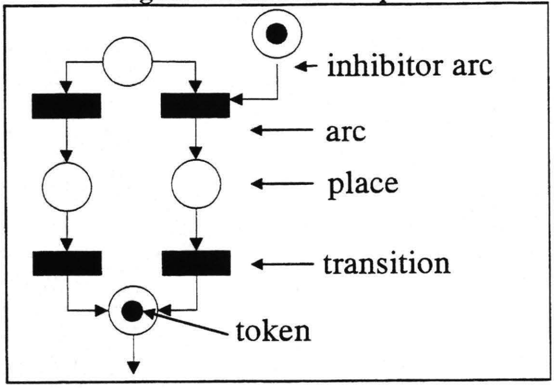
EVALUATING THE NOTATIONS

Five notations have been evaluated using the cognitive dimensions with respect to modeling causal sequences from accident scenarios. Only petri nets and why-because graphs are explored here for demonstration purposes. However, Appendix 1 provides a summary of all the evaluations carried out, including state diagrams, fault tree and event trees.

Petri nets

Petri nets (a directed graph) is a notation that models the flow of information in systems (Figure 2). It has five basic symbols: places, transitions, arcs, tokens and inhibitor arcs. Places are used to represent the state of a system; transitions represent events in a system and cause the state of a system to change. Tokens are used to show what places are holding true at anyone time, thus, representing the current state of a system. Arcs are used to link events to states, depicting the flow through the system from event to state. Finally inhibitor arcs prevent a transition from firing. The syntax (which also allows concurrent sequences) is very simple: state-event-state-event..

Figure 2: Petri Net Template



Using the dimension, role expressiveness of lexicon and syntax for guidance, having only five symbols suggests that it is quite simple for a user to learn, especially for the novice. We can also see in Figure 2 that the form of the symbols within the simple syntactical structure makes each type of symbol easy to identify. However, complex petri net models may be more difficult to read. This depends on the size of of the represented domain and the level of detail required. For examples see Johnson, McCarthy et al. (1995) and Hill and Wright (1997).

A comparison of the information structures within accident scenario texts to that of petri nets show that there is 'non-close' mapping between the two. Accident descriptions themselves are very complex, and this is mirrored in the richness of the textual description. On the other hand, the petri net language is very concise. Any description has to fit within its simple lexicon and syntax. This means the complexity of an accident description has to be fitted onto this structure (Hill and Wright 1997). On the other hand, petri nets can be used to show concurrency allowing a user to model multiple event sequences occurring at the same time in the accident scenario. Overall, however, a user may be made to force-fit the domain onto the petri net structure because of the mapping difficulties encountered.

Petri nets models tend to be viscous. Its syntax does not make it easy to make revisions (Hill and Wright 1997). If an amendment is made to either a place or transition, or another place or transition is inserted, this can affect the semantics of the representation. For example, if an amendment is made to a place, the surrounding lexical items in the sequence may no longer hold. However, from a design point of view, some kind of tool support may be provided making it easier for users making amendments. For example, when a user inserts a new transition, the tool automatically inserts the proceeding place. Petri net authoring tools are available which may be able to solve this usability issue (for example, ALPHA/Sim petri net tool (CPN group)).

Although not part of the formal definition, there are secondary structure enhancements that can be used. For example, in order to emphasise concurrency, spatial layout can be taken advantage of. Figure 2 shows how visible the relationship is between the two sequences: the nodes at the beginning and the end of both sequences are on the same horizontal axes. There are difficulties, however, with this layout. Novices of the notation may infer wrongly that there are temporal relationships between individual event nodes on the same horizontal axes.

Why-because graphs

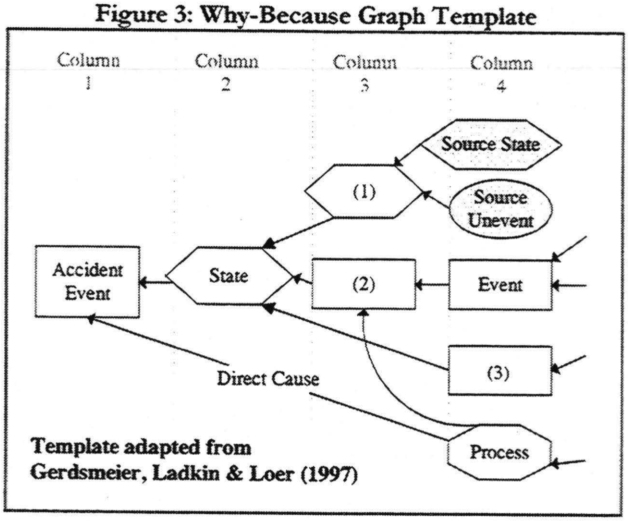
A directed graph designed to explicitly represent causal relations in accident scenarios (Figure 3), why-because graphs have nine symbols: states, events, processes, unevents, source states, source events, source processes and source unevents, and arcs. As the names suggest, states represent states, and events represent events in the accident scenario. Processes are a combination of both states and events, unevents are events that possibly should have happened during the flight, but did not. Source nodes are used to highlight possible root causes in an accident, and arcs represent causal flow. Syntax in comparison to petri nets is very flexible. There are no syntactical rules. Any lexical item can follow any lexical item providing the causal sequence is semantically correct.

Using the dimension, role expressiveness of lexicon and syntax, we can still argue that the language is very simple to use and learn: four of the lexical items are repeated as source nodes. Each symbols is also very easy to identify. However, Figure 3 shows that there are arcs crossing each other within the structure of the two-dimensional page. In a more complex model, it may be difficult for a reader to follow all these arcs intersecting each other. An example of a completed why-because graph is found in Gerdsmeier, Ladkin et al. (1997).

In comparison to petri nets, it can be argued that the mapping is closer between the accident scenario and why-because graphs: its lexicon appears more descriptive. The notation also recognises the difficulty in categorising accident sequences into only states and events; processes and unevents have been introduced. The flexible syntax could be said to reflect the complexity of accident scenarios. Suggesting that force-fitting should be less common.

Why-because graphs are a non-viscous notation. Due to the flexible syntax, making an amendment to a representation is easier syntactically. However, due to implications at the semantic level the surrounding lexical items still need to be resolved with respect to the causal sequence.

Figure 3: Why-Because Graph Template



This notation does use secondary structure to a limited degree. The graphical layout (Figure 3) is used to model visually the causal links between the lexical items, and the graphs can be designed to read from left to right, or right to left, for example. However, the relative position of lexemes does not carry any kind of information about the relationship between them. Only the connecting causal arcs show their relationship. For example event (2) in column two could quite as easily occur in the same temporal sequence as event (3) in column four. Furthermore, columns three and four show symbols on the same vertical axes. A novice could be misled inferring temporal relationship between these symbols because of their close proximal space.

There are two ways to get around this difficulty. The first is to make it very clear to users about the rules of spatial layout. Second, it may be possible to adapt the notation to place nodes in the same temporal alignment. In comparison to petri nets, it is the layout that may cause immediate difficulties with this notation, not the language.

Discussion of notations

Using the cognitive dimensions to evaluate petri nets, why-because graphs and the other notations (Appendix 1) has been successful in helping our task of choosing notations to model accident descriptions. First, we have seen that there are trade-offs to be made when choosing notations (Green and Petre 1996). On the one hand, we have seen that the lexicon and syntax of Petri nets has a poor mapping in relation to accident domains. On the other, we have seen that concurrent sequencing can be used to model concurrent event sequences found in accident domains. In comparison, why-because graphs have been shown to have potentially a descriptive lexicon and syntax reflecting the domain more successfully than petri nets. However, the graphical modeling of complex causal relations may cause reading difficulties due to limited use of secondary structure.

Appendix 1 summarises the advantages and disadvantages between the five notations evaluated. For example, comparing fault trees to petri nets, state diagrams and why-because graphs, we can see that the former has modeling difficulties in showing relationships between events explicitly. This is because the fault tree syntax is designed within the constraints of the tree structure. Consequently relationships between events on different branches cannot be shown explicitly3. There are also difficulties with event trees. They are designed to only explore a single sequence of events at a time using the binary to explore different outcomes with respect to an event holding true or false. As a result it cannot be used to show complex networks of temporal or causal sequences. In comparison, petri nets, state diagrams and why-because graphs, being directed graphs, can show event sequences explicitly, thus, indicating these are more appropriate than fault trees or event trees in the modeling of temporal or causal relationships.

The cognitive dimensions can be used to compare petri nets to state diagrams4. We can see (Appendix 1) that the latter does not show concurrent sequences, although they are very similar. Therefore, for our task of representing causal sequences in accident scenarios, petri nets are more suitable than state diagrams.

The evaluation in addition has shown where tool support could be added to aid a user. For example, we saw that tool support could be given to petri net users to aid amendments to models (viscosity). Explicit guidance should also be given to users on the spatial relationships between nodes for petri nets and why-because graphs. This could also be provided by a tool or by simple instructions.

There were nevertheless practical difficulties with using the cognitive dimensions as an evaluative tool for our case study. Cognitive dimensions are a broad-brush technique allowing notations to be evaluated quickly and easily. However, for our case study, we wanted to use the dimensions for a more fine-grained evaluation. This proved difficult without reference to our earlier empirical results. For example, this was found to be the case for the dimensions, closeness of mapping, viscosity and diffuseness, although the first two dimensions could be explored at a high level.

In order to evaluate closeness of mapping of petri nets to an accident domain at a more thorough level, it was required first to have knowledge of the domain, second, knowledge of petri nets,

---

[3] See Villemeur (1992) for an introduction to fault trees and event trees.

[4] See Harel (1986) for an introduction to state diagrams.

followed by an exercise mapping information from the domain onto the notation's lexicon and syntax. During these exercises, the evaluator also needs to be able to appreciate whether mapping difficulties are due to either the notation, or the represented domain (see Hill and Wright 1997). During this type of empirical exercise, the viscosity of the notation can also be explored more deeply.

Finally evaluating the diffuseness of a notation is not a dimension that can be investigated quickly and easily. This is because in order to find out how diffuse a notation is (petri nets, for example), a comparison of the represented domain in at least one other notation is required, for example why-because graphs. There is, otherwise, no way of knowing if a notation is diffuse or terse in its modeling of represented domains. This causes immediate difficulties because it is hard to model a domain at exactly the same level of detail using two different notations because of their different characteristics (Hill and Wright 1999). Also, this once again takes time. As a result of these practical difficulties, diffuseness has not been evaluated for each notation.

CONCLUSIONS

This paper set out to explore whether cognitive dimensions could take the place of time-consuming empirical comparisons. While we have found this not to be the case, they have value as an high level evaluative method. The dimensions themselves are a set of terms that can be used as a vehicle to explore a notation or a set of notations in a structured way. If certain attributes were required in a notation for specific task, the appropriate dimension(s) can be selected to help compare a selection of notations. For example, in modeling accident scenarios, a notation that can model networks of causal sequences is required. The dimensions, closeness of mapping in conjunction with role expressiveness, were used to help compare a number of notations in these areas. The evaluation showed that both petri nets and why-because graphs had suitable attributes, although if they were to be used it is necessary to be aware of the trade-offs made.

However, in order to have a deeper understanding of these issues, it was found that empirical studies still provided some of the answers, for example with closeness of mapping, viscosity and diffuseness. Furthermore, the dimensions can only be used to indicate but not make certain whether or not users' comprehension of accident scenarios is aided using graphical notations. This is an area where experimental work is still required.

A way forward then is to use cognitive dimensions as a high level means of helping to select notations for the task required. If the analysis then requires a fine-grained analysis, traditional empirical methods could then be applied as the next step.

Another way to use the dimensions is as a method to identify usability requirements for tool support of a notation. For example, tool support could help ameliorate the problems with viscosity.

Acknowledgements

The first author of this work is supported by a UK EPSRC research studentship funded through BAE SYSTEMS Ltd. The second author's contribution is supported by the ESPRIT Mefisto project (ESPRIT LTR 24963).

References

Britton, C. and Jones, S. (1999). The untrained eye: how languages for software specification support understanding in untrained users. J. of Visual Languages.

CPN group, C. S. D., University of Aarhus Petri Nets Tool Database, http://www.daimi.aau.dk/PetriNets/tools/quick.html.

Cunniff, N. and Taylor, R. P. (1987). Graphical versus textual representation: an empirical study of novices' program comprehension. Empirical Studies of Programmers: Second Workshop, Ablex.

Gerdsmeier, T., Ladkin, P., and Loer, K. (1997). Formalising failure analysis. Bielefeld, Faculty of technology, Bielefeld University. http://www.rvs.uni-bielefeld.de/research/WBA

Gilmore, D. J. and Smith, H. T. (1984). An investigation of the utility of flowcharts curing computer program debugging. Int. J. Man-Machine Studies 20: 331-372.

Green, T. R. G. (1989). Cognitive dimensions of notations. People and Computers V, Proceedings of the Fifth Conference of the British Computer Society, Human-Computer Interaction Specialist Group, University of Nottingham 508 September 1989, Cambridge University Press.

Green, T. R. G. and M. Petre (1996). Usability analysis of visual programming environments; a 'cognitive dimensions' framework. J. of Visual Languages and Computing 7: 131-174.

Green, T. R. G., Petre, M. and Bellamy, R. K. E. (1991). Comprehensibility of visual and textual programs: A test of Superlativism against the 'match-mismatch' conjecture. Empirical studies of programmers.

Harel, D. (1986). Statecharts: A visual formalism for compelx systems. Israel, Department of Applied Mathematics, The Weizmann Institute of Science.

Hill, J. and Wright, P. C. (1999). Graphically representing causal sequences in accident scenarios: just some of the issues. Visual representations and interpretations, Liverpool, Springer.

Hill, J. C. and Wright, P. C. (1997). From text to petri nets: the difficulties of describing accident scenarios formally. Design, Specification and Verification of Interactive Systems' 97, Granada, Spain, Springer-Verlag.

Johnson, C. W., McCarthy J. C, and Wright, P. C. (1995). Using a formal language to support natural language in accident report. Ergonomics 36(6).

Modugno, F., Green, T. R. G. and Myers, B. A. (1994). Visual programming in a visual domain: a case study of cognitive dimensions. People and Computers IX, Proceedings of HCI'94, Glasgow.

Villemeur, A. (1992). Reliability, Availability, Maintainability and Safety Assessment, John Wiley and Sons Ltd.

Zhang, J. (1996). A representational analysis of relational information displays. International Journal of Human-Computer Studies 45.

APPENDIX 1: USING THE COGNITIVE DIMENSIONS TO EVALUATE GRAPHICAL NOTATIONS FOR ACCIDENT ANALYSIS – A SUMMARY

	Role expressiveness of lexeme & syntax	**Diffuse-ness**	**Closeness of mapping & error proneness**	**Secondary notation**	**Viscosity**	**Overall opinion**
Petri nets (directed graph)	5 symbols Rigid syntax but shows sequential and concurrent sequences. Easy to read.	Quite diffuse	Not very close but shows concurrency providing a closer mapping. A high chance of force-fitting.	Can use 2-D space to emphasise concurrency. Relationships between symbols could be wrongly inferred.	Viscous	Has a very strict syntax with only a few lexical items, but shows concurrency. 2-D space can be used to emphasise relationships.
State diagrams (directed graph)	3 symbols Rigid syntax and single sequences of events only. Easy to read.	-	Not very close but could be adapted to show concurrency. A high chance of force-fitting.	Can use 2-D space to emphasise concurrency.	Viscous	Similar to petri nets with strict syntax and limited lexemes, but concurrency is not shown. 2-D space can be used to emphasise relationships.
Why-because graphs (directed graph)	10 symbols Flexible syntax, descriptive language. Symbols easy to read but not the network of sequences.	Quite diffuse	Quite close: the complex network of relationships can be shown. A moderate chance of force-fitting.	Does not use 2-D space to emphasise relationships. Relationships between symbols could be wrong inferred.	Not viscous	An expressive language with flexible syntax. Good for temporal and causal sequencing 2-D space is not explicitly used to show relationships.
Fault trees (tree structure)	9 symbols Rigid syntax with a binary tree structure. Tree structure does not clearly show relationships between events. Symbols are distinguishable.	Very diffuse	Not close: tree structure.	2-D space already defined using tree structure.	Viscous	The binary tree structure is not reflective of accident event sequences.

	Role expressiveness of lexeme & syntax	**Diffuse-ness**	**Closeness of mapping & error proneness**	**Secondary notation**	**Viscosity**	**Overall opinion**
Event trees (event sequence and binary tree structure)	Natural language descriptions with binary tree showing possible paths through a single sequence. Natural language is easy to read with binary tree.	Very diffuse	Not very: only single sequences of events are represented.	2-D space already defined describing single sequence of events with alternative paths using a binary tree.	Viscous	Event sequences in natural language A binary tree follows possible paths of the events when true/false. Could use for generating other scenarios for hazard identification.

This page is left intentionally blank

Session #2: Collective Activities

Human Factors and Professional Competence in Humanitarian De-mining

M. Bacchi, P. C. Cacciabue

European Commission
Joint Research Centre
Ispra (VA), Italy.
monica.bacchi@jrc.it, pietro.cacciabue@jrc.it

M. Bellezza, A. Re

University of Turin
Department of Psychology
Turin, Italy.
miriam.bellezza@jrc.it, are.@cisi.unito.it

ABSTRACT

This paper presents an overview of the research conducted at the Joint Research Centre (JRC) of the European Commission (EC) and the University of Turin in the field of professional competence of de-miners expert in manual clearance. It discusses the methodology used, difficulties encountered and results obtained.

Keywords

Human Factors, Humanitarian De-mining, professional competence, knowledge elicitation.

INTRODUCTION

Despite the amount of research conducted in order to identify 'new' technologies to support Clearance Operations in Humanitarian De-mining (King, 1998), little has changed in the field. Currently, world-wide operations of detection and clearance are very often carried out manually, with the support of metal detectors and probes (United Nations).

One of the crucial phases of manual clearance is "prodding" which consist of inserting a probe into the soil at a 30° angle, approximately every five centimetres. When a solid object is detected, first some more probing is conducted to get a feel for the shape and size of the object then excavation is carried out to uncover the object. This is a very slow and dangerous method of mine clearance and it clears one square meter of land in approximately 4 minutes.

The human being is the primary actor in field operations. However, Humanitarian De-mining has been ignored by the scientific Human Factors community even though relevant Human Factors (HF) issues are clearly pointed out in articles about technical subjects. Examples of such HF aspects, which have an impact on de-miners performance and clearance operations, are related to attention, risk perception, safety culture, mental and physical fatigue. In particular:

- *Routine and complacency* is one of the chief reasons for 'poor' or unsafe results as they decrease the level of attention of de-miners.
- *Risk perception* is decreased by routine as well as the high rate of false alarms: 100-1000 false alarms for each real mine detected (Bruschini, 1998).

- *National culture* determines the way in which Standard Operational Procedures and safety equipment are applied and used. For example, in some cultures, wearing protective equipment is seen as a sign of weakness (The University of Western Australia).
- *Fatigue* can strongly affect performance during prodding and de-miners can normally carry out this task for 20 minutes before requiring a rest.

During our preliminary study and review of the literature, it has become immediately clear that HF analysis and assessment are necessary for improving safety and efficiency of performance, in addition to the already existing efforts for ameliorating the technological means.

A fundamental standpoint of our research work has been that the current de-mining work is strongly focused on the active presence of humans in the field. This means that no automatic and distant control is yet being implemented, nor supervisory control means are prominently applied in the domain. Therefore, a human centred approach is needed to study de-mining activity, coupled with appropriate consideration of ergonomic analysis.

With the intention of addressing De-miners' Human Factors, and in order to identify performance indicators and training needs, a pilot research has started in 1999 and continues in 2000 to study de-miners professional competencies. This research aims at eliciting de-miners' competence, and is based on two main hypothesis:

- Competence is tailored to and embedded in the specific working context, that covers all aspects of physical and socio-technical dimensions, from the natural environment and geographical features (e.g. vegetation, rocks, small stones, sand, slopes, water etc.), to the organisation and cultural aspects.
- Competence is mainly exploited by the automatic level of behaviour (skill based level), which is perceived as 'obvious' by the people who practices it, therefore can not be elicited through usual techniques.

In the following we will firstly discuss objectives and hypotheses on which our research has been based. We then describe the selected methodological approach and the experts that have been elicited in this preliminary study. Finally, the results of this first phase of research are considered for discussing about future development on *non-technical training* of de-miners.

OBJECTIVES AND HYPOTHESES

Our research has been based on the hypothesis that the de-miners' expertise is made of a core competence that is acquired by training, application of standard procedures and a continuous information about different kinds of mines, new technologies, etc. We can define this as *restricted competence*, which may be further expanded and developed within the working environments in which it is applied. This generates a wider professional competence due to the personal initiative. Therefore it is less formalised and more prone to individual variation. This *expanded competence*, that generates a "contextually related skill", is adaptive to local rules, to implicit demands of the context. It allows a decision making related to the contextual conditions, to local signs and signals, habits and situations specific to the environment in which the activity is performed. Thanks to this *expanded competence*, the expert is able to anticipate critical or dangerous situations.

The de-miner's *expanded competence* has thus become the object of our research, and a number of considerations have been formulated about the subject.

- In a top down perspective, higher organisational factors (i.e. strategies, policies, and training programmes) produce operative instructions that lead the field operator. That means that the operative performance is the result (and a relevant indicator) of the organisation consistency and effectiveness. The health and the safety of the operator turn out both from his/her sensory-motor and cognitive skills and from relationships between operators and the different organisation levels, that must be considered when studying the de-miner's competence.

- The environment in which de-mining is carried out is constantly affected by communication and organisational failures among different systems and services. Small deviations or malfunctioning of these supporting media may lead to a discrepancy between expected and actual system performance and may engender extra "costs" of performance and eventually favour human inappropriate behaviour, i.e., human error.

In substance, the competence of de-miners can then be represented as a "Russian doll" (Oddone, Re, 1991): in the inside we can consider the sensory motor ability, generated by the knowledge and experience on different type of mines, their distribution in land of different nature (field, rocks, swamp, etc), the physical wear and stress affecting the mine (land shape, heat, climate etc). In the middle, we can consider the cognitive processes that govern the de-mining process, the relationship with the local population, with the team and operation unit. Finally, the outer layer contains the organisational culture and the knowledge of training and information system specific of the organisation.

All these elements contribute the *expanded competence* that build on a specific psychological space, very different from everyday life, and not-easily fitted into a stereotype representation.

METHOD

In order to identify the most appropriate method for our research, we adopted the criterion according to which "the choice of method for data collection and analysis should be determined by what 'work' the data need to do for us, that is to say, what aspect of the real world they need to illuminate for us" (Jordan, 1996).

The variety of competencies of interest and the holistic approach taken, favour either the selection a method or a mixture of methods that enable to pinpoint the multidimensional aspect of the hidden part of professional competence.

Different field methods were taken into consideration but were rejected, due to the difficulties for their application. For instance, in-situ question asking or 'think aloud' techniques were considered too invasive for such risky activity as Clearance Operations. On the other hand, other field methods like observation and video-based analysis were first planned and than discarded, at least for this pilot research, because of restrictions dictated by the SOP-Standard Safety Procedure, and in particular by the safety distance from the operation.

Several other techniques were rejected because task oriented and not in accordance with our human-centred and ergonomic approach. For instance, the classical approaches to "task analysis", mainly focused on the operational level which belong to that side of the competence that we define as *restricted competence.*

Finally, among the methodologies which require verbal interaction, "Instructions for One's Double" (Re, 1990), from the 'Knowledge Elicitation' field, was selected. "Instructions for One's Double", developed in the frame of the Ethnographic Approach, is directed to a broad spectrum of competence's components: contextual, social, organisational and cultural (Oddone, 1984).

This methodology, based on verbal interactions, was first developed and tested in Italy in the '70 and then successfully applied in different working environments, such as industries and schools, to analyse operators' and teachers' professional competence respectively.

From the Ethnographic point of view, this method supports the development of a common language between the expert and the researcher, and without the use of any a-priori category of analysis it leads to the identification of the system culture.

From the Knowledge Elicitation side, it guides the interviewee towards the description of what he/she does, how and in which context, singling out what is usually not discussed at all because considered "obvious" and therefore not relevant.

"Instructions for One's Double" guides the expert through a cognitive process which is characterised by 5 subsequent levels of competence' awareness: from the verbalisation of a list of acts and behaviours that the expert carries out during his/her job (the 'script'), to the awareness of how to use his/her competence (now conscious) for learning and teaching others how to learn. As the title suggest, "Instructions for One's Double", the verbal interaction with the expert leads to 'instructions' which allow the researcher to reproduce the expert's behaviours through a simulation.

SUBJECTS

With the aim to start a process of analysis and understanding of the dimension of professional competence of de-miners, we oriented our research to individuals definable as experts. In order to find the right subjects to interview and collaborate with, we thought that the Italian Army was a valuable environment to investigate. This organisation has a unit specialised in recruitment, training and working activities regarding mines and unexploded devices threat.

Unfortunately finding helpful and available subjects for our purpose was one of the most difficult actions to achieve for several reasons. In the first place, we were confronted with the typical reticence that corporate organisations are usually showing when studied or being scrutinised by observers, perceived as external outsiders. In the second place, once the right way to proceed and interact seemed to be found, our major difficulty was identification and availability of de-miners, as they were usually engaged abroad in international peace-keeping missions.

We carried out interviews with three different subjects. The first interview focused on a subject with a non-military working experience, and was mainly aimed to check whether the method was appropriate for this research. Two other interviews with real de-miner experts followed and the method, already experienced, was soon applied and improved. They were ideal subjects for this analysis on competence, as they both acquired many years of working experience in different countries around the world. Moreover, both experts were supervisors of teams working in minefield clearance operations. Obviously, a preliminary work was done in order to agree on interview guiding principles, and to achieve an atmosphere of mutual confidence. For example, anonymity and confidentiality were granted.

RESULTS

We adopted a qualitative approach for the treatment of the verbal material collected by the interviews. In particular, by employing qualitative analysis an attempt was made to capture the richness of themes emerging from respondents' talk (Smith et al., 1995).

The result of the data analysis enabled to spot several points of consideration related to the professional competence. Following the hypotheses settled for this research, it has been possible to recognise different clues and components of de-miner's professional competence, belonging to the so-called *restricted competence*, common to all individuals doing the same profession, and to the *expanded competence*, which is different for each practitioner because it derives from experience in different socio-technical contexts.

First of all, according to the definition of *restricted competence*, we identified elements of competence common to the standard knowledge of procedures established for all the Italian de-miners, and to the basic sensory-motor skills for clearance activities (e.g. visual and tactile exploration of the ground during clearance operations).

On the other side we discovered elements concerning the *expanded professional competence*, which are connected with three main areas:

- The technical competence necessary for the minefield clearance operations. This appears very complex and articulate because it depends on a combination of tangled factors, e.g., the typology of blast mines and tripwires, the position of the device in the ground, the type of soil where the mine is buried and the environment surrounding the operations. As an example, a different competence is necessary to work in Kosovo, where devices are traps made with transparent wires and located inside houses, than in Mozambique where mines are made in iron and buried around villages.
- The competence that enables communication, agreement and comprehension within all the different social groups present in the country. This covers civil population, local army, foreign military groups, civil organisations of humanitarian de-mining etc., and is necessary for the efficient and safe performance of operations. In practice, this competence relates to all aspects of *inter-organisations* interactions. As an example, one expert said that is necessary to be accepted by the local population, while another subject warned about the possibility that the local population may hinder operations for grudges originated after the war.
- The competence that allows the co-ordination, integration and harmonisation of different sectors within the same organisation. This competence, thus, covers all aspects of *intra-organisational* factors. As an example, one subjects mentioned the difficulty in organising the activity a consequence of the complexity in bureaucracy within the same organisation.
- The competence for the management of the teamwork and group dynamics, based, for example, on the capability of establishing interpersonal relationships and serene climate, as well as on the capability of carrying out an adequate co-ordination of tasks, which are all directed to enhancing safety and team performance. Both subjects focused their attention on the centrality of the relationships within teamwork members

From data analysis we pointed out that the elements of *expanded professional competence* are related with the personal working experience of each expert. This reflects a particular structure of competence that shows different features, depending on the variety of working contexts. Furthermore, from the reports obtained from both experts, the *expanded competence* seemed fundamental to train the beginners on the job, and to support other experts with different working experience in different and complex environments.

DISCUSSION

In the first phase of the research, our work was mainly focused on the development of the objective/hypothesis and on the identification of the methodology for data collection and analysis. Once we had selected "Instructions for One's Double", preliminary interviews were carried out, in a different domain than de-mining, in order to test and practice the specific technique. Finally, the data collection process started, and the first few interviews were carried out.

Even if the research was at its preliminary steps, we could highlight some relevant areas of professional competence. In particular, the subjects focused their attention on the importance of interpersonal relation with the team members and the team leader, with the civilians and the local institutions. Moreover, the role of inter- and intra-organisational issues were highlighted as important complementary factors for the development of the *expanded competence.*

In the next phases, our research will dwell into the identification and analysis of the different components of professional competence in Humanitarian De-mining. The study will be characterised by the combined use and confrontation of the "Instructions for One's Double" with another technique, i.e., the "Behavioural Event Interview" by McClelland (Spencer, Spencer, 1993). These two methods can be easily integrated because they show synergetic elements: both can be applied at all different levels of an organisation, from front-line operators to high decision-makers. Moreover, both the methods increase the expert's awareness about his/her own competence and ameliorate his/her ability to transfer expertise during training or every-day working situation.

The final results of our research will support the development of guidelines for the design of Human Factors Training for Clearance Teams.

According to preliminary results obtained so far, objectives and contents of those training courses could focus on:

- *Teamwork*, highlighting the role of good communication among team members,
- The features of successful *leadership*,
- As well as how to approach different *national culture* and establish a *positive relation* with the local population.

From the organisational point of view, the training courses could deal with *co-ordination* and *co-operation* within the organisation and among the different institutions dealing with de-mining in the same area.

Finally, the results could be used to improve the central information system, and facilitate the transfer to any single operator of competencies and knowledge, which are currently distributed and hidden in different parts of the system, by means of a specific knowledge management process.

In this sense, this first research has highlighted the gap between the information officially transferred by the organisation and the knowledge that de-miners informally acquire interacting with the real system.

REFERENCES

Bruschini C. (1998), "A Survey of Research on Sensor Technology for Landmine Detection". *The Journal of Humanitarian Demining*, Issue 2.1.

King K. (1998), "Demining: Enhancing the Process". *The Journal of Humanitarian Demining*, Issue 2.2.

Jordan, B. (1996) Ethnographic Workplace Studies and CSCW. In Shapiro, D., Tauber, M., Traunmüller, R. The Design of Computer Supported Cooperative Work and Groupware Systems. Human Factors in Information Technology 12. Elsevier, North Holland, pp. 17-43.

Oddone I. (1984), "La Compétence Professionelle Èlargie". *Socieété Française*, N 10, pp. 28-33.

Oddone, I., Re, A. (1991). Competenza professionale ristretta e competenza professionale allargata. In *Soggetti lavoro professioni*. Quaglino, G.P. (Ed.), Bollati Boringhieri, Torino, pp. 260-274.

Re A. (1990), "Psicologia e Soggetto Esperto". Tirrenia Stampatori, Torino.

Smith, J.A., Harrè, R., Van Langenhove, L. (1995). Rethinking methods in psychology. Sage, London.

Spencer, L. M., Spencer, S. M. (1993) Competence at Work: Models for Superior Performance, John Wiley.

The University of Western Australia, "An ergonomic aspect of humanitarian demining": www.mech.uwa.edu.au/jpt/demining/tech/dino/ergonomics.html

United Nations, United Nations demining database, www.un.org/Depts/Landmine/open.htm

A Method for Analysing Collective Design Processes

Françoise Darses

Laboratoire d'Ergonomie
Conservatoire National des Arts et Metiers
41 rue Gay-Lussac
75005 Paris, France
darses@cnam.fr

Françoise Détienne

EIFFEL (Cognition and Cooperation in Design)
INRIA (Institut National de Recherche en Informatique et en Automatique)
Rocquencourt B.P.105
78153 Le Chesnay Cedex, France
Françoise.Detienne@inria.fr
http://www.inria.fr/Equipes/EIFFEL-fra.html

Pierre Falzon

Laboratoire d'Ergonomie
Conservatoire National des Arts et Metiers
41 rue Gay-Lussac
75005 Paris, France
falzon@cnam.fr

Willemien Visser

EIFFEL (Cognition and Cooperation in Design)
INRIA (Institut National de Recherche en Informatique et en Automatique)
Rocquencourt B.P.105
78153 Le Chesnay Cedex, France
Willemien.Visser@inria.fr
http://www.inria.fr/Equipes/EIFFEL-fra.html

ABSTRACT

This paper focuses on modelling collective design, especially co-operative activities in co-design situations. Cognitive psychologists have proposed various methods for the analysis of verbal individual protocols, but much less for dialogues in collective work settings. However, many professional activities are carried out by people working together through verbal interactions. A method is proposed for analysing activities of designers during the meetings in which they are working together on a design project (i.e. their dialogues, and their generation and use of external representations). This research has a triple objective: modelling collective-design activities (especially the reasoning component), proposing a data-analysis methodology for cognitive psychologists and other researchers interested by collective design, and extending verbal-protocol analysis to collective work situations. The method will be illustrated through data collected in studies on the co-operation in co-design situations in two different collective-design tasks, software design and design of local area networks.

Keywords

Design, collective-design activities, dialogue analysis, co-design, distributed design, verbal interactions, co-operation, software design, software review, technical-review meeting (TRM), local-area network design.

INTRODUCTION AND OBJECTIVES

This paper focuses on modelling collective design, especially co-operative activities in co-design situations. Cognitive psychologists have proposed various methods for the analysis of verbal individual protocols, but much less for dialogues in collective work settings. However, many professional activities are carried out by people working together through verbal interactions. From a cognitive-ergonomics point of view, we have developed some methodological guidelines related to the analysis of real-life task-oriented dialogues. These guidelines provide a frame for analysing design dialogues in real situations. These guidelines have been constructed and used in the context of two studies on the co-operative processes in co-design situations. Some results of these studies will illustrate them.

DESIGN PROBLEM SOLVING

Cognitive ergonomics does not identify design in relation to a social function or a status, but qualifies as design tasks certain professional activities in which a set of formal characteristics can be identified. Therefore, one can identify numerous professional domains that deal with design. It can be the design of material artefacts (e.g. mechanical engineering, electronics, and architecture) or the generation of symbolic or abstract devices (e.g. planning, computer programming or resource allocation).

The features of design tasks (which relate to the notion of environment-tasks proposed by Newell and Simon, 1972) are well known (Goel and Pirolli, 1989; Simon, 1981):

- Problems tend to be large and complex. They are generally not confined to local problems, and the variables and their interrelations are too numerous to be divided in independent sub-systems;
- one consequence of this complexity is that the resolution of these problems often requires that multiple competencies be put together, which leads to development of collaboration within a single work group;
- there are many degrees of freedom in the problem initial state. This has led to consider design problems as ill-defined problems;
- a design problem has several acceptable solutions, not just one;
- there is not any pre-determined way that leads to the solution. A certain number of useful procedures and design methodologies are known, one can refer to similar projects already studied or to existing prototypes, but, each time, it is necessary to reinvent the steps that separate specifications and production;
- the definition of the problem and the elaboration for the solution are made in interaction. The problem does not exist before the solution - both are built simultaneously;
- the evaluation of the solutions is delayed as much as possible or in any case limited to the final solution. This is because the generation of all alternative design solutions is costly or impossible to carry out and also because objective metrics do not exist. Therefore, final solutions are satisfactory but not optimal.

One important objective of cognitive ergonomics has been to understand how these tasks characteristics determine the building of the problem-space explored by the designers. From this basis, design tasks have been mostly studied in their individual practices. However, the objective of cognitive ergonomics is to study design tasks in their individual and collective practices with a reciprocal effect: it is necessary to understand how collective design practices lead to a modification of the individual practices which are usually carried out.

CO-OPERATIVE PROCESSES IN CO-DESIGN SITUATIONS

Collective practices of design have been studied for a long time from multiple viewpoints: organisational analysis, social psycho-social and psychic analysis (focused on the role of factors such as the degree of trust in the others, the recognition of personal competence, personal development through work, power redistribution, necessity of protection, etc.).

The approach chosen by cognitive ergonomics differs from these types of analysis. While acknowledging that the dimensions described above obviously affect collective work and the development of a design activity, we have chosen to focus more on the cognitive aspects of collective design, and especially the cognitive aspects of the co-operative processes occurring during co-design.

We have proposed a first framework for analysing collective design tasks (Darses and Falzon, 1996; Darses Falzon and Béguin 1996; Visser, 1993). The actors involved in a design process are not all involved in the same way - some are involved in co-design activities while others participate in distributed design activities. These two situations can be found during one single design process and can also be taken in charge one after the other by one single actor. This distinction between co-design and distributed design is useful since each of these forms of collaboration during product development induces different collective processes.

Distributed design situations

In distributed design situations (Béguin, 1994; Karsenty, 1994), the actors of the design are simultaneously, but not together, involved on the same collective process. They carry out well-determined tasks, which have been allocated beforehand to them, and they pursue goals (or at least sub-goals) that are specific to them. They have as an objective to participate as efficiently as possible in the collective resolution of the problem. Distributed design is typical for concurrent engineering, in which the various sides of the production system must function in strong synergy during the product-development cycle.

Co-design situations

In co-design situations, design partners develop the solution together: they share an identical goal and contribute to reach it through their specific competencies; they do this with very strong constraints of direct co-operation in order to guarantee the success of the problem resolution. The competence of the partners can vary depending on the level of competence (e.g. interaction between designers of different seniority) or on the type of competence (e.g. interaction between drafters and engineers). Co-design has been studied by different authors (D'Astous et al., 1998; Karsenty, 1994; Malhotra et al., 1980; Visser, 1993). The two studies presented in this paper deal with co-design tasks.

A crucial mechanism in co-design is cognitive synchronisation. Cognitive synchronisation enables the partners to reach two objectives: (i) to make sure that each has knowledge of the facts relating to the state of the situation (i.e. problem data, solution states, accepted hypotheses, etc.); (ii) to make sure that they share a common knowledge regarding the domain (i.e. technical rules, objects in the domain and their features, resolution procedures, etc.).

Cognitive synchronisation therefore aims at establishing a context of mutual knowledge, at building a common operative system of reference (Karsenty and Falzon, 1992; Terssac and Chabaud, 1990). Cognitive synchronisation activities will vary depending on the amount of shared knowledge. This means in particular that the parity or non-parity of the dialogue (dialogues between pairs vs. expert/novice dialogues or dialogues between subjects with distinct knowledge) will have an effect on the necessity to communicate general knowledge. In dialogues between experienced partners Falzon, (1991) has shown in this domain, how the common-knowledge hypothesis enables an economy in communication by using operative languages, and how partners used repair dialogues when this hypothesis is at fault. These repair dialogues are aimed precisely at levelling general knowledge. The necessity to assure the nature of the common operative system of reference leads each partner in a dialogue to build a model of the other, as has been shown in various studies (Cahour and Falzon, 1991).

DESCRIPTION OF THE DESIGN SITUATIONS

Table 1 summarises the features of the two studies that were conducted. Several facts deserve to be stressed. Both studies deal with real, complex design tasks. The two studies have been conducted independently and at different periods (several years between the two). However, the authors knew and read each other's publications.

Table 1: Summary of the two studies conducted

Domain	**Computer network design**	**Software design**
Nature of the task	Solution production	Solution review
N° of designers	2	4
Domain of expertise	Same expertise: local area networks (LAN) configuration	Same expertise: software engineering
Level of expertise	Different levels: one experienced designer and one less competent designer	Same levels: all participants are competent designers
Situation	Semi-experimental field study: the situation of co-design was a usual practice. The less competent designer had to propose a solution to the network-configuration the problem that he was provided with. The instructions given to the expert demanded that he would collaborate with the less competent designer without taking charge of the problem. The two designers were not face to face but communicated by connected terminals. They shared a common representation of the network being designed, on which both could act.	Field study: the situation of co-design was technical-review meetings (TRMs) in an industrial software-development project. A TRM may occur after each phase in the global software-development process. It requires the presence of several reviewers. It has two main objectives: to verify the current state of the design project and to validate the specifications of the succeeding tasks. This is done on the basis of discussion of a document written in natural and/or programming language. Roles in these meetings are: manager of the meeting, manager of the project, "software-norms" expert/guarantor and presenter of the document (often author of the document under review).
Data collection	Computer-mediated dialogues: computer-mediated dialogues were the only communication mode allowed. Thus, data collected were written dialogues and graphical representations drawn and sent through the computer. The computer-mediated communication situation was used for reasons outside the scope of this text. It probably affected certain dialogue dimensions, such as verbosity or interruptions. However, we do not believe that it affected the dimensions studied in this paper, i.e. collective processes.	Oral dialogues video-recorded: the dialogues between the designers were their "normal" oral dialogues held in order to discuss the design document under review. Each dialogue was transcribed into a verbal protocol. The protocol was annotated with information concerning the nature of the (parts of) documents consulted and produced.

Domain	**Computer network design**	**Software design**
Bibliographic references	Darses, Falzon and Robert (1993); Falzon and Darses (1992), Darses and Falzon (1994)	D'Astous, Détienne, Robillard and Visser (1998); Détienne, F., Visser, W., D'Astous, P., and Robillard, P. N. (1999); Robillard, D'Astous, Détienne and Visser (1998)

METHODOLOGY

The purpose of this section is to give some methodological guidelines related to the analysis of real-life task-oriented dialogues. These guidelines provide a frame for analysing design dialogues in real situations. The frame is made of two distinct levels of coding: Basic level (which corresponds to Individual Units) and Composite level (which corresponds to Co-operation Moves).

The basic coding of the corpus: coding into Units

Identifying Turns

A dialogue is transcribed into a verbal protocol. Each protocol is cut up into a series of individual participant utterances ("turns"), according to change of locutor. Besides the verbal material, the gestural and physical movement aspects of the interaction have not been addressed in these studies.

Coding into Units

These utterances are cut up further into one or more individual "Units" according to a coding scheme developed on a predicate (argument) basis. A given predicate may admit a number of possible arguments (but not any argument). According to the form of the predicate (assertion or request), each unit is modulated (MOD). The default value of a move is assertive. Only if its predicate is a request, this modulation is coded explicitly. Thus, each unit was coded as MOD[ACT/OBJ] (see Table 2).

Table 2: Basic coding scheme

Predicate action implemented by a participant locutor (ACT), e.g.	*Argument* object affected by the action *(OBJ), e.g.*	*Modulation (MOD)*
generate (GEN)	problem data (DAT)	Request
criticise (crit)*	solution ele.s (SOL)	Assertion
inform (INFO)	domain obj.s (DOM)	
*Note: Criticisms may be positive (+), negative (-) or neutral (no sign)		

DIFFICULTIES WITH THE UNITS CODING SCHEME

How to proceed for defining the coding rules?

The first question is then to define predicates and arguments. In the two studies, this definition was accomplished through an iterative process: the analysts coded separately a same dialogue extract and then confronted their coding. It is to be noted that two questions are discussed at this point. First, the level of granularity of the coding scheme (what is the size of the smallest unit?); second, the nature of predicates and arguments. The analysts decide on some rules, choose a second dialogue extract and then iterate the process.

One point needs to be stressed here: it is extremely important that the analysts keep a precise trace of their own dialogue and decision processes. For instance: Why did they finally decide not to use this predicate but rather this one? What is exactly encompassed by this particular predicate? Ideally, the coding rules should be written down after these discussions. By coding rules, we mean not only the coding scheme, but also the rationale behind the scheme. As a matter of fact, we noted that analysts might very well converge on a coding scheme without being totally conscious of the rules they actually use. This phenomenon becomes obvious when asking a third person to code a dialogue.5 Writing down the rules has two benefits. First it helps the analysts in formalising the scheme; second it facilitates the transfer of the coding scheme towards a third analyst if needed.

How to deal with the argumentation level?

An important issue at this stage is also to determine a position relative to argumentation. Consider the following extract, taken from a "network design" dialogue. A designer has proposed to use the ceiling to install the cables. The other designer then says:

"I think we should stick to using the existing shafts. The ceiling can be used but it causes additional works and high costs of installation".

The second sentence has obviously a role of justification of the first one. Additional work and cost justify the proposition of using existing shafts. Thus one coding solution would be to create a Justify predicate with arguments like Procedural Knowledge and/or Design Criterion (for the additional work and cost). However, this means that coding is then not based on the intrinsic signification of the coded unit, but on its relation to a preceding one. As a matter of fact, the second sentence plays a justification role only in this context. In another context, it could be, for instance, a simple transfer of knowledge.

The alternatives are then either to code argumentative roles directly (i.e. at the first coding pass) or to code argumentation only at a later level of analysis. Different choices have been made in the two studies. Retrospectively, the authors believe that the second alternative should be preferred. Other problems are linked to argumentation, which will be developed further below.

How to solve the knowledge ambiguities?

As was mentioned, the dialogues are technical dialogues between professionals, who share much domain knowledge. This has methodological consequences. Coding cannot be achieved without help from a competent professional; interpretation requires domain knowledge. Consider for instance the following utterance (taken from the "network design" dialogues):
"The Suns have Ethernet cards"
This can be interpreted either as "all Sun computers are equipped with Ethernet cards" or as "these Suns are equipped with Ethernet cards". In the first case, it will be interpreted as transfer of a general piece of knowledge; in the second case, it will be interpreted as transfer of a piece of information about the particular situation.

Thus, in the two studies, the assistance of a domain expert was requested. In the "network design" study, the coding of the protocols was checked with a domain expert. In the "software review" study, the analysts elaborated the coding scheme, then explained it to a domain expert who did the actual coding of the dialogues.

---

[5] As a matter of fact, asking the intervention of a third person can be used as a heuristic to check that rules have been explicitly stated.

The Composite level: Co-operation moves

First phase: Grouping Units into sequences

At the composite level, frequent configurations of units are grouped into sequences corresponding to basic co-operative interactions. Such configurations can be formed in various ways, on a qualitative or on a quantitative, a concept- or a data-driven basis.

Psychological expertise may guide a qualitative approach in a concept-driven way. Cognitive models of design (Visser, 1991, 1992) suppose, e.g., that evaluation proceeds by criticism of a solution proposal, possibly preceded by a request for such a criticism and backed up by a justification. The criticism has to be accepted or rejected. This leads us to search in the design protocol for evaluation sequences composed of these different activities.

A data-driven quantitative approach may be used to identify significantly frequent configurations of units. Different types of method may be suitable for this aim. A statistical method, grounded in information theory, is Lag Sequential Analysis (LSA) (Allison and Liker, 1982). LSA enables the identification of units that follow each other, with or without other units in-between. The analysis consists in determining whether or not the frequency of a given unit is independent of the frequency of another one. Sequential structures enable the definition of configurations. Hierarchical clustering (Johnson, 1967) is another statistical method, which allows similarities inside sequential structures to be quantified.

On the one hand, statistical methods can be used to validate groupings that were identified, using "pure" qualitative analysis (as presented above); on the other hand, cognitive models may be of help in interpreting and conceptually validating the configurations identified by statistical approaches.

A combination of qualitative and quantitative approaches is to formulate, on the basis of psychologically inspired hypotheses, grammatical rewriting rules (Gonzalez and Thomason, 1978) of sequences of units, and then apply a statistical method on the result. Such a combined approach can be applied in an iterative way, in several cycles, as long as statistically significant configurations are detected.

Second phase: Formalising co-operation moves

In both studies, a similar scheme was adopted for formalising co-operation moves. A co-operation move is defined as a set of items, coded at a lower level, organised sequentially and with specific attributes.

Table 3 presents an example taken from the network design corpus. Optional units are put between brackets.

Table 3: The "negative evaluation" co-operation move

Designer 1	Solution generation [Request of criticism]
Designer 2	Negative evaluation Justification [Solution generation OR Goal generation]

It is to be noted that this particular co-operation move covers utterances of the two partners. This is not always the case. Sequences may consist of a succession of units emitted by a single partner.

DIFFICULTIES WITH THE CO-OPERATION MOVES CODING SCHEME

Argumentation

In the preceding example, "Justification" is not an item of the initial coding scheme. It stands for several kinds of informative utterances (on procedures or on domain-object characteristics, for instance). These utterances, in that specific context of negative evaluation move, play the argumentative role of justification. Thus, the creation of co-operation moves implies the identification of units that can play a similar role in a specific position within a move. In the example above, the "Justification" role can be played for instance by an Inform (Goal) message. Another example is the "Prevention" role that appears in the "Positive evaluation" co-operation move, which consists in providing, after an initial positive evaluation of the partner's proposal, an information that will help the application of the solution (OK, but pay attention to…). Again, this information can take several forms (Inform (Procedure), Inform (Domain Object Property), and Inform (Problem Data)).

Inferring missing units

Design dialogues are task-oriented. The participants in such dialogues know what they can expect from each other and what others expect from them. This common frame of reference allows them to interpret what is said and to behave in an appropriate way. For instance, if A proposes a solution, A expects the solution to be evaluated by the others, even if A does not explicitly request it. Similarly, if A proposes a solution, and if B then proposes an alternative solution, A may interpret B's proposal as conveying an implicit criticism (a negative evaluation of A's solution). Or, if A proposes a solution, and if B then proposes an expansion of this solution, A will infer that B implicitly supports A's proposal. Design dialogues thus follow principles of operational co-operation. It is important to understand these principles in order to identify correctly the co-operation moves, either to define as optional some units (for instance, evaluation requests), or to infer some missing units (for instance positive or negative evaluations).

Sequence reorganisation

Verbal expression has to be linear, whereas relations between the underlying elements are not necessarily. E.g. the expression of the arguments justifying a (negative) evaluation may precede or succeed the expression of the evaluation. JUST-EVAL and EVAL-JUST are two possible surface forms of the same underlying argumentative movement OPINION-ARGUMENTS.

CONCLUSION AND DISCUSSION

In the two studies presented above and used as illustrations for the method that we are developing for analysing collective design processes, different choices have been made at the outset. Nevertheless, comparable results have been obtained.

In software reviewing

In technical-review meetings (TRMs), we found that, when introduction of a solution is followed immediately by a development, this development consists in changing the form of the solution, and there is an implicit negative evaluation according to a criterion of form. Introduction of a solution can also be followed immediately by either its evaluation alone or its evaluation and development of an alternative solution (in one order or another). Such review activities may, or may not, be preceded by a cognitive-synchronisation exchange. If they are, the evaluation bears mainly upon content criteria.

With respect to the relationship between review and cognitive synchronisation: when review is introduced by cognitive synchronisation, this means that a shared representation of the to-be-evaluated object may be a prerequisite for its review to take place. The argumentative movement is of the type "proposition-opinion".

With respect to the relationship between review and design: the review of a solution, in particular a negative review, often leads participants to make explicit alternative solutions: such a solution may be a justification for the negative review or a solution for the current rejection. The argumentative movement is of the type "opinion-arguments". Note however, on the surface the arguments may be presented, either before the opinion it supports is presented, or afterwards.

We thus notice that the activities of elaboration and of cognitive synchronisation, even if not expected in the prescribed task, are both necessary and useful in the collaboration that takes place through argumentation in TRMs.

In network design

A key feature of the interaction in this design is the evaluative, critical activity, which represents a large part of the expert's contribution: the less experienced designer proposes an element of solution and the more experienced one reacts by evaluating the proposal. Positive evaluations may be followed by solution extensions (additions to the solution state) or by preventive statements (indications of not-to-be-forgotten facts or actions, implied by the solution proposal). Negative evaluations are always accompanied by justifications and often by alternative proposals —an observation also encountered in the reviews in our companion study. Moreover, the experienced designer also acts as a meta-planner, by proposing problems to be tackled or by suggesting to switch to a new design phase.

Thus, it can be seen that the interventions of the experienced designer are not limited to solution proposals (as most classical expert systems behave): evaluation and planning are essential features of the experienced designer's behaviour. When he makes solution proposals, these are often part of a wider assistance move (e.g. negative evaluation).

REFERENCES

Allison, P.D., Liker, J.K. (1982). Analysing Sequential Categorical Data on Dyadic Interaction: A comment on Gottman. Psychology Bulletin, 91, 393-403.

Béguin, P. (1994). Travailler avec la CAO en ingénierie industrielle: de l'individuel au collectif dans les activités avec instruments. Thèse de Doctorat d'Ergonomie, C.N.A.M., Paris.

Cahour, B. and Falzon, P. (1991). Assistance à l'opérateur et modélisation de sa compétence. Intellectica, 2, 12, 159-186.

Darses, F., Falzon, P. and Robert, J.M. (1993). Cooperating partners: investigating natural assistance. In G. Salvendy and M.J. Smith (Eds.), Human-computer interaction: Software and hardware interfaces. New York: Elsevier.

Darses, F. and Falzon, P. (1994). Activités de coopération au cours de la résolution de problèmes de conception. Workshop sur la Coopération à la Conférence Internationale sur les Systèmes Experts et leurs Applications, Paris, 30-31 May 1994.

Darses, F. and Falzon, P. (1996). La conception collective: une approche de l'ergonomie cognitive. In G. de Terssac et E. Friedberg (Eds.). Coopération et Conception. Octarès: Toulouse.

Darses, F., Falzon, P. and Béguin, P. (Eds.) (1996). Proceedings of COOP 96, Second International Conference on the Design of Co-operative Systems, Juan-les-Pins (France), 12-14 juin, Sophia-Antipolis: INRIA.

D'Astous, P., Détienne, F., Robillard, P. N., and Visser, W. (1998). Types of dialogs in evaluation meetings: an analysis of technical-review meetings in software development. Proceedings of

COOP'98, Third International Conference on the Design of Co-operative Systems, Cannes (France), 26-29 May, 1998.

Détienne, F., Visser, W., D'Astous, P., and Robillard, P. N. (1999). Two complementary approaches in the analysis of design team work: the functional and the interactional approach. Position Paper presented at CHI99 Basic Research Symposium, Pittsburgh, Pennsylvania, USA, May 15-16 1999.

Falzon, P. (1991). Co-operative dialogues. In J. Rasmussen, J. Leplat and B. Brehmer (Eds.) Distributed decision-making. Chichester: Wiley.

Falzon, P. and Darses, F. (1992). Les processus de coopération dans les dialogues d'assistance. Actes du XXVII° Congrès de la SELF. Lille, 23-25 septembre 1992.

Goel, V. and Pirolli, P. (1989). Motivating the notion of generic design within information-processing theory: the design problem space. AI Magazine, Spring, 18-36.

Gonzalez, R.C., Thomason, M.G. (1978). Syntactic Pattern Recognition: An Introduction. Addison-Wesley.

Johnson, S.C. (1967). Hierarchical Clustering Schemes. Psychometrika, 32, 241-254.

Karsenty, L. (1994). L'explication d'une solution dans des dialogues de conception. Thèse de Doctorat de Psychologie Cognitive Ergonomique, Université Paris 8.

Karsenty, L. and Falzon, P. (1992). Cadre général pour l'étude des dialogues orientés-tâche. Actes du XXVII° du Congrès de la SELF. Lille, 23-25 septembre.

Malhotra, A., Thomas, J. Carroll, J., and Miller, L. (1980). Cognitive processes in design. International Journal of Man-Machine Studies, 12 (119-140).

Newell, A. and Simon, H. (1972). Human Problem Solving. Englewood Cliffs, N. J.: Lawrence Erlbaum.

Robillard, P. N., D'Astous, P., Détienne, F., and Visser, W. (1998). Measuring cognitive activities in software engineering. Proc. of ICSE98, 20th International Conference on Software Engineering, Kyoto (Japan), April 19-25, 1998.

Simon, H. (1981). The structure of ill-defined problems. Artificial Intelligence, 4, 181-201.

De Terssac, G. and Chabaud, C. (1990). Référentiel opératif commun et fiabilité. In J. Leplat et G. de Terssac (Eds.) Les facteurs humains de la fiabilité dans les systèmes complexes. Paris: Octarès.

Visser, W. (1991). Evocation and elaboration of solutions: Different types of problem-solving actions. An empirical study on the design of an aerospace artifact. In T. Kohonen and F. Fogelman-Soulié (Eds.), COGNITIVA 90. At the crossroads of Artificial Intelligence, Cognitive science, and Neuroscience. Proceedings of the third COGNITIVA symposium . Amsterdam: Elsevier.

Visser, W. (1992). Designers' activities examined at three levels: organization, strategies & problem-solving. Knowledge-Based Systems, 5, 1, 92-104.

Visser, W. (1993). Collective design: A cognitive analysis of co-operation in practice. In N. F. M. Roozenburg (Ed.). Proceedings of ICED 93. 9th International Conference on Engineering Design (Volume 1), The Hague, August 17-19, 1993. Zürich: HEURISTA.

Studying cognitive effort and processes involved in computer graphics

Nathalie Bonnardel & Pascal Gaden
Université de Provence
Research Center in Psychology of Cognition, Language and Emotion
29, avenue Robert Schuman 13621 Aix en Provenc, France
nathb@up.univ-mrs.fr **pgaden@infonie.fr**

ABSTRACT

Measuring cognitive effort or mental load (or workload) and characterizing the evolution of cognitive processes have been main research topics for years in cognitive ergonomics, though no consensual method for studying such aspects has been proposed. Therefore, we first present a method that is traditionally used both to measure cognitive effort and study cognitive processes involved in writing activities: Kellogg's "triple task". Then, we argue that this method can be adapted to study professional activities. In order to illustrate the use of such a method, we present a study conducted with designers who are used to performing their tasks on a computational tool: computer graphics artists. This study aims at knowing more about (1) cognitive efforts required to perform cognitive treatments involved in computer graphics tasks and (2) the evolution of such cognitive treatments as the problem-solving progresses. The results we obtain are compared to descriptions of cognitive processes of designers working without computational tools and, based on these findings, we point out difficulties that seem to be due to the use of computational tools. Such an illustration of the use of the triple task allows us to conclude about the interest of this method to study real-life activities.

Keywords

Cognitive effort, mental load, design problem-solving, cognitive processes, computer graphics.

OBJECTIVE

The objective of this paper is twofold:

- To propose a method to both study cognitive effort or mental load required to perform real tasks and determine the evolution of cognitive processes while conducting a specific task;
- to discover more about design activities, and especially design activities conducted with a computational tool, as is the case in computer graphics.

Therefore, we are going to define the notion of "cognitive effort" or "mental load", present different methods that can be used to measure it and describe more precisely the method we chose to use to study design activities. Then, we will point out the current challenges of studying design activities performed using computational tools and, in order to know more about such activities, we will present a study that aimed at analyzing both cognitive effort and treatments involved in a design activity developed in the area of computer graphics.

TOWARDS DEFINING AND MEASURING MENTAL LOAD OR COGNITIVE EFFORT

Some definitions

The concept of "mental load" — also called "mental workload", "cognitive load" or "cognitive effort" — is a very old one, proposed in psychology of work in the years 60's. Before explaining such a notion, we can first distinguish the notion of mental load or cognitive effort from the feeling of tiredness (in general or specifically of mental tiredness): one can use important cognitive resources to perform a task without feeling tired. In addition, the term of tiredness may refer mainly to long term effects of work, whereas mental workload refers more to short term effects (Leplat, 1997). In order to go further, three notions seem important to be defined (see Halford, 1993):

- "*cognitive resources*", which correspond to the mental energy available for a given individual (or operator), at a given moment and to perform specific treatments:
- "*Cognitive capacity*", which refers to the maximal amount of resources that a given individual can mobilize;
- "*cognitive load*" ("mental load" or "cognitive effort"), which traduces the amount of cognitive or attentional resources used to perform a given task.

Especially, in accordance with the views of certain authors (see, for instance, Spérandio, 1988; Leplat, 1997), we consider that the cognitive load of a given individual (or operator) in charge of a given task does not only depend on requirements of the task at hand but also on different factors linked, for example, to the operators in charge of the task (e.g., their level of experience and the strategies they develop) or to the situation of work (e.g., the nature of the tools available to perform the task, the organization of work).

Concerning this last aspect, different terms can be used, such as mental load (or workload), cognitive load or cognitive effort. Personally, we prefer to avoid the term "load", which can refer to both a condition and a consequence of the activity (see Leplat, 1997). Therefore, we will use the term "cognitive effort" in the following parts of this paper.

Methods to measure cognitive effort

Since cognitive effort appears to be a main characteristic of human activities, how can we measure it? Several methods can been listed and characterized (for some of the following indicators, Spérandio, 1988; Leplat, 1997):

- *Indicators linked to the feeling of cognitive effort*, which consist of asking directly the operator to express his or her feeling about the cognitive effort or load, through for instance scales of subjective evaluation;
- *indicators based on the work activity*, such as changes of postures, characteristics of a visual exploration or changes of "operating behaviours" (Spérandio, 1988). This last kind of indicators consists in changes in the methods of work spontaneously used by the individuals or operators (see, for instance, Spérandio, 1977; de Viviès, 1999). Thus, in reaction to an increase of the task requirements, the operator can change of strategy or operating behaviour in order to decrease his or her cognitive load or effort;
- *physiological or neurophysiological indicators*, such as measures of the cardiac activity or of event-related brain potentials and, especially, the amplitude of the P300 for cognitive or attentional resources (see, for instance, Israel, Chesney, Wickens & Donchin, 1980). These indicators allow the determination of modifications of the organism or of the electro-encephalographical activity due to the task in progress;

- *indicators based on the performance of a task other than the main one*, as it is the case for the method of the "dual task" or "added task". This method, which has been used for more than 25 years, consists in asking the participant to perform, in addition to the main task, a secondary task — though the priority is explicitly assigned to the main task (see, for instance among numerous research works, Kahneman, 1973). This paradigm is based on descriptions of the human-being as a system of treatment with a limited capacity. Such a limitation is variable, for instance, according to the individuals or to the cognitive processes or operating behaviours they develop (see Spérandio, 1988). However, it is acknowledged that attentional resources available in working memory are limited (Baddeley & Hitch, 1974). Another hypothesis, which previously underlay the method of the "dual task", has been questioned: the hypothesis of a unique channel with no differentiation in the resources. The alternative hypothesis of multiple resources has been adopted by several authors (see Wickens, 1984). Especially, the model of Baddeley (Baddeley & Hitch, 1974) presents the architecture of the working memory as composed of three systems: the "central executive", which controls the attention and supervises the 2 other systems (called "slaves systems"), i.e. the "phonological loop" (for pronounceable information elements) and the "visuo-spatial sketchpad" (for visuo-spatial information elements).

All these indicators provide us with obviously only indirect information elements about the individual's cognitive effort and processes. Though their interest is dependent on the type of tasks to study, we can briefly point out some limitations of such indicators. First, indicators linked to subjective experience can be particularly difficult to interpret since they are influenced by other factors, such as the motivation of the individuals or the satisfaction they get from their work. Indicators based on the work activity can be relatively difficult to identify depending on the task at hand and they are also difficult to interpret. For instance, a change of operating behaviour can be performed to decrease a too important cognitive cost or effort, but it can also constitute an answer adapted to a current state of problem-solving (Spérandio, 1988). The link between classical physiological indicators and the cognitive activity seems to us difficult to establish, but event-related brain potentials seem to constitute more interesting indicators of individuals' cognitive effort or attentional resources, though their reliability may not always be guaranteed. Anyway, we consider that the identification of variations, whatever they are (operating behaviours, physiological or neuro-physiological data, reaction times to a secondary task), remains insufficient to characterize the operators' cognitive processes. Indeed, we argue that such methods need to be combined with others in order to determine which cognitive processes are associated to variations of cognitive effort. Therefore, we suggest the use of the paradigm of the "triple task", developed by Kellogg (see, for instance, Kellogg, 1987, 1994, 1998), which presents the interest of allowing the analysis of both cognitive processes involved in a given activity and cognitive efforts they require.

Description of Kellogg's "triple task"

The paradigm of the "triple task" (Kellogg, ibid.) is, firstly, based on the "dual task" that we evoked previously and, thus, on the hypothesis of limited attentional resources (Kahneman, 1973). The idea underlying this paradigm is that the residual attentional resources not used to perform the main task can be used to perform the secondary task, without profoundly modifying the main task (this aspect will be discussed in the following paragraph).

To measure residual resources, in Kellogg's paradigm, the participants are interrupted by auditory probes ("beeps") while performing the main task (e.g., writing a text) and they have to react as quickly as possible to these stimuli. More precisely, reaction times of each participant are firstly measured independently of any other activity, in order to define a mean basic reaction time specific to each participant. This basic reaction time is then calculated from the reaction times measured while performing the main task, in order to define "weighted reaction times" associated with cognitive processes. Therefore, the more significant the weighted reaction time is, the more resources are required by the interrupted cognitive process.

In addition, in order to analyze the evolution of cognitive processes, the participants in these studies have to indicate, through a "retrospection" task, the cognitive processes that have been interrupted by probes. For this last task, each participant has to choose among several labels the one corresponding to the cognitive process that has been interrupted. Especially, to study writing processes, the proposed labels correspond to processes described by Hayes & Flower (1980). Therefore, the participants have to perform the following triple task:

- Writing a text (main task);
- reacting to probes;
- indicating the cognitive process which has been interrupted.

This paradigm has been used for years to study mainly writing activities, but also to compare them with other cognitive activities, such as reading, learning and playing chess (see Kellogg, 1986; Britton, Glynn, Meyer and Penland, 1982, cited in Kellog, 1986). Our objective will be to show that it can also be used to analyze professional activities.

Limits of this method?

Though methodological criticisms have been expressed about the technique of the dual task (Fisk, Derrick & Schneider, 1986-87), several experiments have shown that the two additional tasks we just described (reacting to probes and indicating the interrupted cognitive processes) do not damage the main activity (see Piolat, Olive, Roussey & Thunin, 1999). These experiments are based on a comparison between "natural" conditions of writing (i.e. without any specific technique) and "unusual" conditions (main task associated to the two additional tasks), and they showed that:

There is no deterioration of the quality of the productions obtained. Writers can allocate enough attentional resources to preserve their writing objectives and they can define "commitments" in managing their attentional resources, which allow them both to produce a text with the same quality than in natural conditions and to perform the additional tasks (Kellogg, 1987);

Though the writers can feel bothered by the interruptions, especially when the probes are particularly frequent, the mobilization of the writing processes and the management of the task in its whole are not modified by the interruptions (Piolat, Olive, Roussey & Thunin, 1999). Therefore, we decided to adapt Kellogg's triple task to our area of design.

A STUDY OF COGNITIVE EFFORT AND TREATMENTS INVOLVED IN COMPUTER GRAPHICS

A current challenge

In today's workforce, there are many skilled professionals involved in design activities (e.g., in areas such as software design, mechanical design, or architectural design) and such skilled designers will become increasingly prevalent in the years ahead. In addition, in professional areas, the design activities are more and more realized using computational tools. Thus, there are very pragmatic reasons to know: (1) how designers perform tasks using such computational tools and (2) whether these tools effectively facilitate their design activities. Towards this end, numerous research efforts in computer science and artificial intelligence are focused on creating various types of knowledge-based design support environments. In the same time, several studies have been conducted in cognitive psychology and cognitive ergonomics to determine the influence of computational tools, such as CAD tools, on the designers' activities (see, for instance, Whitefield, 1986). These studies are usually based on observations associated with comments from the designers on their activities. Some interesting results have been obtained, but we still need to know more about the evolution of the cognitive processes required during a computer supported design activity.

In order to characterize the evolution of cognitive processes and the cognitive effort they require, we carried out an experiment in the area of computer graphics. We consider this area as interesting since the products the designers have to develop must be innovative as well as satisfy certain requirements.

Pre-experimental and experimental situations

Our experimental situation was based on a pre-experimental analysis performed in order to identify the cognitive treatments involved in the design of computer graphics products for an Offset printing office.

Pre-experiment

A pre-experimental analysis, based on interviews and observations, allowed us to analyze the tasks and activities of computer graphics artists. The interviews aimed at determining the nature of the main cognitive treatments these designers performed during their specific design tasks. Results of this analysis showed that the activities of computer graphics artists are centered on two main characteristics of the products they have to design: the shape and the colour of these objects. It also appeared that the designers could consider these objects at different levels of abstraction, which is in accordance with general results about design activities (see, for instance, Rasmussen, 1984). Therefore, we decided to distinguish global treatments and detailed treatments. In addition, we observed computer graphics artists while using a computational tool: the Illustrator software. This tool allows designers to perform innovative tasks (the use of this computational tool begins with a white page, that designers fulfill as they wish). The observations showed that they spontaneously interrupted their graphical activities on the computational tool, to assess what they were creating. In addition, these evaluations appeared to be performed either globally on the object or locally, i.e. with regard to specific and detailed points of view.

Experimental task

Based on these results, the triple task we asked the designers to perform in our experiment was computer-driven (based on the use of Scriptkell - Piolat, Olive, Roussey & Thunin, 1999) and it consisted of the following three tasks:

- Firstly, to solve a design problem representative of the usual tasks these computer graphics artists had to perform: to design a logo for a photography office. In order to induce innovative activities from the designers, this task had never been asked for by the employees' company;
- Secondly, during the same time, the designers were interrupted by auditory probes and had to react to them by pressing a button;
- Finally, to indicate the cognitive treatment they were performing while interrupted by the probe, designers chose one among the six cognitive treatments defined during the pre-experimental analysis:

    1. Global shape of the object to design;
    2. detailed shape;
    3. global colour (which corresponds to the choice of a shade of colour for the object);
    4. detailed colour;
    5. global evaluation of the object;
    6. local evaluation.

The experimental situation began by explanations from the experimenter. In addition, the designers had to perform two small training tasks corresponding to the two additional tasks. After having performed them, the designers had to perform the real experimental task (i.e. to design the logo), in about 1 hour.
Four professional designers specialized in computer graphics participated in our experiment. They were all working in the same Offset printing office, and they had all a similar level of expertise in graphical design activities as well as in the use of the Illustrator software.

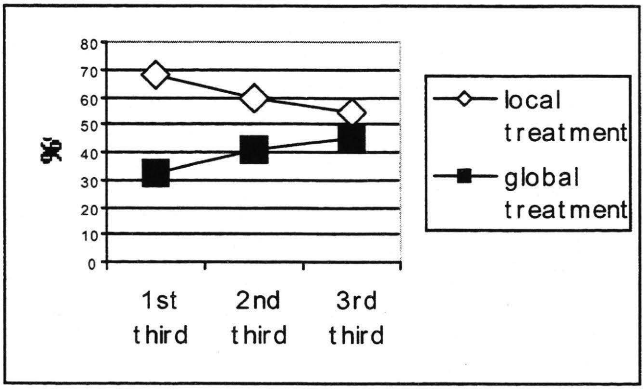
Results

Evolution of the design activity

The design activity is usually described as based on a progressive refinement of the designer's mental representation (Rasmussen, 1984). Consequently, we could expect that the designers would be first performing global treatments, which can correspond to a conceptual or abstract representation of the design solution (i.e., here, the logo), and then more local treatments. In order to characterize the evolution of the design activity, we decided to divide the total time of this activity into three parts (or "thirds"), in order to get three views of the cognitive treatments performed. Contrary to our expectation, we observed that (see Figure 1):

- Firstly, the designers who participated in this study performed more local treatments than global treatments (in mean 60.5% of the designations correspond to local treatments vs. 39.4% for the global treatments);
- secondly, the local treatments decrease as the design problem-solving progresses, whereas the global treatments increase.

Figure 1: Percentage of designation of global and local treatments according to the evolution of the design activity (thirds)



More precisely, we analyzed the evolution of the different cognitive treatments identified during the pre-experiment. Thus, we can see on Figure 2 that:

- The detailed treatments of shape and colour of the logo are more performed by the designers at the beginning of the task than at the end, whereas we observe an opposite tendency for the global treatment of colour (but not for the global shape);
- The global and local evaluations both increase as the task is progressing.

Figure 2: Percentage of designation of cognitive treatments with regard to the evolution of the design activity (thirds)

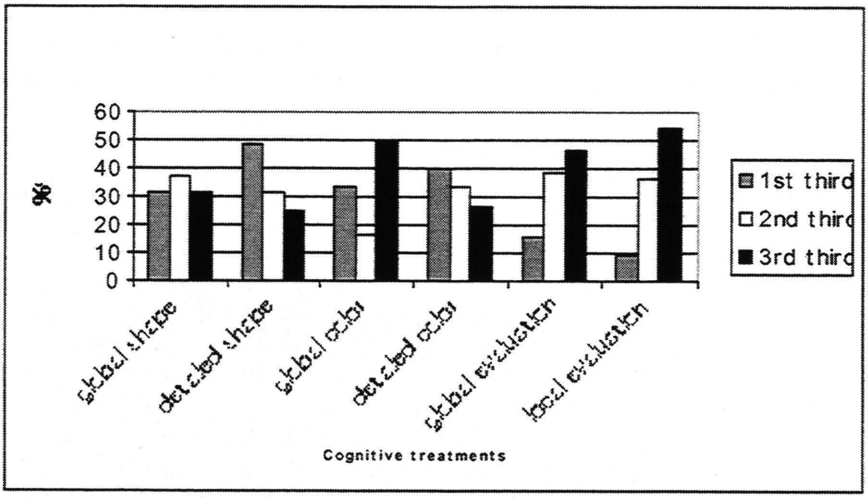
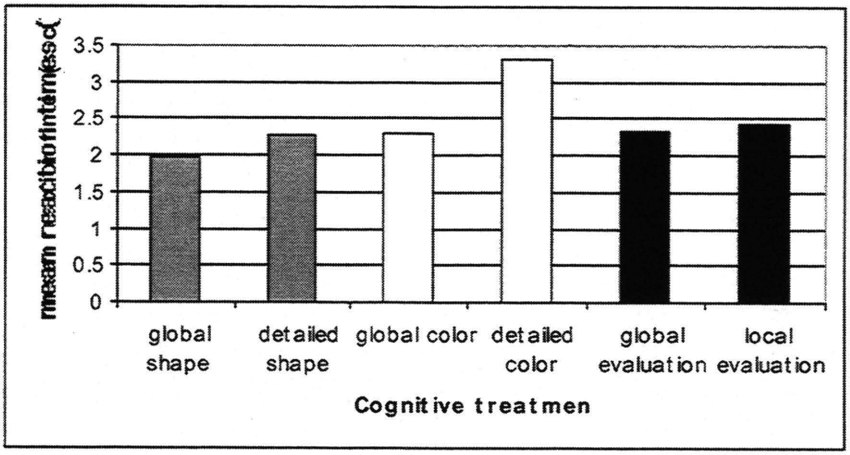


Figure 3: Mean reaction times according to the cognitive treatments



Cognitive effort

Concerning the cognitive effort, the results we obtained (see Figure 3) show that global treatments are associated to shorter reaction times than local treatments, whatever the global treatments are about (i.e. shape, colour or evaluation). Therefore, the global design of the logo seems to be easier to perform by the designers than the design of detailed features.

DISCUSSION

The results we obtained are going to be compared with previous results about the design activity and with results related to writing tasks (which can also be considered as specific design activities). Then, we will express some reflections about the influence of the use of computational tools on the design activity.

Comparison with previous results about design activity

Concerning the design activity itself, our results firstly seem to show an opposite tendency to previous descriptions of design activity. Indeed, the design activities are usually described as schematic at the beginning and as, progressively, more and more precise. Contrary to these descriptions, our subjects began their activities by performing local and precise treatments.

Secondly, our results seem also different from descriptions of the design activity as based on the intertwining of both processes of solution generation and solution evaluation (see Bonnardel, 1999). Such differences with previous descriptions of the design activity could result from the use of a computational tool.

The effect of computational tools

Our results are in accordance with previous ones and show that the use of computational tools, which can consist of CAD tools or computer graphics tools (as the one used in our study), compel designers to make decisions about detailed features of a design object in order to represent it graphically on the computer. Such decisions may occur too early and therefore, contradict the strategies developed by designers when they do not use computational tools. Indeed, previous research showed that the designers try to postpone decisions as late as possible to keep the space of solutions large (see Lebahar, 1983). Moreover, we showed in our study that the treatment of detailed features of the object required the most important attentional resources. This result may be specifically linked to the use of a computational tool, which involves the manipulation of complex functions of the software.

In addition, the use of a computational tool may make evaluations of the object in the course of the design activity difficult. The designers who participated in our study seem to be focused on the graphical representation of the objects (especially, to create detailed features) and therefore, only assess their design at the end of the development of important parts of the objects.

These two observations allow us to point out risks associated to the use of computational tools: if the designers' decisions are not appropriate, they may discover it only late (and, maybe, too late). Since such inconveniences seem difficult to avoid, we suggest to associate a "critiquing" system (or module) to computer graphics tools (see Fischer, Lemke, Mastaglio & Morch, 1991). Such a critiquing system would thus support designers in assessing earlier their design, for instance, as soon as a design decision opposite to certain constraints is made. The interest of these systems is that they can support designers in evaluating their products without hindering or interfering with the designers' innovative capabilities. In addition, it has been shown that they support both experienced and novice designers, though they have a different effect depending on the designers' level of experience (Bonnardel & Sumner, 1996; Sumner, Bonnardel & Kallag-Harstad, 1997)

CONCLUSION

The study we conducted in the area of computer graphics shows that the paradigm of the triple task is interesting both to identify cognitive treatments involved in real professional activities and to determine the cognitive efforts required by these treatments or processes. Moreover, we can point out the fact that this method is relatively easy to use, since it does not require a heavy equipment (only a portable computer is needed).

However, we saw that, contrary to the analysis of writing activities, which can be conducted on the basis of cognitive processes previously identified, the use of such a method in a real professional activity can only occur after performing a classical analysis of the task and activity of the operators.

Thus, we can conclude that when integrated in an ergonomic approach — and, therefore, associated to other classical methods (such as interviews and observations) — the triple task we described is particularly useful to obtain precise and precious information elements about the cognitive functioning of designers and, potentially, of other types of operators.

REFERENCES

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), The psychology of learning and motivation (vol. 8). London: Academic Press, 281-301.

Bonnardel, N. (1999). L'évaluation réflexive dans la dynamique de l'activité du concepteur. In J. Perrin (Ed.), Pilotage et évaluation des activités de conception. Paris : Editions L'Harmattan, 87-105.

Bonnardel, N. & Sumner, T. (1996). Supporting evaluation in design. Acta Psychologica, 91, 221-244.

Britton, B.K., Glynn, S.M., Meyer, B.J.F., & Penland, M.J. (1982). Effects of text structure on the use of cognitive capacity during reading. Journal of Educational Psychology, 74, 51-61.

Fischer, G., Lemke, A. C., Mastaglio, T., & Morch, A. I. (1991). Critics: An emerging approach to knowledge-based human computer interaction. International Journal of Man-Machine Studies, 35(5), 695-721.

Fisk, A. D., Derrick, W. L., & Schneider, W. (1986-87). A methodological assessment and evaluation of dual-task paradigms. Current Psychology Research & Reviews, 5, 315-327.

Halford, G.S. (1993). Children's understanding. Hillsdale, NJ : Lawrence Erlbaum.

Hayes, J.R., Flower, L.S. (1980). Identifying the organization of writing processes. In W. Gregg & R. Steinberg (Eds.), Cognitive processes in writing, Hillsdale, NJ: Erlbaum, 3-30.

Israel, J.B., Chesney, G.I., Wickens, C.D. & Donchine, E. (1980). P300 and tracking difficulty: evidence for multiple resources in dual task performance. Psychophysiology, 17, 259-273.

Kahneman, D. (1973). Attention and Effort . Englewood Cliffs, NJ : Prentice-Hall.

Kellogg, R.T. (1986). Designing idea processors for document composition. Behavior Research Methods, Instruments, & Computers.18, 118-128.

Kellogg, R.T. (1987). Effects of Topic Knowledge on the Allocation of Processing line and cognitive effort to writing process. Memory and Cognition, 15, 3, 256-266.

Kellogg, R. T. (1994). The Psychology of Writing. New York : Oxford University Press.

Kellogg, R. T. (1998). Components of working memory in text production. In M. Torrance & D. Galbraith (Eds.), Cognitive demands of writing. Amsterdam: Amsterdam University Press.

Lebahar, J.C. (1983) Le dessin d'architecte. Roquevaire : Parenthèses.

Leplat, J. (1997). Regards sur l'activité en situation de travail: Contribution à la psychologie ergonomique. Paris: PUF.

Piolat, A., Olive, T., Roussey, J.-Y., & Thunin, O. (1999). SCRIPTKELL : An experimental tool assisted by computer for measuring cognitive effort and time processing during writing and other complex cognitive activities. Behavior, Research Methods, Instruments & Computers, 31(1), 113-121.

Rasmussen, J. (1984). Strategies for state identification and diagnosis in supervisory control tasks, and esign of computer-based support systems. Advances in Man-Machine Systems Research, 1, 139-193.

Spérandio, J.C. (1977). La régulation des modes opératoires en fonction de la charge de travail chez les contrôleurs de trafic aérien. Le Travail Humain, 40(2), 249-256.

Spérandio, J.C. (1988). L'ergonomie du travail mental. Paris: Masson.

Sumner, T., Bonnardel, N. & Kallag-Harstad, B. (1997). The cognitive ergonomics of knowledge-based design support systems. Proceedings of the International Conference CHI'97 - Human Factors in Computing Systems (Atlanta, March 22-27), 83-90.

de Viviès, X. (1999). Points de vue et type de représentation des règles. Deux niveaux de difficulté pour la résolution de problèmes. L'Année Psychologique, 99, 271-293.

Whitefield, A. (1986). An analysis and comparison of knowledge use in designing with and without CAD. In A. Smith (Ed.) Proceedings of CAD'86, Londres: Butterworth.

Wickens, C.D. (1984). Engineering psychology and human performance. Columbus : C.E. Merrill.

Session #3: Controlling Processes

Difficulty And Safety During The Management Of A Severe Incident Management Sequence

Anne Edland

Swedish Nuclear Power Inspectorate,
SE-106 58 Stockholm, Sweden
anne.edland@ski.se

Ola Svenson

Department of Psychology, Stockholm University,
S-106 91 Stockholm, Sweden
osn@psychology.su.se

Erik Hollnagel

University of Linköping,
SE-581 83 Linköping
eriho@ikp.liu.se

ABSTRACT

Different judgmental approaches were applied simultaneously to assess the safety of a hazard management process in a process industry. In doing so, a steam tube rupture incident in a nuclear power plant was first modelled in a sequential task analysis. Difficulty, likelihood of failure and severity of failure consequences were rated for each of the stages of the analysis. Time available for each of the stages, were estimated in a simulator and by operators. The results showed how the approach could be applied and how potential weak spots in the management process were identified. The contribution illustrates the advantages of a multi-method approach to safety analysis of a hazard management process and underlines the need of conscious training of usually automatized largely unconscious routines.

Keywords

Incident management process, safety, human factors, assessment, time, difficulty, failure.

INTRODUCTION

When preparing for severe accident management in a process industry such as a nuclear power plant, there are some accident scenarios that are particularly interesting from a safety point of view. In the nuclear industry they are typically selected because they loom large in probabilistic risk analysis (PRAs) of a plant. One of the scenarios involves a steam generator tube rupture. The steam generators in a pressurized water nuclear power reactor (PWR) transmit energy from the primary circuit to the secondary side that drives the turbines and the generator. A steam tube rupture starts a very interesting accident management sequence, not only from a safety perspective, but also from a human factors perspective. As just mentioned, a steam generator tube rupture (SGTR) has a prominent role in PRAs of pressurized water nuclear power reactors and this role has been validated by incidents that have taken place at nuclear power plant sites over the world.

From a human factors point of view the SGTR sequence is important because it involves a number of manual sub-processes involving human information processing and action. Because of the important role played by the operators during the SGTR sequence, this sequence can also be used as a prototype sequence that can be used to develop generic methods for human reliability analysis and for quality assessments of these methods.
Specifically, the human reliability assessments (HRAs) obtained for a SGTR can be inserted in existing PRAs to replace earlier estimates.

The purpose of the present study is to develop an approach that analyses the reliability of the sub-processes in a SGTR incident with particular emphasis on the human factor component. The assessment of human reliability for different segments of the sequence will be made in different ways. The first of these is based on the work carried out by Hollnagel who developed the method of cognitive task analysis, CORA (Hollnagel, 1998). Based on this method, human factors experts will rate the human reliability of the different sub-processes needed to manage a SGTR sequence. The second approach involves instructors at a simulator-training center who will contribute a second set of ratings of the human reliability of the same subrocesses. Third, a group of nuclear power plant operators, who are very familiar with handling the sequence, will provide more detailed ratings related to the reliability of the different human system interactions in the sequence. Therefore, the analyses of the data from the group of operators will be carried out in more detail than for the other groups. Finally, a comparison will be made between the different approaches.

In the following, the SGTR sequence will first be described in general terms. There then follows the decomposition of the sequence into sub-processes. The different ways of assessing human reliability will be presented next. The main interest will be devoted to the detailed assessments of different aspects of safety provided by the group of nuclear operators.

Pressurized water nuclear reactors have two systems, the primary and the secondary loop systems, transferring energy from the nuclear process to the turbines coupled with generators producing electricity. The primary system passing the nuclear core circulates water under high temperature and pressure. In the steam generators (SGs) this water passes through a great number of narrow tubes surrounded by water and steam of the secondary system. The pressure is lower in the secondary system than in the primary system. Water from the secondary system is fed into the steam generators and transformed to steam which continues in pipes to the turbines. The Swedish PWRs have three steam generators each, but this number varies across plants in other countries.

The pipes of the primary system leading from the core to the steam generator tubes are very wide and by design contain no valves which could be closed, isolating the reactor vessel from the steam generators. Connected to the primary system, there is a pressurizer, with relief valves opening if the pressure becomes intolerably high. The pressurizer has a measurable water level below its content of steam. All the remaining matter in the primary system is water under high pressure. Thus, the status of the primary system is to a great extent reflected in the level and pressure in the pressurizer, which is connected to alarms and safety functions.

Reactor coolant pumps circulate water in the cold legs of the primary system to the reactor from the steam generators. In case of a cooling failure, there are high-pressure safety injection systems and also a low-pressure safety injection system connected to the primary system. On the secondary side, the feed water loop supplies the steam generators with water that is transformed to steam fed to the turbines. If feedwater fails, there is an auxiliary feed water system that can be used to prevent the steam generators from boiling dry.

If there is a leakage between the primary and secondary systems, water flows from the primary into the secondary system. This leads to loss of water and pressure from the primary system and increasing pressure in the secondary system coupled with the inflow of very hot water into the secondary system. This is a very unfortunate situation that may lead to serious consequences. Because it is impossible by design to isolate the primary system from leaking steam generator tubes, isolation has to be made through isolating the failing steam generator from the rest of the secondary system. This in turn leads to increased pressure in the isolated steam generator that may lead to the opening of pressure relief valves.

Generally speaking, the main goals for the incident management process in a steam tube rupture incident are to bring down the pressure in the primary system so that it balances that of the secondary system and to stop the nuclear process. Very briefly, a steam tube rupture is normally followed by a reactor trip (manual or automatic). After this, the event leading to the reactor trip has to be classified by the operators. The leaking steam generator has to be identified and isolated. The reactor then has to be cooled down to enable a depressurisation that will stop the leak flow. After having balanced the pressures in the primary and secondary systems, the reactor has to be brought to safe conditions.

The top event in the corresponding probabilistic risk assessment (PRA) applied here, is water overflow in the damaged steam generator that in turn may lead to core damage. This top event was chosen because it has a greater probability of occurring than core damage and it makes the top event more realistic to the groups of judges (human factors experts, instructors and operators). When a steam generator overflows in a SGTR sequence, there will be a direct path from the primary side to the environment outside through the safety relief valves of the steam generator. Primary steam-water will then leak out into the surroundings and cause contamination and other damage.

To partly recapitulate, the overall goal of the operators in a SGTR incident is to manage the plant and bring it into a safe operating state. Balancing secondary and primary pressures of the damaged steam generator and stopping the outflow of primary water to the secondary side is the top priority to achieve safe operating conditions. As mentioned earlier, the pressure of the primary system must be lower than the pressure of secondary side of the ruptured SG, and the secondary side pressure of the SG must be lower than the set point of the SG safety relief valves. This can be referred to as pressure balance. Another important and necessary goal appearing early is the removal of residual heat from the reactor. This is done in order to bring down the pressure in the primary system.

When there is an indication of tube rupture, operator action has to be taken. The emergency operating procedures dealing with a tube rupture are found in what the Westinghouse manufacturer calls E3. Other relevant emergency operating procedures are e.g., E0 used to identify the tube rupture sequence and other procedures applicable to achieve pressure balance in case of other events or failures in the plant during the accident management sequence.

To summarize, in order to manage a SGTR incident, the operators have to carry out the following main sub-processes. (1) Identification of the steam generator (SG) with a rupture. (2) Isolation of the damaged steam generator. (3) Stop of feed water (FW). (4) Reduction of temperature of primary system (RCS, cool down). (5) Re-establishing pressurizer (PRZ) level. (6) Stop of Safety Injection, SI (7) Successful accomplishment of pressure balance.

The chief goal of the present study is to find a way to analyse the hazard management process associated with a SGTR sequence, in order to find weaknesses in that process.

METHOD

Participants

Two human factors experts, 2 simulator instructors and 24 nuclear power operators participated in the study. Twelve of the operators were reactor operators, 7 supervisors and 5 assistant reactor operators at the same Swedish nuclear power plant. All operators had worked in the control room of the plant (average 15 years). Each one of the operators had participated in simulator training of some kind of tube rupture simulations on the average about 27 times (mean of self-ratings).

Human factors experts and instructor ratings

The instructors made expert judgements of the degree of difficulty of the 7 different sub-processes, based on personal experience, knowledge and expertise. The human factors experts made judgements of difficulty using a more formalised procedure. The considerations used in this procedure were; (1) Time available for each sub-process (based on simulator exercises), (2) Number of steps/actions included in each sub-process and (3) The cognitive activities included in each sub sequence (based on Cognitive task analyses introduced by Hollnagel and applied to the present sequence by Hollnagel, Edland and Svenson, 1996). All three considerations for analysis were first made for each sub-process. After this, an informal integration was made within each sub-process taking into account the time available, the number of steps, and the cognitive activities of which the steps/operations consisted.

Operator ratings

A questionnaire based on the SGTR procedures was composed and the questions followed the seven main sub-processes presented above representing the SGTR scenario. As mentioned earlier the seven sequences are: (1) Identification of the steam generator (SG) with a rupture, (2) Isolation of the damaged steam generator, (3) Stop of feed water (FW), (4) Reduction of temperature of primary system (RCS cool down), (5) Re-establishing the pressurizer (PRZ) level, (6) Stop of SI, and, finally (7) Successful accomplishment of pressure balance. All ratings were given on scales ranging from 1 to 10 with 1 corresponding to small probability/ not serious at all/ not difficult at all/ not good at all, and 10 corresponding to very high probability/ very serious/ very difficult/ or very good. (The naming of the extreme points of a scale depends on each individual question).

The questionnaire also included questions concerning the work situation during the tube rupture event. The questions concerned the following; How (1) time pressured, (2) stressed, (3) activated, (4) passive did the subjects feel they would be during the event and how (5) co-operation demanding, and (6) well-organized they judged the situation to be and to what extent there was a (7) well-functioning assignment and delegation of work. Finally, the subjects were asked to assess to what extent the event was characterised by (8) high mental load, (9) lack of information, (10) track keeping of many things at the same time, (11) doing many things at the same time, (12) having control of the tempo, and finally (13) making a lot of decisions. All questions in this part were answered on a scale from 1 (do not agree at all) to 10 (agree fully).

The form was distributed to operators who participated in training sessions at a plant specific simulator. The operators filled in the questionnaire in approximately 90 minutes. As mentioned earlier, all operators were familiar with the SGTR sequence from experience gathered in several full-scope simulations of the sequence.

RESULTS

Difficulty

The human factors experts (N=2), instructors (N=2) and operators (N=24) rated the difficulty of each of the 7 sub-processes on scales from not difficult at all to very difficult. The human factors experts produced joint ratings after a discussion among themselves. The two instructors followed the same procedure. Both these groups used 5-step scales. In contrast, the operators all produced individual ratings through the whole experiment. They used 10 step scales for difficulty. The mean ratings were computed in the operator group and compared to the ratings of the human factors experts and instructors.
To achieve compatibility across scales, the ratings were standardized in each group through dividing each rating (jointly produced in each of the two first groups and the mean in the operator group) by the standard deviation across the 7 sub-processes within each group. Table 1 shows the results for three different categories of difficulty ratings.

Not surprisingly, the difficulty varies across sub-processes for all groups. The most difficult sub-process is the last, achieving pressure balance, for all groups. Stop of SI, is judged as the least difficult for the human factors experts and instructors while stopping feedwater is the least difficult for the operators. Although differences seem to exist, the small numbers of Human Factors experts and instructors do not allow statistically significant conclusions.

Pearson correlations between the scales in Table 1 indicate significant ($p<0.05$) relationships between the instructor judgments and the other two sets of judgments (0.76 with the operators and 0.78 with human factors experts). The correlation between human factors expert and operator ratings was 0.53 (non-significant). When partial correlations were computed controlling for the ratings not included the result was the following, 0.65 (instructors and operators), 0.68 (instructors and human factors experts), and -0.15 (human factors experts and operators).

In the following, the analysis will concentrate on more specific assessments from the operators of the different sub-processes in the SGTR scenario.

Table 1: Human factors experts and Instructors judgements and assessments from operator of degree of difficulty (standardized values, see text for explanation)

SUB-PROCESS	DEGREE OF DIFFICULTY		
	Human Factors	Instructors	Operators
(1) Identification of the steam generator (SG) with rupture.	4.64	2.55	3.54
(2) Isolation of the steam generator with rupture.	3.86	1.70	2.87
Stop of feed water (FW).	3.09	1.70	1.79
(4) Reduction of temperature of primary system (RCS, cool down).	3.86	1.70	2.67
(5) Re-establish the pressurizer (PRZ) level.	3.09	1.70	3.43
(6) SI stop.	1.55	0.85	2.31
(7) Pressure balance.	4.64	4.25	5.13
Standard deviation across sub-processes	0.65	1.18	0.85

Difficulty, Risk and Failure Consequences

In addition to the comparisons along the difficulty dimension, the failure consequences and risk of failure are also interesting to study. Table 2 gives the operators' mean difficulty ratings and the mean seriousness and risk of process failure ratings of the different sub-processes. The ratings were all given on 10 point scales. The standard deviations are given in the parentheses.

Table 2 shows that level of difficulty is not perfectly correlated with the estimated risk of process failure. Identification is judged to be the second most difficult process while the risk of failure of identification is at minimum. Seriousness of an identification failure is judged as the second most serious failure. The most serious consequence follows a failure in the SI stop process. However, in this case the SI process is very low both on difficulty and risk of failure.

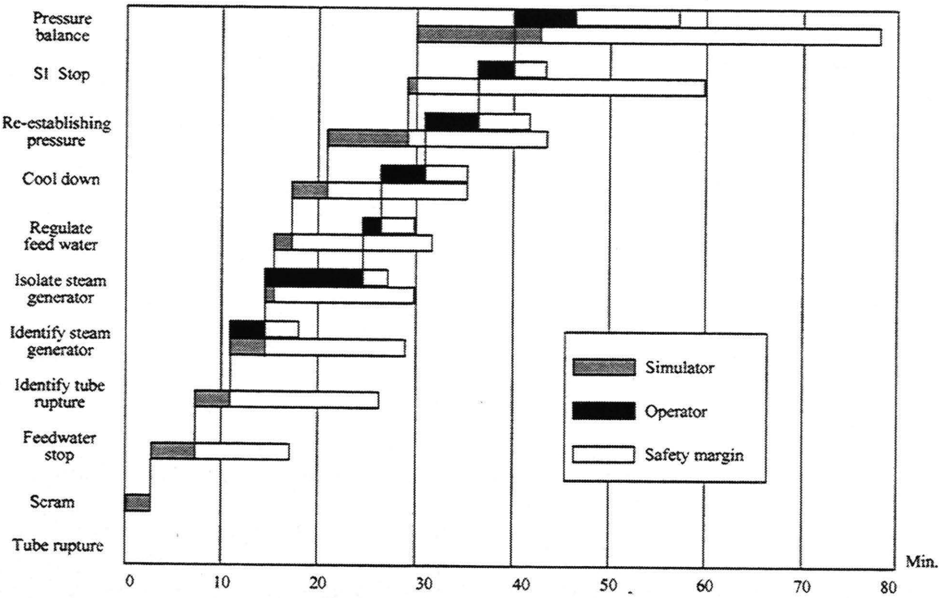
The SGTR scenario was run on a computer simulator (CENTS) in order to establish estimates of maximally available time for the different sub-process in a successful management process (c.f. Andersson & Edland, 1996). Input to the CENTS simulation was time used for the different sub-processes (that were identified in a session with simulator instructors running a full scope simulator). This time pattern represents optimal performance with most efficient use of time.

For each sub-process the maximum time available for performing that sub-process in which a top filling could be avoided was determined (in a computer simulation of the physical processes of the plant) (see Figure 1). In parallel with the optimal instructor times, the operator's judgments of time typically needed were also marked in Figure 1. Thus, a comparison can be made between the estimated times from the simulator (CENTS) and the times judged by the operators (Andersson & Edland, 1996). In general, operators judged the safety margins as shorter than the simulator data indicate. However, for isolation, operators believed that normally they would use up most of the estimated maximally available time. This in turn would give the operator less time for the following sub-processes and could induce later feelings of time pressure.

Table 2: Overall averages of the assessments given by the operators (n=24) to three questions. (1) The difficulty of each sub-process (2) The probability of failure with a step or within a sequence (3) The seriousness of the consequences. The answers are made on a scale from 1 (not difficult/probable/serious at all) to 10 (very difficult/probable/serious) The numbers in parenthesis are standard deviations. In contrast to Table 1 the values in this table were not standardized

Sub-Process	Level of difficulty	**Probability of sub-process failure**	Seriousness of consequences
Identification	3.00 (1.17)	1.46 (0.66)	9.25 (1.80)
Isolation	2.43 (0.84)	2.33 (0.82)	7.71 (2.68)
Feedwater stop	1.52 (0.67)	1.71 (0.91)	8.92 (0.97)
Cool down	2.26 (1.01)	2.00 (0.98)	8.57 (1.95)
Re-establish pressure	2.91 (1.24)	2.38 (1.10)	8.91 (1.44)
SI Stop	1.96 (0.77)	1.83 (0.87)	9.58 (0.58)
Pressure balance	4.35 (2.33)	2.71 (1.23)	7.38 (1.88)

Figure 1: Physical safety margins, simulator estimates and operator judgements of time needed to perform different sub-processes in a steam tube rupture scenario



Characteristics of the work situation

The operators made assessments of the working situation during a SGTR scenario. Table 3 shows the averages as well as averages for shift supervisors, reactor operator and assistant reactor operators. Looking at the results it is first important to note that 5 represents a neutral position. The overall average results indicate disagreement with the characteristics of "passive" (1.42), "missing information" (2.25) and "self control of the tempo" (2.13). The sequence is characterized by an active role played by the operator in a time-pressured situation, rather stressful and with a low degree of self-control of tempo but with sufficient information available. Furthermore, the situation is considered well organized and well functioning.

This self-declared assessment of the work situation was shared by all categories of operators with, as could be expected, the shift supervisors having greater values for most mental resource demanding characteristics. To, summarize, the feelings of low self-control of tempo and high time pressure provides a suboptimal situation for the operators. In this kind of situation the demands on prior training, technical and organizational support are very high.

Table 3: Assessments of the work situation during a SGTR scenario for shift supervisors, reactor operators and assistant reactor operators. The scale runs from 1 (does not agree at all) to 10 (fully agree)

Work situation	Overall average	Shift supervisors	Reactor operators	Assistant reactor operators
Time pressured	**9,08** (1,02)	**9,43** (0,79)	**9,00** (1,28)	**8,80** (0,45)
Stressed	**7,13** (1,62)	**7,43** (1,40)	**7,08** (1,93)	**6,80** (1,30)
Active	**9,54** (0,66)	**9,71** (0,49)	**9,50** (0,80)	**9,40** (0,55)
Passive	**1,42** (1,84)	**1,00** (0,00)	**1,83** (2,59)	**1,00** (0,00)
Co-operation demanding	**9,21** (0,93)	**9,43** (0,79)	**9,00** (1,13)	**9,40** (0,55)
Well organised	**9,58** (0,58)	**9,71** (0,49)	**9,58** (0,67)	**9,40** (0,55)
Well functioning distribution of task	**9,54** (0,59)	**9,57** (0,53)	**9,58** (0,67)	**9,40** (0,55)
High mental load	**7,67** (2,43)	**8,00** (2,77)	**7,42** (2,57)	**7,80** (1,92)
Missing information	**2,25** (1,22)	**1,86** (0,90)	**2,50** (1,57)	**2,20** (0,45)
Keeping track of many things at the same time	**6,67** (2,44)	**5,57** (2,51)	**7,00** (2,45)	**7,40** (2,30)
Perform many things at the same time	**5,83** (2,87)	**4,86** (2,61)	**6,42** (2,81)	**5,80** (3,56)
Self control of the tempo	**2,13** (2,35)	**1,29** (0,49)	**2,25** (2,53)	**3,00** (3,39)
Many decisions	**7,21** (2,62)	**7,29** (2,43)	**7,25** (2,83)	**7,00** (2,92)

CONCLUDING REMARKS

The present study has illustrated how different experts judge the difficulty of different sub-processes in a severe incident. The inter-rater correlations were reasonably high (about 0.75) with the instructors and the other two groups. Partial correlations between operators and human factors specialists, with instructor influence controlled for, were not significantly different from zero. Thus, human factors specialists and operators express in their ratings partly different sets of information. Both sets of information are likely to be of value when evaluating the risks of the situation. For example, the identification of the correct steam generator may not be as easy as assumed by the operators themselves and the pressure balance can be more difficult than the human factors specialists believe.

When the operators were asked about the probability of an error and the seriousness of that error if it should happen, the failure of SI stop was more likely to lead to a more serious consequence than an error of identification (Table 2). Again, this is a result supporting the earlier findings exposed in Table 1. Thus, the operators seem to have a low level of vigilance towards an error in detecting the correct steam generator. A simulating training scenario involving such a difficulty could enhance operator vigilance in this respect.

It is interesting to note that expert performance can be seen as either (1) largely consciously controlled and automatized or (2) largely unconsciously controlled. When behavior has become automatized it is not so easy for the performer to judge the likelihood of a failure (Hedenborg & Svenson, in press) or to improve performance (Ericsson, personal communication). Therefore, it is important for expert operators not to rely on their automatized performance routines only, but to reflect on this performance and train in a conscious mode for improvement readiness in case of an accident.

The rated safety margins in Figure 1 show that operators believe that they can identify the correct steam generator as fast as physically possible and that the operators have less time than they believe to check that they isolate the correct steam generator. In most of the other cases the operators seem to have quite realistic views on the time safety margins they work with.

In summary, the present study has shown how operators judge a severe incident in terms of difficulty work conditions and time available. The procedures used in the present contribution can be applied to any incident sequence to find out where operators underestimate or overestimate parameters (e.g., time available) in the situation. Analyses like the present one can inform an analyst about which the risky and safe segments are in any hazard management process.

ACKNOWLEDGMENTS

The authors are grateful for comments and assistance from Petra Söstrand, Bengt Ljungqvist and Johan Enkvist. Funds from the Swedish Nuclear Power Inspectorate and NKS supported this study.

REFERENCES

Andersson, S-O., & Edland, A. (1996) NKS/RAK Subproject 3, Integrated safety analysis, steam generator tube rupture, method description. NKS/RAK-1(96)R4, Vattenfall GES 35/96.

Hedenborg, M-L & Svenson, O. (In press) "Errors in subtractions: judged and observed frequencies" Scandiavian Journal of Psychology, in press.

Hollnagel, E. (1998) Cognitive reliability and error analysis method – CREAM. Oxford UK: Elsevier.

Hollnagel, E., Edland, A., & Svenson, O. (1996) A cognitive task analysis of the SGTR scenario, NKS/RAK-1 (96)

A Three-stage Information Analysis to Support Predictive Human Reliability Analysis

Wondea Jung

Integrated Safety Assessment Team,
Korea Atomic Energy Research Institute, P.O.Box 150, Yusong, Taejon, Korea
wdjung@kaeri.re.kr

Wan C. Yoon

Department of Industrial Engineering,
Korea Advanced Institute for Science and Technology, Kusong-Dong, 373-1, Yusong, Taejon, Korea

ABSTRACT

This paper suggests a new information analysis method that helps human reliability analysts investigate the task structure and context systematically from a cognitive engineering perspective. The method consists of three stages: the scenario analysis, the goal-means analysis, and the cognitive function analysis. Through the three-stage analysis, a systematic observation is undertaken from the macroscopic view on task situations to the microscopic view on specific cognitive processes. It would be possible and desirable to use the proposed method as a part of an advanced human reliability analysis (HRA) method for a better prediction of human errors.

Keywords

Task analysis, human reliability analysis, performance influencing factors.

INTRODUCTION

Regarding the conventional HRAs, several limitations, including theoretical problems as well as weaknesses in the quantification technique, were discussed from the various perspectives and levels of depth (Apostolakis et al., 1990). In order to resolve the limitations of the conventional HRAs, new methodologies that focus on cognitive errors have been or are being developed for HRA.

The recently developed 2nd generation HRA methods include MERMOS (Pierre, 1999), CREAM (Hollnagel, 1998), ATHEANA (USNRC, 1998), and HRMS (Kirwan, 1997). The latter three methodologies provide detailed taxonomies and analysis processes for error prediction and quantification based on a cognitive model. The taxonomies include performance influencing factors (PIFs), cognitive stages, and detailed error types.

However, there are still difficulties and uncertainties doing HRA since it basically depends on the analyst's expertise and judgement. Most HRA methodologies including the newly developed ones are model-based; the possible error types and probabilities can be derived by the values of the model parameters. Being supported by scarce empirical data, the values of the model parameters have to be determined on the basis of the analyst's knowledge on human performance in the given tasks and their context. Therefore, the outcome of HRA may only be warranted by a proper acquisition of relevant information about the tasks and situations. The contents of information that may affect the quality of the analysis include the relevant situation factors, task attributes, task requirements and the cognitive elements of the tasks. For the appropriate acquisition of such information, a well-defined task analysis process must precede the main analysis phase of the HRA.

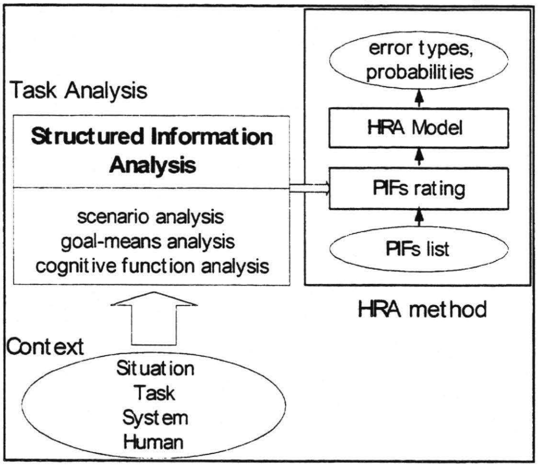
In order to assist the analysts in collecting and structuring information on tasks and contexts, this paper suggests a three-stage structured information analysis (SIA) as a task analysis method for HRA. It aims at helping analysts to investigate the task structure and context systematically from a systems- engineering viewpoint. The SIA is especially useful for the analysis of emergency tasks in nuclear power plants (NPPs) in which complex interactions between error causes can occur. This method emphasizes the contextual information as well as the cognitive process of the tasks. Figure 1 shows the schematic relationship between the SIA and the overall process of HRA.

TECHNICAL BACKGROUNDS

Cognitive model

Error can be defined as the deviation from the normal expectation occurring while an individual, group or organisation performs given task. For error analysis, taxonomies, error mechanism and cognitive model are needed to describe the process and related influence factors to error occurrences. The cognitive process is different depending on the combination of task characteristics or PIFs, and accordingly, the related psychological phenomena are classified. Rasmussen's SRK framework presents that the pattern of cognitive processes varies according to the skilled level of the task. In this study, we use the five cognitive stages; detection (D), observation (O), state identification (I), goal setting (G), and planning/execution (E). The cognitive model is the one used in the K-HPES (Yoon et al., 1996), which is a system for analyzing past error events. The cognitive model is basically similar to the one of CREAM or ATHEANA, except for the classification of information detection stages into detection and observation. This classification is based on the definition that detection is the term representing the passive process and observation is for the active process.

Figure 1: The process of HRA supported by SIA



PIFs and information for HRA

The selection of PIFs affects the HRA result directly since the quantification of HEP is conducted on the basis of the assessment of the pre-defined PIFs. We selected the influence factors affecting error occurrences in emergency operation tasks of NPPs for HRA. The comprehensive knowledge of the task structure and situation can be obtained through the analysis of the major factors affecting error possibility. There are a great number of factors influencing error occurrence.

However, the factors that should be paid special attention in view of the prediction of the error types and possibilities can be organized within a manageable level. To be applicable, it is desirable to minimize the number of factors without omitting the important structural information or PIFs. In this study, 11 PIFs were finally adopted from the prior study (Kim et al., 1999).

Besides PIFs, there is some additional information needed for the SIA. Such information includes task steps, cognitive stages and patterns of cognitive function related to the task steps, input/output information, and the implied goals. Those information items are not PIFs, but should be collected and analysed for error analysis. All items, including PIFs, to be analysed in SIA are arranged in Table 1, categorizing them into three parts.

Table 1: Information items for the SIA

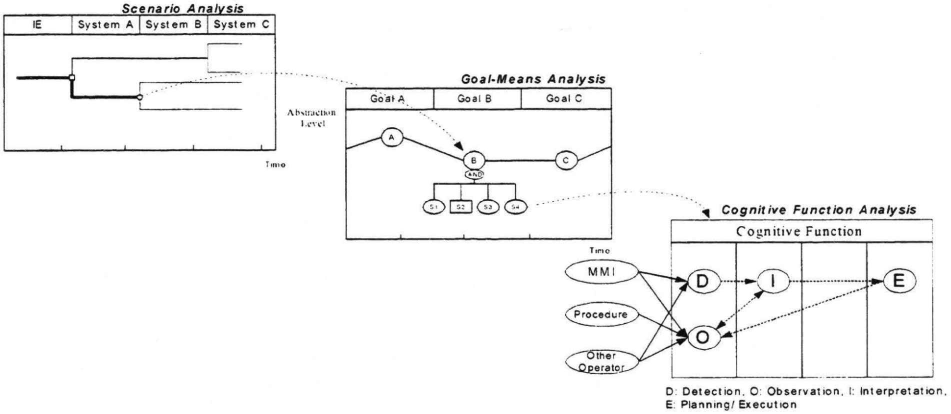
Categories	Information items
Scenario analysis	Accident history (initiating event, preceding/concurrent tasks)
	Status and trends of major parameters
	Status of major systems/components
	Allowable time
	Environment (location, clothes, tool, heat, noise, etc.)
	Plant policy/safety culture
	Team cooperation and communication
Goal-means analysis	Implied goals (attention flow)
	Task steps
	Action time
	Familiarity (experience, simulator training)
	Task characteristics (dynamic/step-by-step)
	Availability and usability of HIS (alarm, indicator, display)
	Availability and quality of procedure
	Involved operators (SRO, RO, TO, etc.)
Cognitive function analysis	Input/output information
	Related cognitive stages
	Pattern of cognitive function

STRUCTURED INFORMATION ANLAYSIS FOR HRA

Even though task analysis is an essential step of the HRA processes, HRA methodologies including newly developed ones do not explicitly provide the technique for task analysis itself. It is generally the analyst's mission to choose an appropriate task analysis technique among the various ones that were originally developed for purposes other than HRA. An exception is the framework of task analysis suggested by Kirwan (1998) for conducting HRA. However, the framework was designed for conventional HRAs, with no provision of an information collection technique for cognitive error analysis, one of the main concerns of the 2nd generation HRA.

The level of information on the task structure and context varies. There is the macroscopic contextual information such as accident scenarios, and on the other hand, the microscopic cognitive information such as cognitive stage and input/output information. In this paper, in order to systematically collect and analyze such various levels of information, we divided all of the relevant information into three parts considering the meaning, interrelationship and abstraction level of information. An integrative and schematic structure for information analysis was developed for each part. The first part of SIA is scenario analysis, the second is goal-means analysis representing the structural relation between the task goal and steps, and the third is cognitive function analysis representing the cognitive pattern and information flow. Using those three analyses, a systematic observation is undertaken from the macroscopic viewpoint of the task to the microscopic standpoint on the specific cognitive processes.

Figure 2: The Framework of SIA and Three Stages



Consequently through this process, the structural vulnerability of the given task can be identified and a more concrete foundation on the objective assessment of the related factors can be made with resolving strategies for such vulnerability. Figure 2 shows the three stages of SIA and the interrelationship among them.

Scenario analysis stage

The scenario analysis shown in Figure 2 is a process of collecting and integrating the macroscopic contextual information on the given task. The accident progression of the NPPs varies according to the response of the relevant systems and operators after a reactor trip, and, even for the same task, the situation such as the previous and concurrent task, and the status of the systems or components can be different by scenario. In addition, within a single scenario, some error events can take place sequentially and there may exist interdependence between error events. To predict error types and possibilities, the analysts should basically understand the situational context around the given task. From this scenario analysis, the following information can be identified, i.e., the initiating event taken place, operating status of the safety systems, and previous or concurrent tasks. The time in which the operator perceived the sign for the task and the available time for the task completion are examined in this analysis to estimate the allowable diagnosis time for the given task.

Goal-means analysis stage

The goal-means analysis shown in Figure 2 is the schematic representation of the information related to the goal-means of the task. This is the application of Rasmussen's (1986) abstraction paradigm of knowledge to task structure analysis. The figure represents the task steps centering on the abstraction level of knowledge and the interconnectivity. It specifies the task attributes and requirements in a top-down fashion. In the goal-means analysis, the collected information contains the following: task goal set by the operator's cognitive process, task steps and related abstraction level, logical relations between the task goal and steps, the operator's attention flow preceding and following the given task, etc. Besides the structural pattern of the task, conventional HRA PIFs are assessed in goal-means analysis. Such PIFs include each task step and corresponding responsible person, execution time, task familiarity, task characteristics, availability and quality of procedures, and quality of human system interaction (HSI).

Cognitive function analysis stage

The cognitive function analysis takes charge of the most microscopic information related to the cognitive function for a given task step. For each task step identified in the goal-means analysis, the cognitive function is analysed as necessary. The cognitive model used in this study consists of five cognitive stages: detection, observation, state identification, goal setting, and planning/execution. According to the type of each task step, the pattern of cognitive functions is different, and so are the related cognitive stages. The analysts navigate the operator's cognitive process and the information flow for doing a task step. Through this analysis, the cognitive characteristics with cognitive patterns can be identified. If the task type is skill-based, the cognition flow is directly moved to the execution stage from detection. However, if knowledge-based, it traces all the cognitive stages. In the cognitive function analysis, input information detected and observed from the system and corresponding output from the decision-making processes, and the pattern of cognitive functions can be analysed.

APPLICATION OF SIA

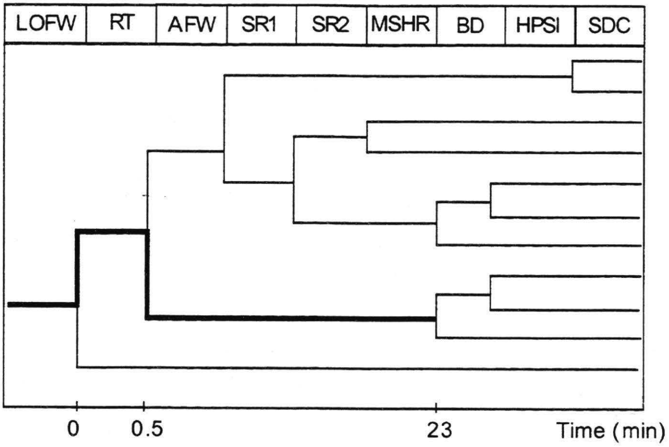
As an application of SIA, the 'bleed and feed operation' task is selected, which is one of the tasks required in emergency operating procedure (EOP) of NPP. The 'bleed and feed operation' is required for eliminating the residual core heat, in the case of the total loss of all feedwater. The task to be analysed is 'the operator identifies the total loss of all feedwater, makes the decision of the bleed and feed operation, and then initiates the reactor coolant system (RCS) bleeding by opening the safety depressurized system (SDS) valves'.

Scenario analysis of 'Bleed and Feed'

The situational condition of the 'bleed and feed operation' is that the reactor trip was successful, but the auxiliary feedwater system (AFWS) has failed to operate. The scenario associated with this situation is represented with a bold line in the scenario analysis in Figure 3. The collected information in the scenario analysis for the given task is as follows:

- Initial event and related scenario: loss of main feedwater * reactor trip * loss of auxiliary feedwater;
- related procedures: EOP-01, EOP-02, EOP-05, and FRP (functional recovery procedure)-06;
- previous/concurrent tasks: manual operation and recovery of AFWS;
- status and trend of major operating parameters:
    - o Pressurizer : pressure(↑↑), level(–);
    - o Steam generator: pressure (–), level (↓↓↓);
    - o RCS average temperature (–);
- status of safety system: AFWS is unavailable;
- timing and available time: 10~15 min. after reactor trip, available time is 8~13min;
- task location and working environment: MCR, adequate;
- level of safety culture/management: the task is educated and exercised in a simulator biannually.

Figure 3: Scenario analysis of 'bleed and feed'



Goal-means analysis of 'Bleed and Feed'

In the goal-means analysis, a detailed analysis on the task itself is performed. As shown in Table 2, the 'bleed and feed operation' task contains the task step 6 and 7 of EOP-05 and FRP-06, HR-04, but situation assessment and decision making of the task are directly associated with the steps 6 and 7 of EOP-05. If the detailed task steps are identified, then the task goal is analysed and the logical relationship between the goal and the task steps is represented (refer to Figure 4). The collected information in the goal-means analysis can be summarized as follows:

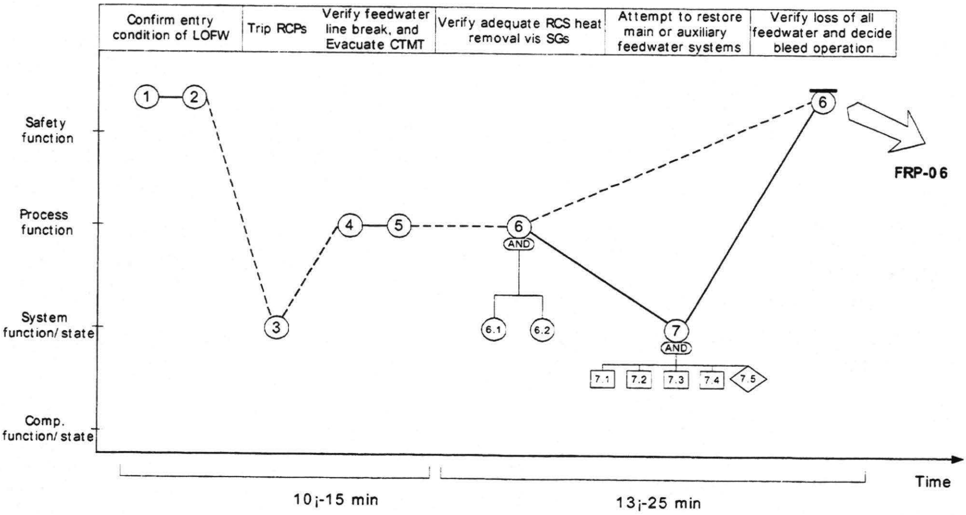
- Task goal: assessment of loss of all feedwater, decision on the bleed and feed operation;
- task steps and responsible operator: EOP-05, steps 6 and 7. SRO/RO/TO;
- expected response time: 13~25 min;
- availability and quality of procedure: symptom-based and event-based well-designed procedure;
- level of MMI: appropriate (indicators/alarms)
- task characteristics: team task, decision task with priority problem.
- Task familiarity: relatively well known, but rare experience;
- Expectation: doubt as to the spurious signal on the AFWS during the early phase.

The important task step in the 'bleed and feed operation' task can be considered 'EOP-05 step 6', in which operators recognise the necessity of the 'bleed and feed operation' and make a decision on the transfer to the recovery procedure. At this point, there may be a time delay in decision making, since SRO can face difficulty in setting priority between two goals, i.e. the economic loss and the plant safety.

Table 2: Procedural steps of the task for "bleed and feed"

EOP-05 (Loss of all feedwater)		
Step	Instructions	Contingency Actions
5	Evacuate the containment	
6*	Verify adequate RCS heat removal via the steam generators by checking ALL of the following:	
6.1	Steam generator has wide range level greater than or equal to 2% in at least one SG OR being restored by a total feedwater flow greater than 35 L/sec to one or both SGs AND	IF both SG wide range level are less than 2% AND the total feedwater flow is NOT being restored greater than 35 L/sec to one or both SGs, THEN implement the FRP-06, HR-04 OR
6.2	RCS Tc temperature do NOT increase by more than 10°F (5.6 °C)	IF RCS Tc temperature increase uncontrollably by 10°F (5.6 °C) or greater, THEN implement the FRP-06, HR-04
7*	Attempt to restore main or auxiliary feedwater systems to operation by ensuring ALL of the following:	
7.1	*1.1.1.1.1 ...*	
-		
FRP-06, HR-04 (bleed and feed operation)		
Step	Instructions	Contingency Actions
1	Confirm that bleed and feed operation is required by ANY of the following:	

Figure 4: Goal-means analysis of 'bleed and feed'

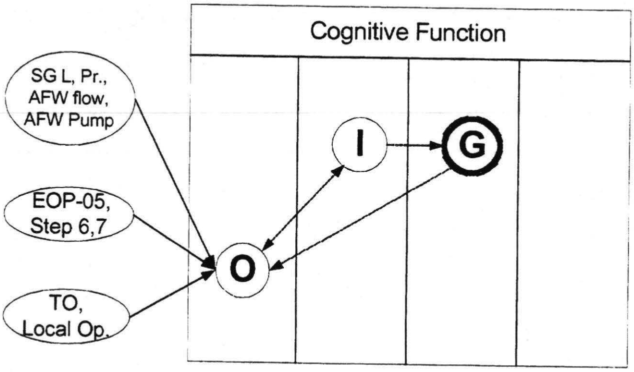


Cognitive function analysis of 'EOP-05, step6'

In the 'bleed and feed operation' task, 'EOP-05, step 6', in which the transfer to the recovery procedure is decided, a higher level of cognitive process is required. In step 6, operators check the possibility of RCS heat removal through steam generators, and, if it is not possible, finally make a decision on the transfer to the recovery procedure. When we apply the simplified cognitive model described above, the observation (O), state identification (I), and goal setting (G) stages are highly associated with step 6.

The input information related to the task in step 6 is obtained from various channels such as other operators or instrumentation. Such information includes the steam generator water level, the flow rate of auxiliary feedwater, and the status of the main and auxiliary feedwater pumps. Another information channel is the procedure that states the relevant emergency task. On the basis of the delivered or observed information, the SRO should make state identification and a decision on the priority between the transfer to the recovery procedure and the recovery of AFWS. Figure 5 shows the schematic representation of the cognitive flow and pattern, and the important cognitive stages and related input/output information associated with 'EOP-05, step 6'. Based on the analysis, step 6 can be summarized in that the stage of goal setting (G) stage is judged to be a higher burden to the operators. If the decision for the 'bleed and feed operation' is not made promptly due to the recovery of the auxiliary feedwater system, there is a high possibility, for the scenario stated above, that the task would not be completed within the available time.

Figure 5: Cognitive function analysis of step 6 conclusion



O: Observation, I: Interpretation, G: Goal Setting

Human reliability analysis (HRA) is conducted to predict the possible error types and their probabilities on given tasks with specific contexts. The power of this prediction is most importantly determined by the analyst's knowledge on the tasks and their contexts. However, most of the HRA methodologies, including the newly developed ones, focus on the provision of error models and taxnomies while leaving the consideration of contextual information mostly to the hands of the analyst.

This paper suggested SIA as a task analysis method for HRA to assist HRA analysts in collecting and structuring such information on tasks and contexts. The method emphasises the contextual information as well as the cognitive process of tasks. The SIA consists of three stages: the scenario analysis, the goal-means analysis, and the cognitive function analysis. The SIA does not simply help the individual assessment of the information related to the task steps or PIFs, but more importantly offers a systems-engineering view on the task structure. While the PIFs assessment in HRA has been a one-dimensional and enumerative factor analysis, the SIA framework allows analysts to view the problem structure stereoscopically considering information from the macroscopic contextual level to microscopic cognitive level. The result can be graphically represented to show the structural organisation of a task. This enables the analyst to interpret the results of the task analysis intuitively and to get insights on the types and possibility of potential error. It should be possible and desirable to combine the SIA with the emerging advanced HRA methods such as CREAM and ATHEANA for a better prediction of cognitive errors.

Since the multi-stage analytic process of SIA may be overly laborious, we are currently developing a computerised support system to mitigate the efficiency problem. The method is currently being applied to the analysis of the emergency operational tasks of NPPs.

ACKNOWLEDGEMENTS

This research was supported by "The Mid- and Long Term Nuclear R&D Program" of MOST (Ministry of Science and Technology), Korea.

REFERENCES

Apostolakis, G. E., Mancini, G. and Van Otterloo, R. W. (Eds.) (1990) Special issue on human reliability analysis. Reliability Engineering and System Safety 29, 283-385.

Hollnagel, E.(1998) CREAM - Cognitive Reliability and Error Analysis Method. Elsevier Science Publishers.

Kim, J.W., Jung, W. (1999) Proposal of performance influencing factors for the Assessment of human error under accident management situations. Proceedings of International Conference of PSA. Washington DC.

Kirwan, B. (1997) The development of a nuclear chemical plant human reliability management approach: HRMS and JHEDI. Reliability Engineering and System Safety 56, 107-133.

Kirwan, B. (1998) Human error identification techniques for risk assessment of high risk system – part2: towards a framework approach. Applied Ergonomics 29, 229-319.

Pierre, L. B., Francesco, C. and Corinne, B. (1999) MERMOS, a second generation HRA method: What it does and doesn't do. Proceedings of International Conference of PSA. Washington DC.

Rasmussen, J. (1986) Information processing and human-machine interaction: an approach to cognitive engineering. North-Holland. New York.

USNRC(1998) Technical Basis and Implementation Guidelines for a Technique for Human Event Analysis (ATHEANA). Draft Report for Comment. NUREG-1624.

Yoon, W.C., Lee, Y.H. and Kim, Y.S. (1996) A model-based and computer-aided approach to analysis of human errors in nuclear power plants. Reliability Engineering and System Safety 51, 43-52.

SPEED: a model for the Task Analysis Process

Gabrielle de Brito
EURISCO
European Institute of Cognitive Sciences and Engineering
4, avenue Edouard Belin, 31400 Toulouse

Université René Descartes, Laboratoire d'ergonomie informatique de Paris V, Rue des Saints-Pères,
75006 Paris
Gabrielle.Debrito@onecert.fr
http://www-eurisco.onecert.fr

ABSTRACT

This article presents a model (SPEED) for analysing the cognitive task involved in the use of written procedures. There are seven distinct steps in this task, each involving a different activity. In each of these stages aviation examples are given.

Keywords

Procedure-following, cognitive task analysis, operator's cognitive activity.

INTRODUCTION

Actors such as pilots, designers and manager cooperate in enhancing flight safety. The introduction and evolution of new technology has lead operators to use increasingly complex, automated and abstract systems (Amalberti & Wioland, 1997). In order to improve safety and help operators cope with this growing level of complexity, and to prevent errors in human-machine interaction, information aids have been developed. They provide guidelines for task execution in specific contexts. This has resulted in the production of a large amount of written procedures and instructions in the work place. Elaborate proceduralisation is an essential characteristic of complex systems such as the aeronautical domain and civil aviation in particular. In the United States of America, Degani & Wiener (1990) focused their study on how pilots use procedures (check-lists) in normal flight situations. We have completed this study by observing the use of procedures (do-lists) in abnormal and emergency situations while carrying out analysis of pilot's cognitive activity during procedure following6. We first provide a brief status of written procedures used in aeronautics, followed by an explanation of the methodology used to carry out our study. In the last part, we shall present SPEED, a model that can be used to make a psycho-cognitive analysis of the task of the use of written procedures.

THE STATUS OF WRITTEN PROCEDURES IN AERONAUTICS

Check-lists and do-lists used in new generation aircraft (type Airbus A320, Boeing 777, McDonnel-Douglas MD11) have different, but nonetheless complementary status. They can be considered as (i) an aide (a reminder for relevant actions, diagnosis assistance, and an aide for training), (ii) a guide for precise and officially established action aiming to provide necessary information for carrying out a task, (iii) an instruction that must be respected, (iv) an instrument for coordination between crew members, technical systems and the company and finally, they can be seen as (v) a legal

---

[6] The study reported in this paper was financed by the French Civil Aviation Authority (DGAC). Complete results of this analysis can be found in three EURISCO reports (Karsenty, Bigot, & de Brito, 1995), (de Brito, Pinet, & Boy, 1998).

reference in the event of an accident. Considering these various statuses, we are faced with an unresolved dilemma:

- On one hand, procedures are necessary, because they perform crucial functions of codification and transmission of knowledge, formalisation of good practice, and quality assurance. Pilots must respect them literally in order to take the right actions;
- on the other hand, considering their incompleteness (non exhaustive, lacking the degree of implicitness), they cannot ensure alone the reliability of all socio-technical systems (Roth, Bennett, & Woods, 1987). Pilots are thus obliged to adapt them, or to transform them according to the current situation.

This dilemma incited us to look closer at existing deviations between the task prescribed by designers and the activity effectively carried out by pilots during instruction following.

METHODOLOGY

The theory behind this study is that at least some deviations from procedures can be explained by pilots' cognitive activity and the cognitive needs that this gives rise to. This section draws up a conceptual framework to characterise the cognitive activity involved in following written procedures, and to develop more precise theories on the link between cognitive functions and different modes of execution of written procedures.

Task Analysis and pre-analysis of possible deviations

We started by analysing the totality of instructions linked to the use of a procedure, and identified in parallel the totality of possible deviations linked to these instructions. This analysis served as a support for elaborating a questionnaire and an observation grid.

The questionnaire

The questionnaire contains 35 questions. The majority of questions consisted of (i) questioning pilots on certain deviations, (ii) justifying replies given. Where possible, closed questions or multiple choice were used to facilitate the reply. Ten companies using new generation Airbus were identified. Of 606 questionnaires sent, 227 were returned completed.

Detailed comment analysis

The detailed analysis of replies provided useful qualitative data. A systematic collection of expressed ideas was undertaken question by question, in order to gather all ideas expressed on a given question.

Simulator Observations

All our observations were undertaken in mobile flight simulators (Full Flight Simulators) at the training centre "Airbus Training" in Toulouse, with pilots in recurrent training (line pilots who return to the training centre twice a year for training in specific flight situations).

These observations (35 sessions representing approximately 140 hours) do not aim to represent the totality of pilot activity in the cockpit. The objective is to show how procedures are used by pilots.

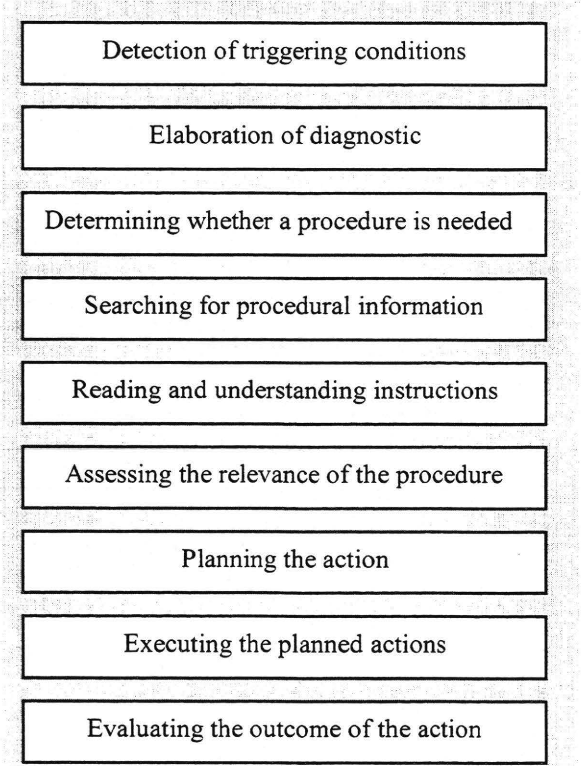
SPEED: A COGNITIVE MODEL

Following a procedural document in order to carry out a task involves correctly performing a series of actions. To transform the information of the procedure into actions, the operator uses jointly, sequentially and/or selectively various resources: systems knowledge, document characteristics,

habits, etc. However, it is very difficult to distinguish and to evaluate the different cognitive levels involved in instruction following because these various levels are interdependent.
In order to identify the different cognitive processes used by the operators, we built a psycho-cognitive model of the activity of following written procedures. This model called SPEED (Suivi de Porcédures Ecrites dans les Environnements Dynamiques: Written Procedures Following in Dynamic Environments) consists of nine different stages corresponding to stages involved in the use of procedures.

The nine stages of SPEED7 are presented in figure 1.

Figure 1: Steps of instruction following



The sequential display of the model can evolve according to several factors such as the environmental context, the evolution of the system, time available, internal/external interruptions, the degree of expertise of the operator, etc. Not only does it aim at improving procedures or the prescription of operator behaviour, but it also attempts to point out the main factors accounting for the ineffectiveness of certain written procedures by taking into account sociotechnical aspects of the use of instructions, the constraints of use linked to the task to be carried out, specificities and characteristics of operators, and interactions with systems.

---

[7] A more detailled version is presented in de Brito, (1999) and in Veyrac & de Brito (1999)

Detection of triggering conditions

Factors which determine the use of instructions are numerous (Wright, Creighton, & Threlfall, 1982). In the case of abnormal situations, it is often the detection of a change of status by an alarm that triggers the use of procedures. Alarms (although they do not actually guide actions) predetermine the start of procedures to resolve a current failure. They must be detected and interpreted quickly. This has lead to numerous studies centred on the cognitive aspects of the interpretation of alarms (Stanton, 1994; Edworthy & Adams, 1996; Wickens 1992; and Hale, 1990). Chatty, Athènes, & Bustico (1999) point out that the main objective of an alarm is awareness and identification of an incident. An alarm has thus two functions: to attract the operator's attention and to provide him/her with information that will incite a behaviour modification (Wogalter, 1994).

For abnormal and emergency situations, a deterioration of the system is announced by alarms (sound and/or visual) and a computerised message that informs the pilot about the nature of the failure and lists the actions to be taken. Because pilots are usually assisted during this phase, certain problems may appear:

- When a failure unforeseen by the system occurs, 43% of pilots admit to having difficulty in immediately detecting that they are in a problem situation;
- when interfaces present information and/or conflicting signals, pilots have trouble evaluating information to be taken into account and must analyse the situation;
- when alarms are masked or drowned out by other stimuli;
- when there are false alarms or when signals do not mean anything to the pilots.

The detection of indicators allowing the evaluation of a situation depends above all on the operators' attention. But stress, workload, personal factors (e.g. tiredness) reduce attention and can lead to a failure to take the alarm into account. The repetition of certain alarms may lead pilots to neglect related messages. The more the operators are familiar with situations, the less they perceive them as being dangerous (Wogalter, 1994).

Elaboration of diagnosis

In new generation aircraft, the objective of procedures is to avoid the diagnosis activity by directly prescribing what pilots have to do in a given situation. However, our study showed that pilots perform this activity nevertheless, whether before, during or after the application of a procedure. The questionnaire allowed us to identify two cases: (1) if the pilot has time, he/she will try to understand the dysfunction before applying the procedure; 75% of the pilots admit that they perform a diagnosis before executing the current do-list (2) if the pilot is in an emergency situation, he/she first applies the procedure, but while applying it, will try to understand the dysfunction. Moreover, it is the execution of the first actions described in the procedure that will help him/her to understand the current situation by comparing his/her own action plan with that presented to him/her.

Diagnosis is thus influenced by situation awareness which allows pilots to understand a given situation. A large part depends on:

- The perception that the operator has of his/her working environment;
- the operator 's objectives;
- the pilot's expectations on how the systems work based on his/her knowledge and experience. Certain authors (e.g. Hoc, 1987) have shown that the operator gives preference to information that confirms his initial diagnosis than to information that does not;
- characteristics of the situation: time pressure or estimation of incurred risks;

- what the operator can do and knows how to do. In the case of rapid processes, the operator gives precedence to diagnoses that allow him/her to envisage at once actions for handling the situation according to relevant (Amalberti & Valot, 1990).

Problems of non-application or difficulties of application of written procedures can also be explained by the divergence between the pilot's diagnosis and that of the engineer, which is implicit in the procedure.

Determining whether a procedure is needed

In a given situation, unless the operator has a clear understanding of his/her competence and the capabilities of the technical system, he/she may fail to see the need for instructions. This depends on the operator's assessment of his/her know-how and competence. These assessments may be overestimated, correctly estimated, or underestimated. In the last case, the operator is overconfident and unaware of the need for information. This is often the case when action feedback is deferred.

The decision to refer to written procedures depends on three factors:

- Urgency of the situation: if the operator feels that an immediate solution is not required, he/she may proceed by trial-and-error;
- usefulness for the situation: reverting to written procedures seems to be connected to the complexity of the task and the operator's experience. The less the operator perceives the task to be difficult, the less he/she will judge assistance to be necessary. Experienced pilots feel less and less the need to use written procedures;
- the estimated risk of the situation: the more the situation is perceived as dangerous, the higher the probability that the operator will read the accompanying instructions. Wogalter & Young (1994) complete these results by pointing out that the more operators are familiar with a situation encountered, the less they perceive them as being dangerous.

Succeeding without assistance can provide a means of self-appraisal and may even represent a challenge that operators set for themselves. Indeed, some pilots (4 %) admit to perceiving the challenge "do it alone" as a means of testing their individual capacities in problem situations. Palmer (1995) and Degani & Wiener (1994) cite similar cases.

Pilots may also see the use of procedures as a means of confirming their own evaluation, when they are not sure of a procedure, or when they have made errors (Szlichcinski, 1980).

In an emergency situation, 11% of pilots say they do not waste time looking for a procedure if they can not find it quickly. This evaluation depends on the benefit provided by the procedure compared to the difficulty of putting it into operation. If there is nothing to be gained by using the procedure, its assistance value is non-existent.

Searching for the procedure

Easy access to information is a prerequisite to using procedures. To find the right instructions, operators must determine in which category the current situation or incident belongs and relate it to the classification system of the procedures. This is done with elements such as the table of contents, use of titles or indexations to hypertext links, etc. According to (Heurley, 1994), how an operator looks for information depends also on how it is used. If it concerns a new document, the operator orients his search, otherwise he goes directly to the required information. The repeated use of a procedure tends to entail a certain automation of the corresponding activity, and can produce a decline of active validation strategies (Gersick et Hackman 1990, quoted by Leplat, 1998).

In Airbus aircraft, an automatic mode in real time allows the pilot to immediately begin the execution of instructions without having to localise the procedure he/she needs. 68% of pilots find the access to do-lists "very satisfactory" or "satisfactory". These results are directly linked to the difficulties that 26% of pilots encounter in finding do-lists that do not appear automatically:

- When the incident is not foreseen by the automatic system;
- when the failed system does not have sensors directly connected to the ECAM;
- in all situations where procedures are accessible only at the request of the operator;
- when the electronic checklist refers to a paper checklist, which is a problem for 52% of pilots. Pilots point out that in stressful situations this difficulty may lead to not finding the do-list or to selecting an incorrect one. Added to this is the subsequent difficulty connected to following paper procedures: the risk of omitting certain steps, inaccurate completion, can only be seen by one crew member at a time, etc.

Reading and understanding instructions

According to linguistics, we adopt here a psychological point of view considering that the understanding of a text is a twofold process (Kintsch & vanDijk, 1978):

- A stage of literal understanding which is a representation of the meaning of the text. This "surface understanding", consists in making sense of instructions from the sense of the terms. This first stage, refers to the intelligibility of a procedure. This is where the syntaxico-semantic aspects, the morpho-dispositionnels aspects (e.g. use of diagrams), the lexical and semantic aspects, and typo- dispositionnel (e.g. layout and formatting of text) aspects of written procedures should be taken into account;
- A stage of contextualised understanding, or appropriation. Understanding a procedure is assimilating a coherent representation of the action more or less explicitly described. Contextualised understanding allows thus to envisage an action with regard to the representation of the situation, including the systems status, other procedures in use, crew availability, etc.

We can say that an action is "always situated" (Suchman, 1987) and thus, in a certain way, is unique, whereas the text of the procedure is intended to cover the majority of cases. But problem solving consists in defining a situation according to the interpretation of the relevant elements of the situation. Operators must understand the situation and the procedure to be able to apply written instructions. About 75% of pilots have already questioned prescribed actions. The main reason given (66% of these pilots) is not understanding the instructions. These results confirm and complete those of Wright (1981) who puts forward four main factors linked to not understanding:

- The use of terms which have the different meaning for users, i.e. acronyms and abbreviations;
- the non exhaustive nature of written procedures: the absence of justification of actions to be taken does not always enable pilots to understand a procedure (14%), current situation (9%), evolution of the status of the system (7%), even more so when the procedure is complex or when there are several simultaneous failures or interactions between systems;
- an unsuitable format of presentation. 27% of pilots say the ECAM is not organised according to operational requirements;
- the presence of ambiguities or errors between the different formats of presentation;
- It is almost impossible to decide on the right level of information of instructions, since this depends on each pilot, and evolves according to his/her experience. Ideally, the design of adaptable, multi-level procedures would make it possible to obtain different levels of detail, in particular, an experienced operator could ignore information that does not interest him/her.

Assessing the relevance of the procedure

The relevance of a procedure is the capacity it gives the operator to resolve a given situation. When the consequences of certain actions are not easy to assess, or when they are known only after the action has been carried out, the action's relevance, and thus following the procedure itself may be questioned (Hale, 1990; Battman & Klumb, 1993).

Assessing relevance also depends on what procedures the pilot expects. If the contents of the written procedures deviate too far from what the pilot expects, or his/her view of what the task should involve, 21% say they simply refuse to apply them.
Evaluating the cost required for the application of an instruction in a given situation (Heurley, 1994), also taken into when assessing the relevance depends on:

- The degree of constraint linked to the procedure Zeitlin (1994). 12% of pilots do not consider it important to follow literally instructions relative to conditions considered as reversible or unimportant;
- the possible degree of conflict between several procedures;
- the institutional consequences of the (non) application application of the procedure;
- the importance of the information: 26% of pilots select certain items and 41% group together items presented and carry out only actions considered as most relevant. So it is necessary to restrict procedures to the bare minimum: only information indispensable to the resolution of the problem should appear.

Action planning

- Planning actions in a dynamic environment means managing not only actions but also the surrounding conditions (Suchman, 1987). Consequently, it is necessary to priorize every action, according to an evaluation of risk and degree of urgency of foreseen actions. The term "planning" refers to three requirements of pilots work:
  - Managing the most opportune moment to apply procedures;
  - managing simultaneous tasks: 49% of pilots admit they execute check lists early, 57% admit they postpone checklists, depending on the operational situation;
  - managing available resources to execute procedures (sensory channels, attention capacity, mobilizable skills, etc.) according to the availability of each crew member. The last requirement leads 45% of pilots to postpone the execution of checklists, and 52% of do-lists if one of the pilots is temporarily unavailable.

Written procedures attempt to describe actions as precisely as possible. Consequently, adapting to real situations is poorly tolerated, even though unavoidable. Real pilot expertise lies as much in action planning as in adapting actions to a given situation in real time.

Executing planned actions

During the execution of a procedure, every instruction must be carried out. The execution of instructions is described as a process of implementation of the mental schemata set up during the planning phase. This supposes that the operator is capable of remembering the situation model elaborated during the understanding phase for a sufficiently long period of time.

Our study has shown that one of the main problems linked to the execution of the normal checklists is the omission of some items. Indeed, 88% of pilots admit to having already forgotten an item when they execute a checklist. The three main reasons are: interruptions and distractions that they cause, time pressure, and fatigue. Pilots point out that it is often the combination of these factors (time pressure + interruption, time pressure + fatigue, etc.) that causes the omission. In our study, it appeared that these omissions were more particularly linked to the nature of the interruption. Indeed, pilots may forget to complete checklists after an interruption by the Air Traffic Control (68% of cases) and after the occurrence of an alert message on the ECAM (89% of cases). The routine aspect of checklists and their conditions of execution are partly responsible for loss of control of the activity. This can also be at the origin of omissions or incorrect verifications (ex.: announcing a value as right although the real value is wrong).

Incorrect execution or non execution of instructions can result in partial treatment of the procedure: instructions are read, understood but are not executed (Fischer, Orasanu, & Wich, 1995); (Wogalter & Barlow, 1990). Between the intention to carry out an action and its execution, "breakdowns" can intervene: habits, automatisms, and interferences can interrupt an intended action. Furthermore, errors of interference (Reason, 1990) can disrupt the execution of a procedure by provoking mix ups or reversals. A mix up occurs when an action must be executed while the operator is already occupied with a previous one. The new action is carried out as if it was the previous one (example: press on all flashing buttons and verify that button A is ON = the operator mixes up the two actions and presses on button A): the recurrent influence of an action over the subsequent action. Reversals occur when several actions carried out simultaneously become mixed up. For example, when operators are required to push flashing buttons and to pull out those that are lit, and the operator does the opposite: a reversal of actions.

Evaluating the outcome of the action

The process of evaluating the outcome of an action, is a form of after-the-fact control: it is situated after the execution and provides the cognitive system with feedback that allows the operator to evaluate the success or failure of actions and, if necessary, to reorientate the activity. Procedures allow the evaluation of actions carried out with regard to actions not yet done. The first evaluation is that of the deviation from the goal to be obtained. If the evaluation of the deviation from the goal is a means of detecting an error or an incompletion of the procedure, it is not always sufficient to correct the error or to continue the procedure. A second evaluation is the recognition of critical situations where the indications point to an erroneous action. These situations can lead to the analysis of the action, aiming to identify errors, and to questionning the representation of the actions which were at the origin of the error.

Conclusion

We introduced this work by explaining that we need to know about the process of following (or not) written procedures in order to improve our knowledge of the operator/instruction system in emergency situations, to identify the factors that affect it, and to provide a structured representation of it. To do this, we used a model based on the activity of the pilots engaged in following written procedures.

For each step of the activity of following procedures, we presented various factors that might lead pilots to not comply with the prescriptions as envisaged. Accordingly, we showed that pilots encountered problems when searching for procedures that do not appear automatically on the system display. The same problem occurs when the ECAM becomes crowded with information of differing nature and importance: according to pilots, finding relevant information for the situation in progress can be complex and difficult.

Pilots' knowledge, expertise, and know-how significantly influence the following of written procedures. These cognitive functions enable them to evaluate a given situation, to categorize information presented, to evaluate the relevance and the feasibility of information presented, to plan and to execute adequate actions at the proper time.

The analyses presented in this paper show that several different factors are involved in following written procedures. The use of SPEED has also allowed us to show that instructions associated to the use of procedures are not always adapted to the current situation, in particular emergency situations.

REFERENCES

Amalberti, R., & Valot, C. (1990). Champ de validité, pour une population de pilotes, de l'expertise de l'un d'entre eux. Le travail humain, 53(4).

Amalberti, R., & Wioland, L. (1997). Human error in aviation. A. safety (Ed.), International aviation safety conference , (IASC-97) (pp. 91-108), Rotterdam airport the netherlands, Suekka.

Battman, W., & Klumb, P. (1993). Behavioural economics and compliance with safety regulations. Safety Sciences, 16, 35-46.

Brito de, G. (1999). Practices in Procedure Following: A First Classification, Human Error, Safety & System Development (HESSD'99), Liège, Belgium: 7-8 June 1999.

Brito de, G., Pinet, J., & Boy, G. (1998). Etude SFACT sur l'utilisation de la documentation opérationnelle dans les cockpits de nouvelle génération. Etude de l'interface Homme-Machine (n° T-98-054). Toulouse, France. : EURISCO Contactor Report.

Chatty, S., Athènes, S., & Bustico, A. (1999). Votre attention s'il vous plait ! Eléments d'un espace de conception de signaux visuels. C. Editions (Ed.), Interaction homme Machine (IHM'99) (pp. 17-24), Monpellier, France: 22 au 26 Novembre 1999.

Degani, A., & Wiener, E. (1990). Human factors of flight-deck checklists: The normal checklist : NASA contractor report 177549, Moffett Field, CA: NASA Ames Reasearch Center.

Degani, A., & Wiener, E. (1994). On the design of flight-Deck Procedures: NASA-Ames Research Center 177642. Moffett Field, CA: NASA Ames Reasearch Center.

Edworthy, J., & Adams, A. (1996). Warning Design. A research prospective. London: Taylor & Francis.

Fischer, U., Orasanu, J., & Wich, M. (1995). Expert pilots' perceptions of problem situations, 8 th - International Symposium on Aviation Psychologie (pp. 777-782), Columbus: Ohio State University: April 24-27.

Hale, A. R. (1990). Safety rules OK? Possibilities and limitations in behavioural safety strategies. Journal of Occupational Accidents, 12, 3-20.

Heurley, L. (1994). Traitement de textes procéduraux. Etude de psycholinguistisque cognitive des processus de production et de compréhension chez des adultes non experts, Thèse de doctorat, Université de Bourgogne.

Hoc, J. M. (1987). Psychologie cognitive de la planification. Grenoble: PUG.

Karsenty, L., Bigot, V., & de Brito, G. (1995). L'utilisation des procédures écrites dans un avion de nouvelle génération (T-95-023). Toulouse, France. : EURISCO Report.

Kintsch, W., & vanDijk, T. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Leplat, J. (1998). A propos des procédures. Performances humaines & techniques, mai-juin 1998 n°94, 6-15.

Reason, J. (1990). Human error. Cambridge.: Cambridge University Press.

Palmer, E. (1995). "Oops, it didn't arm" A case study of two automation surprises, 8 th - International Symposium on Aviation Psychologie, Columbus: Ohio State University: April 24-27.

Roth, E. M., Bennett, K. B., & Woods, D. D. (1987). Human interaction with an "intelligent" machine. International Journal of Man-Machine Studies, 27, 479-525.

Stanton, N. (1994). Human Factors in Alarm Design. London: Taylor & Francis Ltd.

Suchman, L. (1987). Plans and situated actions. Cambridge: Cambridge University Press.

Szlichcinski, K. P. (1980). The syntax of pictorial instructions. In P. Kolers, M. Wrolstad, & H. Bouma (Eds.), Processing of visible

Veyrac, H., & de Brito, G. (1999). Task analysis of instructions following in work situation. submitted at Safety Sciences.

Wickens, C. (1992). Engineering psychology and human performance: Harper collins publishers Inc.

Wogalter, M. (1994). Factors influencing the effectiveness of warnings, Public Graphics, Lunteren, The Netherlands: 26-30 September 1994.

Wogalter, M., & Barlow, T. (1990). Injury severity and likehood in warnings, Human Factors Society 34 th Annual Meeting (pp. 580-583), Santa Monica CA: Human Factors Society.

Wogalter, M. S., & Young, S. L. (1994). The effect of alternative product-label design on warning compliance. Applied Ergonomics, 25(1), 53-57. 1105

Wright, P. (1981). "The instructions clearly state... " Can't people read ? Applied Ergonomics, 12(3), 131-141.

Wright, P., Creighton, P., & Threlfall, S. (1982). Some factors determining when instructions will be read. Ergonomics, 25(3), 225-237.

Zeitlin, L. (1994). Failure to follow Safety instructions: faulty communication or risky decisions? Human Factors, 36(172-181).

Traffic manoeuvrability and cockpit display characteristics determine whether commercial airline pilots can maintain self separation in realistic scenarios of en-route flight

Kip C. S. Smith

Department of Psychology
Kansas State University
Manhattan, KS 66503
kip@ksu.edu

P. A. Hancock

Human Factors Research Laboratory
University of Minnesota
Minneapolis, MN 55455
peter@hfrl.umn.edu

ABSTRACT

Twenty-four currently certified commercial airline pilots participated in an experiment that tested an extreme implementation of the concept of 'free flight.' Their task in the experiment was to assume complete responsibility for maintaining self-separation with other aircraft. This responsibility currently resides with air traffic control (ATC). The purpose of the experiment was to determine whether experienced commercial airline pilots working with complete information could safely assume responsibility for all decisions regarding routing and separation. The experiment manipulated two factors – traffic intent and the cockpit display of traffic information (CDTI) - in a completely randomized between subject design. Our main dependent measure was the number of 'conflicts,' violations of separation criteria, incurred. The pilots were largely unable to maintain separation. The majority of conflicts were incurred in the experimental condition in which traffic changed course without declaring their intent to change and in which the display concealed those unannounced changes. These results strongly argue against the complete and immediate transfer of responsibility for separation maintenance from air traffic controllers to commercial pilots. For an extreme vision of free flight to take wing, pilots must have access to information about changes in traffic intent that updates as such changes occur. Current CDTIs and those being tested at leading research institutions, including our own, have yet to meet this need.

Keywords

Free flight, pilot performance, self-separation, cockpit display of traffic information (CDTI).

INTRODUCTION

This paper reviews a portion of an experiment on pilot performance in 'free flight.' Free flight is defined as any change away from the current system of centralized ground-based air traffic control towards a more distributed control system that includes pilots and airline dispatchers (RTCA, 1995).

The goal of the experiment was to test whether experienced pilots would be able to maintain separation without the assistance of air traffic control. To this end, we presented realistic simulations of air traffic in the free-flight environment to experienced commercial pilots, provided them with complete and accurate information, and gave them full authority to select and implement maneuvers to maintain separation (Smith, Scallan, Knecht, and Hancock, 1998). Such a situation might arise were free flight implemented in a manner that removes the air traffic controller from the active control loop. The complete reallocation of control to the flightdeck simulated in our experiment represents an end-member in the continuum of potential implementations of free flight. This implementation of free flight was intentionally extreme in order to provide a highly rigorous test of the free flight concept.

Measures of pilot performance included the locations and velocities of all aircraft from which we calculated aircraft separation. The data presented here are the number of trials in which pilots failed to maintain separation. Comparison across treatment conditions reveals constraints on pilot performance in simulated free flight.

METHOD

Participants

Twenty-four currently-certified commercial airline pilots (with an average of 14,000+ flight hours) volunteered to participate in the experiment. They received no remuneration for their time. Six pilots participated in each of four cells in a completely randomized between-subject design.

Materials

The experimental materials were a series of 12 realistic en-route air traffic scenarios. The flightplans for all traffic in the 12 scenarios including the pilot's 'ownship[8]' were designed to challenge pilots by putting them in situations that might be encountered in free flight and that would provoke decisions to resolve impending conflicts between aircraft. The 12 scenarios took place in 12 different regions of the continental United States. The experiment was conducted in the fixed-based glass-cockpit simulator at the Human Factors Research Laboratory of the University of Minnesota.

Design

The pilot's (participant's) task was to proceed safely and efficiently along the flightplan and to maintain separation between the ownship and all traffic.

The experiment crossed two factors at two levels. The two factors were traffic intent and the traffic-proximity coding used by the CDTI in our simulator. The two levels of traffic intent were 'maintaining level flight' and 'unannounced change'. In the level-flight condition, all traffic could safely be relied upon to adhere to their flightplans and altitudes. In the unannounced-change condition, selected traffic changed headings and/or altitudes without warning. These changes were, however, fully within the reasonable parameters of the flight envelope of the aircraft described (i.e., traffic did not suddenly appear or disappear or make impossible ascents or descents). This manipulation was designed to make the intent of traffic more predictable in the level flight condition and to test the effect of unannounced changes in intent on the pilots' ability to maintain self-separation.

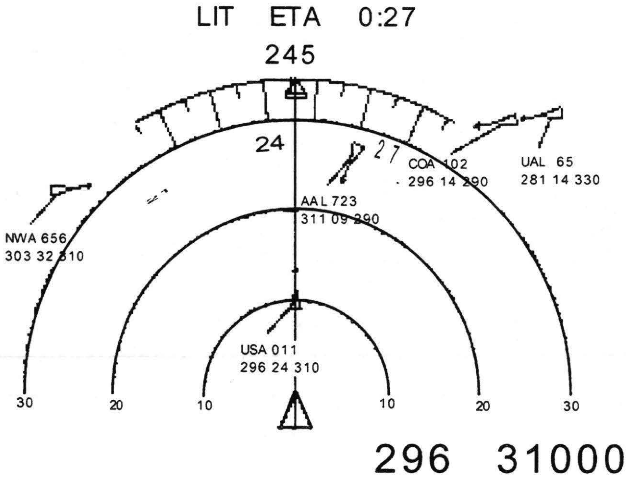
The CDTI symbology and controls and the two levels of traffic-proximity coding were patterned after the CDTI displays used by researchers at NASA (e.g., Cashion et al., 1997; Dunbar et al., 1999;

---

[8] The term 'ownship' refers to the pilot's own aircraft, that is, the aircraft under the pilot's control.

Johnson, et al., 1997) and NRL (e.g., van Gent, et al., 1997). The two levels of CDTI traffic-proximity coding were 'color' and 'black and white.' In the example shown in Figure 1, black and white have been reversed to facilitate legibility. In the color condition, the color of traffic symbology (aircraft location and datatag) on the CDTI indicated the traffic's proximity to the pilot's ownship. Aircraft not at the pilot's flight level were coded white. Aircraft at the pilot's flight level but at a distance greater than 30 nautical miles were coded blue. Yellow symbols indicated aircraft at the pilot's flight level and at a distance less than 30 miles but greater than 5 miles from the pilot's ownship. Aircraft shown in red were in technical violation of the Federal Aviation Administration (FAA) minimum criteria for aircraft separation, that is, at the pilot's flight level and at a distance less than 5 miles. In the black and white condition, the color coding for proximity was turned off; all traffic icons and datatags were shown in white. For a more complete description of the CDTI, see Scallen, Smith, and Hancock (1997).

Figure 1: The CDTI used in the experiment (with black and white reversed)



The location of the pilot's ownship is the peak of the triangle at the bottom of the display. The concentric arcs indicate distance from the ownship in nautical miles. The large numbers at the bottom of the display indicate the ownship's indicated airspeed (knots) and altitude in feet. The compass rose indicates the ownship's heading in degrees from magnetic north. The alphanumerics at the top of the display reveal that the ownship is expected to cross the LIT (Little Rock, Arkansas) waypoint in 27 minutes. Traffic is shown as triangles with intent vectors that point in the direction of travel. The alphanumerics next to the traffic icons are called 'datatags.' The first line is the aircraft's callsign; the second line indicates its indicated airspeed (knots), heading/10, and altitude/100 in feet.

Procedure

Prior to beginning each scenario, the pilots received a paper map showing the filed flightplan and the ownship's approximate en-route position at the beginning of the scenario. The experimental instructions told the pilots (1) their goal was to navigate the aircraft in a manner consistent with the 'filed' flightplan and to proceed toward their destination, (2) the flightplan information, including all waypoints, had been previously entered, into the flight management system (FMS), (3) each scenario began with the aircraft at cruise altitude and part way through the flightplan, and (4) they had full authority and responsibility to make and execute decisions for routing (altitude, heading, and airspeed) and self-separation. The only constraint placed on the pilots was that they work to maintain a minimum separation of 5 miles laterally and 1000 feet vertically. The pilots were also told that ATC was acting in a back-up role and to expect no ATC communication. We relaxed the criterion of 2000 vertical feet above FL290 (the barometric flight level corresponding to 29,000 feet) in order to discover whether pilots would seek novel flight levels, e.g., FL341. [They never did.] The pilots were also asked to end the scenario by informing the experimenter when they felt the 'coast was clear' with no source of potential conflict in sight.

Throughout the experiment, pilots had access to CDTI controls that allowed them (a) to toggle the position of the ownship between the center and the bottom of the display, (b) to change the maximum range of the display in logarithmic increments from 30 to 300 nautical miles, and (c) to suppress the alphanumeric data tags.

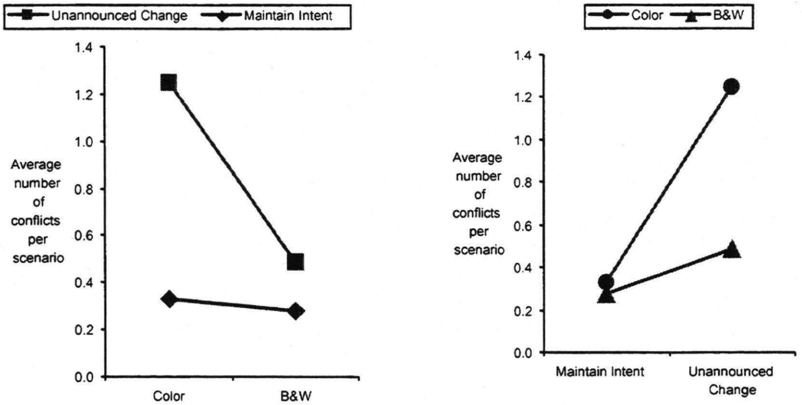
Measures

Performance data were collected as the pilots flew the traffic scenarios. Dependent measures included the locations and velocities of all aircraft. A salient piece of information that can be extracted from these data is the number of conflict situations encountered. The count of conflicts is a telling measure of success at maintaining self-separation. For a pilot to maintain self-separation, the minimum distance between his aircraft and all traffic had to exceed, at all times during the scenario, the FAA's separation criterion of five nautical miles laterally and 1000 feet vertically.

RESULTS

Each pilot encountered 12 scenarios and one 'designed conflict' in each scenario. Figure 2 shows the average number of conflicts incurred per scenario. The upper graph is meant to highlight the manipulation of traffic intent. The lower graph is meant to highlight the manipulation of CDTI proximity coding. However, both highlight the strong interaction between traffic intent and CDTI proximity coding. They reveal that the average number of errors increased dramatically in the unannounced change – color CDTI condition.

Figure 2: Graphs showing the interaction between traffic intent and CDTI proximity coding



The interaction between the unannounced change and color CDTI conditions supercedes the main effects.

In the unannounced change/color CDTI condition, the pilots committed 90 errors, 18 more than were designed into the scenarios. The presence of these 'additional' conflicts explains why the average number of conflicts per scenario exceeded 1.0. When we replayed the experimental sessions, we found that pilots who failed to maintain separation often executed maneuvers that inadvertently put the ownship in conflict with more than one aircraft. Their attempts to evade the one conflict presented by the scripted flight plan often generated multiple conflicts with other (additional, different) aircraft.

Analysis of variance found the interaction between traffic intent and CDTI proximity coding to be significant F(1, 20) = 9.18, $p \leq 0.0066$. Both main effects were found to be statistically significant. However, differences between them are attributable solely to the high number of errors incurred in the unannounced change – color CDTI condition. A Tukey multiple comparison procedure found the unannounced change – color CDTI condition to be significantly different than the other three and the others not significantly different from each other.

IMPLICATIONS

At first glance, these results appear to bode ill for pilots and for free flight. However, it says as much about the difficulty of our scenarios as it does about the pilots' abilities. We designed our scenarios to be extraordinarily challenging in order to push the pilots to the limit and to discover the constraints on their performance at maintaining self-separation.

The disparity in performance in the unannounced change/color CDTI condition is the product of a mismatch between the information needs of the pilot, our operationalization of unannounced change, and our design of the CDTI. First, all pilots indicated during the post-experimental briefing that they wanted information about traffic intent, e.g., Is AAL666 going to stay at that flight level? Or is it going to descend into my path? Second, in the unannounced change condition, aircraft changed heading and/or altitude without warning, that is, without signaling intent. From the pilots' perspective, this traffic would change course or altitude suddenly and unexpectedly. Third, our CDTI used color to highlight traffic at the pilot's flight level. Color served to focus pilots' attention on traffic within their ownship's flight level and away from traffic that might be climbing or descending into that flight level. The mismatch is the use of color to highlight less-than-critical information. The pilots taught us that the critical pieces of information in these scenarios are (a) changes in intent and (b) the likelihood of a conflict situation. By focusing on aircraft at the pilot's flight level, our CDTI did not provide the critical data.

In both the color and b/w conditions, the only cue to a vertical change in traffic intent (e.g., the begin of a turn or of a climb or descent) was the scrolling of numbers in the altitude field of an aircraft's datatag. The data shown in Figure 2 suggest that color-coding that focuses solely on the ownship's flight level overwhelmed this relatively subtle cue of changing intent. In sum, the color-coded traffic-proximity display drew the pilots' attention away from unexpected vertical moves by traffic that subsequently generated conflict situations. The reduced level of performance in the unannounced change-color CDTI condition illustrates a simple but classic lesson in human factors:

1. CDTIs for free flight must alert the pilot to changes in traffic intent such as the beginning of a climb or descent. Information about changing traffic intent must be made at least as salient as information about traffic proximity.

We do not wish to imply that there is anything intrinsically wrong with color displays per se. Rather, it may well be that a color-coded system that represents a proximity warning in both lateral and vertical separation may well be helpful. Indeed, other elements of our experimental program suggest that a display based on time-to-contact may well be an efficacious solution (Knecht and Hancock, 1999). What is clear is that the total replacement of ground-based air-traffic control is probably unadvisable at this stage. Further, the process of transition to a fully distributed decision-making system will undoubtedly take a number of years and design iterations to perfect.

In examining the ramifications of the present work it is important to look ahead to what a final system may look like. We refer to our preferred model as one of hybrid control (Hancock, 1996). Let us initially compare ground and air transportation. Commercial air transport, as we have seen, has a strong central control with only limited pilot flexibility in the case of emergency (e.g., TCAS warnings). In contrast, ground transport provides almost complete driver autonomy. We advocate a system architecture constructed at the confluence of these operational approaches. That is, strong local autonomy under stable, low-demand or routine operations and an active, strongly centralized (ATC?) control system under high demand, high density or emergency conditions. On a sunny day with uncrouded airspace, the pilot will control the skies. When constraints tighten, control will revert to centralized, normative arbiters (Smith and Hancock, 1995). (There is good reason to believe the brain also operates using a similarly hybrid architecture.)

A hybrid system promises to provide an effective compromise between the potential confusion engendered by fully distributed control and the potential overload on a single, central control node. Nevertheless, it cannot be the complete answer. What happens when a pilot (or pair of pilots) recognize an error in central control? Surely they must be free to correct such a situation. Tools and procedures to support such preemption or summary transfer of control at different points in space and time should become a primary focus of human factors research. The research will have to pay particular attention to developing an understanding of such poorly articulated yet critical areas as shared mental models and that collage of constructs collectively known as "situation awareness."

The great carrot offered by "free flight" is the draw of efficiency through the use of airspace that presently remains an untapped resource. Even in the face of a continuing increase in demand, the en-route airspace for self-navigated aircraft will remain an open blue sky for the foreseeable future. However, at or near major airports, the ascending demand promises to rapidly overwhelm available resources (both of technologic and cognitive). Constraints can tighten quickly in the terminal environment. Therefore, we should not rule out too quickly some algorithmic control and must hasten efforts to facilitate the transition to hybrid architectures.

Our findings point out the shortcomings of the current generation of CDTIs designed to facilitate the transition and point us in the direction toward which we must work. The immediate empirical issue remains: How can a CDTI be designed to overcome the unannounced change-color confound reported here? How to emphasize changes in intent without obscuring information about proximity? We continue to develop and test prototype CDTIs in an effort to answer these questions.

ACKNOWLEDGEMENTS

The study was supported by grants 93-G-048, 98-G-013, and 99-G-020 from the Office of the Chief Scientific and Technical Advisor for Human Factors, AAR-100, of the United States Federal Aviation Administration. Dr. Thomas McCloy was the technical monitor. Mr. Rickey Thomas assisted with the statistical analyses. The opinions are the authors' and should not be construed as positions taken by the FAA or any Federal Agency or employee.

REFERENCES

Cashion, P., Mackintosh, M., McGann, A., & Lozito, S. (1997). A study of commercial flight crew self-separation. Proceedings of the 16th AIAA/IEEE Digital Avionics Systems Conference, Vol. 2, 6.3-19 – 6.3-25. New Jersey: Institute of Electrical and Electronic Engineers.

Dunbar, M., Cashion, P., McGann, A., Mackintosh, M., Dulchinos, V., Jara, D., Jordan, K., & Lozito, S. (1999). Air-ground integration issues in a self-separation environment. Proceedings of the 10th International Symposium on Aviation Psychology. Columbus, OH: Ohio State University.

Hancock, P.A. (1996). On convergent technological evolution. Ergonomics in Design, 4 (1), 22-29.

Johnson, W., Battiste, V., Delzell, S., Holland, S., Belcher, S., & Jordan, K. (1997). Development of an initial cockpit situational display for free flight: The AMES display. Proceedings of the 1997 SAE/AIAA World Aviation Congress. Moffett Field, CA: NASA Ames Research Center.

Knecht, W., & Hancock, P.A. (1999). Separation maintenance in high stress free flight using a time-to-contact-based cockpit display of traffic information. Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomic Society, 16-20.

RTCA. (1995). Final report of the RTCA board of director's select committee on free flight. RTCA, Inc., Washington, D.C.

Scallen, S. F., Smith, K, & Hancock, P.A.(1997). Influence of color cockpit displays of traffic information on pilot decision making in Free Flight. Proceedings of the 9th International Symposium on Aviation Psychology. Columbus, OH: Ohio State University.

Smith, K., and Hancock, P.A. (1995) Situation awareness is adaptive, externally-directed consciousness. Human Factors, 27, 1, 137-148.

Smith, K.C.S., Scallan, S.F. Knecht, W., & Hancock, P.A. (1998). An index of dynamic density. Human Factors, 40, 69-78.

van Gent, R.N.H.W., Hoekstra, J.M., & Ruigrok, R.C.J.. (1997). Free flight with airborne separation assurance. Proceedings of the Confederation of European Aerospace Societies' 10th European Aerospace Conference: Free Flight.

Session #4:
Planning and Doing

Degree of automation and its influence on the development of mental representations

Cornelia Ryser

Institute of Work Psychology
Swiss Federal Institute of Technology (ETH)
CH-8092 Zurich, Switzerland
ryser@ifap.bepr.ethz.ch
http://www.ifap.bepr.ethz.ch

Gudela Grote

Institute of Work Psychology
Swiss Federal Institute of Technology (ETH)
CH-8092 Zurich, Switzerland
grote@ifap.bepr.ethz.ch
http://www.ifap.bepr.ethz.ch

ABSTRACT

The theoretical and methodological background of a project is presented which aims at studying the relationship between the degree of automation of a technical system, the strategies applied by the operators while controlling the system, and their mental representations of the system. The empirical results of the study will be presented during the conference.

Keywords

Degree of automation, mental representations, control strategies.

INTRODUCTION

The increasing degree of automation of technical systems in virtually all working (and private) domains has profoundly changed the tasks and the role of humans working in human-machine systems. Sheridan (1987) coined the term of "supervisory control" to describe the human task in automated systems, where the operator performs a cognitive task which consists essentially of planning the task execution, programming the system, monitoring the automated process, intervening in the process in case of deviations, and learning from experience.

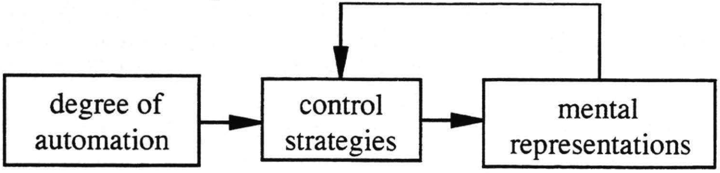
Ironically, while aiming at eliminating the human operator as far as possible from the processes, the increasing automation, leading to increasing complexity of technical systems, assigned the human operator an even more crucial position inasmuch as he is now primarily accountable for the control of uncertainties, i.e. for unexpected situations for which no technically provided algorithms exist: He plays the role of a creative problem solver who overcomes the still insufficient flexibility of technical systems (Bainbridge, 1987; Grote, 1997).
In order to be able to meet the requirements of this task, humans need to have control over the technical system (e.g. Grote, 1997; Hoc, 1993; Hollnagel, 1998; Ulich, 1998), i.e. the system must provide opportunities for comprehension of the process and prediction of its behavior as well as the possibility to influence it (Grote, 1997). These three prerequisites for control are not independent of each other: To make adequate use of the possibilities to control the process, one needs to understand it and to be able to anticipate the system's probable behaviour or the

possibilities one has to face unforeseen situations (e.g. Hollnagel, 1998). But one can only develop such an understanding and the ability for prediction by actively influencing the process itself (Grote, 1997).

The consideration of these three prerequisites for control and their interdependencies makes clear that control encompasses subjective as well as objective aspects: On the one hand, the system must provide opportunities for control (objective aspect). The degree of automation of the technical systems defines the potential amount of control the human operator has over the system by determining the ways an operator (potentially) interacts with the technical system. In other words: the degree of automation of the system determines the (potential) control strategies the operator applies when controlling the system.

On the other hand, however, these objective control opportunities need to be used by a human operator (subjective aspect). The control strategies the operator actually applies when controlling the process depend on his/her mental representation of the system (e.g. Brehmer, 1990; Hollnagel, 1993; Kluwe & Haider, 1990; Wickens, 1992). At the same time, however, the way the operator interacts with the system determines the kind of mental representation he will develop (e.g. Hacker, 1998; cf. also Kessel & Wickens, 1982) (cf. figure 1).

Figure 1: Assumed relationship between degree of automation, control strategies and mental representations of the technical system



In the present paper a study carried out to investigate the relationship between degree of automation, control strategies and mental representations will be presented. The focus of the paper does not lie on the results of the study which are still being processed (and will be object of the presentation at the conference). The aim of the paper is to present the theoretical and methodological background of the study.

Degree of automation

The term degree of automation refers to the way, functions are allocated between human operator and technical system. Several taxonomies and scales of degrees of automation exist in the literature (cf. e.g. Billings, 1997; Endsley & Kaber, 1999; Sheridan, 1987). The present study, however, bases on the criteria provided by the KOMPASS method to assess human-machine function allocation (see Grote et al., 1999; Grote et al., 2000). The term "degree of automation" does not just refer to the actual execution of the function at issue, but primarily to the allocation of decision authority within the system, i.e. the way decision authority with respect to the process itself as well as to access to information from the process is assigned to the human operator and/or the technical system.

Much empirical work has been done on the consequences of increasing automation on human tasks, behavior and cognition. Most of the studies involving degree of automation refer to its influence on system or human performance, workload, situation awareness and safety. However, there is a lack of empirical work on the influence of degree of automation on mental representations.

Control strategies

The "control strategies" describe the way the operator interacts with the technical system. The term "strategy" is used here in a rather broad sense, not necessarily as referring to some kind of cognitively predetermined behaviour, but more as a way to describe the operator's line of action related to certain aspects of the process control task.

Basing on the three main subtasks of supervisory control: monitoring of the process, intervention in the process and fault diagnosis, three aspects of control strategies are distinguished: monitoring strategies, diagnosis strategies and intervention strategies:

Monitoring strategies refer to all activities of the operator related to the supervision of the process, i.e. the (continuous) observation and documentation of process parameters, the diagnosis of the process state and the detection of variations and disturbances (cf. e.g. Moray, 1986; Parasuraman, Molloy, Mouloua & Hilburn, 1996);

- *fault diagnosis strategies* refer to everything the operator does in order to find out the cause of deviations from normal process state respectively evolution;
- *intervention strategies* include all actions of the operator by which he actively influences the process; two aspects of intervention strategies are considered:

On the one hand, the term "intervention" refers to the operator's activities of (re-)allocation of the functions between the system and himself, i.e. the use he makes of automation (cf. e.g. Lee & Moray, 1992, 1994; Riley, 1994). On the other hand, the strategies of the operators when actually influencing the process itself are considered, i.e. manipulations of valves, change of set points etc. Aspects of intervention in the process are, for instance, the choice of the parameter on which to intervene, the order and time interval of interventions as well as their "dosage" (Härtner, 1988).

This analytical and somehow arbitrary differentiation between the three aspects of the process control task may be questionable and an unequivocal allocation of the operator's activities to one of the categories can sometimes be difficult or even impossible. However, the distinction appears to be useful as an aid for the analysis of the behavioral and interview data (cf. below).

The distinction of the three aspects, does not describe types of control strategies, yet. Thus, two general types of strategies often cited in the literature, which can be considered as orthogonal to the above mentioned aspects, are introduced: feedback, closed-loop or reactive control on the one hand, and feedforward, open-loop or anticipative control on the other hand (e.g. Brehmer, 1990; Hollnagel, 1998; Wickens, 1992). Whereas the former refer to control decisions as response to actual information from the process, the latter refer to decisions taken by the operator on the basis of the anticipation of future system states. These strategies differ with reference to the amount of cognitive resources that they require and – which is not independent – to the types of mental models that they imply for process control. Studies have shown that the application of these two strategies is related to expertise (e.g. Brehmer, 1990; Moray, 1999), i.e. the tendency to switch from a feedback to a feedforward strategy with increasing experience, but also to the type of process to be controlled. It can be expected that the degree of automation of a technical system influences the strategy applied by the operator in so far as a high degree of automation induces a more reactive strategy whereas an anticipatory strategy is more likely to be applied for a low degree of automation, although the operator's expertise may play an important role as well.

An additional factor that seems particularly important to be considered when discussing about control strategies, is the temporal aspect (cf. e.g. De Keyser et al., 1998; Grosjean, 1999; Hoc, 1996). In particular dynamic environments, like the ones which are object of the present study – that evolve also without active interventions of the operator and/or that are characterized by feedback delays – call for feedforward control as the optimal control strategy (Brehmer, 1990).

Mental representations

Quite a broad consensus exists on the idea that process control is somehow related to a "mental representation" of the system to be controlled. However, there is little consensus about what such mental representations are. In fact, a huge variety of concepts and definitions can be found in the literature which differ from each other relating to the names used to designate them (e.g. mental models, internal representations, representational systems, operational images ("operative Abbildsysteme") etc.), the contents and definitions associated to each concept, the kind of systems and tasks to which they apply, the methods with which they are assessed etc. For the present work, the two terms "internal representation" and "mental model" are used as synonyms. The often cited definition of mental models by Rouse & Morris (1986) is taken as a basis: "Mental models are the mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future system states" (p. 351). Schumacher & Czerwinski (1992) provide a definition which is very similar, but they constrain explicitely its applicability to three specific "entities", which are also valid for the present study: (i) The definition applies only to physical entities and processes and (ii) these physical entities must be complex and (iii) dynamic.

From the proposed definition it is clear that mental models must represent some kind of knowledge representation, although Rouse and Morris (ibid.) stress the importance to differentiate the concept of mental model from that of "knowledge" in general.
Hacker (1998) describes by means of an illustrative example what it means for a plant operator to have a mental model: "He 'knows' the processes going on, i.e. he 'knows' how the parameters are connected, he has an 'idea' of the structure of the inner, not observable parts of the plant, he 'knows' numerous signals that show him process states which require his intervention, he 'disposes' of the necessary measures (actions), he 'knows' the possible consequences of certain actions, their conditions, temporal parameters and probabilities" (p. 187, trad. C.R.). These types of knowledge can be summarized in two general categories: factual knowledge, i.e. "know-what" or "how-the-system-works knowledge", and action knowledge, i.e. "know-how" or "how-to-use-a-system knowledge" (e.g. Hacker, 1998; Putz-Osterloh, 1993).

On what kind of knowledge structure, respectively on which contents the operator's mental representations base, depends on the type of control task he performs (Hacker, 1998; Kluwe, 1997a). This means that one of the crucial characteristics of mental models – as the concept is used in the present context – is that they are especially developed while (actively) interacting with their environment. This fact has at least two implications:

a) It implies that *different mental representations* of the same system can be developed *depending on the way the operator interacts with the system* (Hacker, 1998). The consequence is that the degree of automation of a technical system, while determining the way the operator (potentially) interacts with the system, must affect the operator's mental representation of the system.

b) Various *types of mental representations* must be distinguished. Rasmussen (1986), for instance, states that a person can have several different mental models of the system he/she is interacting with at different levels of abstraction. Depending on his/her goals and intentions, a person uses a different mental model of the system with which he/she interacts. These mental models can be described along two orthogonal dimensions: (i) *the abstract-concrete dimension*, representing the functional properties of a system. Five levels of abstraction are distinguished (pp.16-19): the level of physical form, the level of physical function, the level of generalized function, the level of abstract function, and the level of system purpose; (ii) *the whole-parts relationships* that relate to the level of detail at which the system is represented on each of the abstraction levels.

The idea that each operator has different mental models which can be described along an abstraction hierarchy is also supported by Moray (e.g. 1999) in his "lattice theory of mental models": "A person working in such a system (maintenance operators performing fault diagnosis) moves up and down the hierarchy, and across levels of part-whole decomposition analysis during such a task. The implication is that users have, not one, but many mental models of a given system, and that they will call up whatever model seems appropriate to the task of the moment" (p.237).

All this also means that, depending on the kind of task he has to perform, the operator "activates" a different kind of mental model. In order to allow him to perform his task effectively and efficiently, this mental model must be functional (Norman, 1983; Kluwe, 1990) or adequate (Grote, 1997).

This fact has important consequences for the design of human-machine systems. Process control tasks belong to the correspondence-driven work domain (Vicente, 1990), i.e. an objective reality, independent of the human-computer dyad, acts as the driving constraint for human-computer interaction. According to Vicente (1990), this calls for a systems perspective in the definition of requirements for system design: the interface should be designed as to allow the operator to develop a veridical mental model of the reality. If we consider the discussion about different types of mental models, however, it is clear, that it is difficult or even impossible to define what such a veridical mental model is. Experience shows that attempts to implement descriptions of mental models of operators into technical systems to support the human operator in the execution of his task, often fail due to the impossibility of an identification of the unique – veridical – mental representation adequate for all situations and for all system users.

One example that proves this impossibility to describe a universally valid model of the operator is provided by the present study. The system used in it encompasses dozens of variables whose relations are complex and cannot be easily described in a clear set of mathematical equations. The initial attempt to identify one objective model of the system to be compared with the mental representations of the subjects failed. There are as many different models of the system as there are different aspects to focus on (e.g. the mathematician's, the physicist's or the control engineer's models). Thus, it is not possible to describe the model of the operator either.

This fact has consequences for the design of technical systems as well as for operator training. Design and training should not be based on the idea of one unique and ideal mental representation but rather support the development and use of different representations, depending on the situation and on the operator's training level and preferences.

After having defined mental representations and described their characteristics, one crucial issue remains to be addressed: the question of their assessment. The notorious difficulty to measure mental representations has been widely discussed (e.g. Cooke, 1994; Leplat & Hoc, 1981; Rutherford & Wilson, 1992) and has produced a big variety of different methodologies. For capacity reasons, here, the various methods cannot be discussed in detail. Only the two methods or groups of methods which were applied in the study are briefly mentioned.

A group of often used techniques are verbal reports (see Bainbridge, 1979, for an overview). Several authors have discussed the issue of possible distortions of verbal reports and the conditions under which their validity can be considered as satisfactory (e.g. Leplat & Hoc, 1981; Ericsson & Simon, 1993).

A further method that can be applied to assess the operator's mental representation is a structural or causal diagram where the subjects have to draw on a diagram the relations between the process parameters (e.g. Funke, 1992).

To overcome at least some of the problems of the assessment of mental representations, a multi-method approach was chosen for the present study, i.e. the different aspects of the mental representations were assessed by means of different kinds of data. The methodological approach will be presented next.

METHOD

Setting and subjects

Quasi-experiments were carried out with a simulated production plant for methanol synthesis in use for the training of future operators in a Swiss chemical company. The simulator setting allows the variation of the degree of automation of the plant between manual control, automated control and mixed function allocation as well as the implementation of several variances and disturbances of varying severity. The experiments took place during the regular simulator training courses of all the students in the second year of apprenticeship (five classes, 47 apprentices all together).

The tasks

The students were introduced in the use of the simulator during about 1,5 days before the actual experiments were carried out. During this preparation they were given theoretical background about the chemical process of the methanol synthesis and were introduced theoretically and practically into the handling of the simulator and the methanol synthesis plant.

During the experiments, each student had to perform 3 phases of a control task. Each control phase lasted 30 to 45 minutes. In each phase, the apprentices were required to control the process either manually or automatically or they were given the freedom to choose the degree of automation they judged as the most appropriate. In each phase, furthermore, variances or disturbances were implemented.

Between the control phases or at the end of the whole series, the subjects were asked to work on several written tasks for the (partial) assessment of their mental representations and their perception of the process situation: Besides a repeated administration of a questionnaire about the actual state of the process, the subject's experience of control during the control task and his/her (possibly changing) preference for one of the degrees of automation, the subjects had to write shift protocols and instructions about the functioning of the plant for a new colleague, and to draw a structural diagram representing the causal links between the process parameters.

The group of subjects was divided into two subgroups that each were assigned to one of two versions of the experimental design: In version A the subjects changed between all the three possible control modes; the formation of these subjects' mental representations, thus, was based on their experience with different degrees of automation. In version B, instead, the subjects experienced only either the manual or the automatic control mode before being asked about their mental representation.
A few weeks later, the subjects performed once more the control task and answered the questionnaire about the actual process state and their personal experience of control.

During the control tasks, for 10 of the 47 apprentices more detailed data were collected: Actionpaths were drawn on the basis of behavioral protocols of the subjects' actions during process control. Subsequently to the entire series of control tasks, during an interview session, these 10 subjects were asked to verbalize the phases of the control tasks based on the action paths. Table 1 gives an overview on which data were acquired and how they relate to the issues addressed in the study.

Table 1: Overview of acquired data and measured variables

Degree of automation	
- automatic control	- all (possible) controls in the auto mode
- manual control	- all (possible) controls in the manual mode
- free choice of control mode	- subject chooses control mode for each component
Control strategies	
- Behavioral protocols/actionpaths	
- subsequent verbalization	
Mental representations	
- structural diagram	- factual knowledge
- written instruction for new colleague	- factual knowledge (actionknoledge)
- shift protocols	- factual knowledge
- verbalization of action paths	- action knowledge/factual knowledge

The acquired data are currently being analysed. The data processed so far provide some (provisional) evidence that the degree of automation actually has some influence on the development of mental representations of the studied subjects. The results will be presented at the conference.

REFERENCES

Bainbridge, L. (1979). Verbal reports as evidence of the process operator's knowledge. International Journal of Man-Machine-Studies, 11, 411-436.

Bainbridge, L. (1987). Ironies of automation. In J. Rasmussen, K. Duncan & J. Leplat (eds.). New Technology and Human Error (pp.271-283). Chichester: Wiley.

Billings, C.E. (1997). Aviation Automation. The Search for a Human-Centered Approach. Mahwah, New Jersey: Lawrence Erlbaum.

Brehmer, B. (1990). Strategies in real-time, dynamic decision making. In R.M. Hogarth (ed.). Insights in Decision Making. Chicago: The University of Chicago Press.

Cooke, N.J. (1994). Varieties of knowledge elicitation techniques. International Journal of Human-Computer Studies, 41, 801-849.

De Keyser, V., d'Ydewalle, G., Vandierendonck, A. (eds.) (1998). Time and the Dynamic Control of Behavior. Seattle, Bern: Hogrefe & Huber.

Endsley, M.R. & Kaber, D.B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics, 42, 462-492.

Ericsson, K.A. & Simon, H.A. (1993). Protocol Analysis. Verbal Reports as Data. (revised edition, 1st ed. 1984). Cambridge, Massachusetts & London: MIT Press.

Funke, J. (1992). Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung. Berlin: Springer.

Grosjean, V. (1999). Temporal awareness: pivotal in performance? Ergonomics, 42, 1443-1456.

Grote, G. (1997). Autonomie und Kontrolle. Zürich: vdf Hochschulverlag.

Grote, G., Wäfler, T., Ryser, C., Weik, S., Zölch, M. & Windischer, A. (1999). Wie sich Mensch und Technik sinnvoll ergänzen. Die Analyse automatisierter Produktionssysteme mit KOMPASS. Schriftenreihe Mensch-Technik-Organisation (ed. E. Ulich), Vol. 19. Zürich: vdf Hochschulverlag.

Grote, G., Ryser, C., Wäfler, T., Windischer, A. & Weik, S. (2000). KOMPASS: a method for complementary function allocation in automated work systems. International Journal of Human-Computer Studies, 52, 267-287.

Hacker, W. (1998). Allgemeine Arbeitspsychologie. Psychische Regulation von Arbeitstätigkeiten. Bern: Huber.

Härtner, C. (1988). Analyse, Bewertung und Ursachen von Strategien bei Prozesssteuerung an einem rechnersimulierten verfahrenstechnischen Wartenarbeitsplatz. Doctoral Thesis. Technical University Dresden, Germany.

Hoc, J.-M. (1993). Some dimensions of a cognitive typology of process control situations. Ergonomics, 36, 1445-1455.

Hoc, J.-M. (1996). Supervision et Contrôle de Processus. La Cognition en Situation Dynamique. Grenoble: Presses Universitaires de Grenoble.

Hollnagel, E. (1993). Models of cognition: procedural prototypes and contextual control. Le Travail Humain, 56, 27-51.

Hollnagel, E. (1998). Context, cognition and control. In Y. Waern (ed.). Co-operative process management. Cognition and information technology (pp.27-52). London: Taylor & Francis.

Kessel, C.J. & Wickens, C.D. (1982). The transfer of failure-detection skills between monitoring and controlling dynamic systems. Human Factors, 24, 49-60.

Kluwe, R.H. (1990). Problemlösen, Entscheiden und Denkfehler. In C. Graf Hoyos & B. Zimolong (Hrsg.). Ingenieurpsychologie (Enzyklopädie der Psychologie, Themenbereich D, Serie 3, Vol. 2, pp.121-147). Göttingen: Hogrefe.

Kluwe, R.H. (1997a). Acquisition of knowledge in the control of a simulated technical system. Le Travail Humain, 60, 61-85.

Kluwe, R.H. (1997b). Informationsverarbeitung, Wissen und mentale Modelle beim Umgang mit komplexen Systemen. In K. Sonntag & N. Schaper (eds). Störungsmanagement und Diagnosekompetenz. Schriftenreihe Mensch Technik Organisation (ed. E. Ulich), Vol. 13. Zürich: vdf Hochschulverlag.

Kluwe, R.H. & Haider, H. (1990). Modelle zur internen Repräsentation komplexer technischer Systeme. Sprache & Kognition, 9, 173-192.

Lee, J.D. & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. Ergonomics 35, 1243-1270.

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

Leplat, J. & Hoc, J-M. (1981). Subsequent verbalization in the study of cognitive processes. Ergonomics, 24, 743-755.

Moray, N. (1986). Monitoring behavior and supervisory control. In K.R. Boff, L. Kaufman, J.P. Thomas (eds.). Handbook of Perception and Human Performance. Vol. II, Cognitive Processes and Performance (pp.40-1 - 40-51). New York: Wiley.

Moray, N. (1999). Mental models in theory and practice. In D. Gopher & A. Koriat (eds.). Attention and Performance XVII. Cognitive Regulation of Performance: Interaction of Theory and Application (pp.223-258). Cambridge & London: The MIT Press.

Norman, D.A. (1983). Some observations on mental models. In D. Gentner & A.L. Stevens (eds.). Mental Models (pp.7-14). Hillsdale, New Jersey: Lawrence Erlbaum.

Parasuraman, R., Molloy, R., Mouloua, M. & Hilburn, B. (1996). Monitoring of Automated Systems. In R. Parasuraman & M. Mouloua (eds.). Automation and human performance: theory and applications (pp.91-115). Mahwah, New Jersey: Erlbaum.

Putz-Osterloh, W. (1993). Unterschiede im Erwerb und in der Reichweite des Wissens bei der Steuerung eines dynamischen Systems. Zeitschrift für experimentelle und angewandte Psychologie, 40, 386-410.

Rasmussen, J. (1986). Information Processing and Human-Machine Interaction. Amsterdam: Elsevier.

Riley, V. (1994). Human Use of Automation. Doctoral thesis. Faculty of the Graduate School of the University of Minnesota.

Rouse, W.B. & Morris, N.M. (1986). On looking into the black box: Prospects and limits in the search for mental models. Psychological Bulletin, 100, 349-363.

Rutherford, A. & Wilson, J.R. (1992). Searching for mental models in human-machine systems. In Y. Rogers, A. Rutherford & P.A. Bibby (eds.). Models in the Mind. Theory, Perspective & Application (pp.195-223). London: Academic Press.

Schumacher, R.M. & Czerwinsky, M.P. (1992). Mental models and the acquisition of expert knowledge. In R.R. Hoffman (ed.), The Psychology of Expertise (pp.61-79). New York: Springer.

Sheridan, T.B. (1987). Supervisory control. In G. Salvendy (ed.). Handbook of human factors (pp.1243-1268). New York: Wiley.

Ulich, E. (1998). Arbeitspsychologie (4th Edition). Zürich: vdf Hochschulverlag; Stuttgart: Schäffer-Pöschel.

Vicente, K.J. (1990). Coherence- and correspondence-driven work domains: implications for sytems design. Behaviour & Information Technology, 9, 493-502.

Wickens, C.D. (1992). Engineering Psychology and Human Performance. New York: HarperCollins Publishers.

Exploring the Metaphor of "Automation as a Team Player": taking team playing seriously

Christian Fairburn
HCI Group
University of York
York YO10 5DD, UK
chrisf@cs.york.ac.uk
http://www.cs.york.ac.uk/~chrisf

Peter Wright
HCI Group
University of York
York YO10 5DD, UK
pcw@cs.york.ac.uk
http://www.cs.york.ac.uk/~pcw

ABSTRACT

It is becoming popular to view automated systems as 'team players' in distributed co-operative work environments. This paper scrutinises such a viewpoint in detail, focusing on the implications of being a team player and whether today's automation is really capable of achieving this type of status. We view agent interaction as more than information transfer. True teamplay implies co-operation and active collaboration. We use work concerned with the description of human-human collaboration to highlight the importance of mutuality in true collaboration. A recent aircraft accident in which human-automation interaction was causally implicated is used to illustrate how today's automation falls short of achieving true participation and by implication fails as a team player.

Our aim in particular is to draw attention to the limits of the 'automation as team player' metaphor as a goal of automation design. The key question designers should be asking is not 'how can we make automation a better teamplayer ?', but 'how can we make automation compatible with the teamplaying context in which it will be deployed ?'.

Keywords

Human-computer interaction, computer supported co-operative work, automation, cognitive systems engineering, teamplay.

INTRODUCTION

Computerised systems are becoming increasingly prevalent across all domains of work, from the mundane to the safety critical. Furthermore, today's computerised systems are also becoming more complex. For example, the highly automated systems in a modern commercial aircraft are less like tools with which one can achieve work, and more like a semi-autonomous agent to whom one assigns work (Sarter & Woods, 1992; Brennan, 1998). Today's pilots frequently delegate tasks to automation in advance of their execution (for example, by pre-programming the flightpath using the Flight Management System (FMS)). Humans then take a supervisory role while the automation itself performs the work. Conversely, in some cases designers have given automated systems the authority to override or question human action. In effect the automation takes charge of safety and takes on the supervisory role itself. A good example of this is the automatically enforced predefined protective envelope of performance which is designed to prevent an aircraft from stalling or pulling excessive 'G' (see Hughes & Dornheim, 1995).

This new breed of complex automation is frequently found in complex and distributed work environments (e.g. flight control, nuclear power generation, industrial production processes). These are domains in which the performance of work, often safety critical in nature, relies heavily on the real-time distributed performance of a whole set of inter-related and interdependent sub-tasks. It is important for designers of this new type of automation, and of socio-technical work systems to have some conception of the way in which automation may fit into the performance of joint work and the process of collaboration. As a response to this need, a popular metaphor has emerged which views automation as a 'team player' (e.g. Sarter & Woods, 1992; 1995; 1997).

The team playing perspective would imply that we may be able to view automated systems as intelligent agents who take part in the performance of collaborative or joint work. So far, the teamplay between humans and automation (if such a thing can be said to exist) has to some degree been primarily described in terms of the outcome (success or failure), or information transfer (the automation is 'strong but silent' about its actions and can result in 'surprises' (Sarter & Woods, 1995; 1997)). For sure, valuable suggestions for improvement have been made on the basis of unsuccessful instances of human-automation teamplay but we feel there is a need to take a step back and look in more detail to the nature of what it means to really take part in teamplay. How useful is the team player metaphor and how far can it be pushed before it becomes invalid ?

This paper attempts to link research connected with human-human co-operation to that concerned with the possibility of human-machine co-operation. We draw upon work within the domains of Human Computer Interaction (HCI), Computer Supported Co-operative Work (CSCW), the sociology of work and the study of joint work surrounding conversation, in an attempt to discover whether current descriptions of teamplay between humans can be brought to bear on the 'automation as team player' metaphor.

In order to exemplify our view on what it means to participate, we describe three case studies. Two from the domain of air traffic control illustrate the subtlety of human-human responses to co-ordination failure. The third, a description of an air accident, is used to illustrate how today's automated systems fall short of achieving true participation and by implication, fail as team players.

Cognitive Systems Engineering

Hollnagel & Woods (1983), in their cognitive systems engineering (CSE) approach, make a strong argument for viewing a human-automation system as a cognitive system. They define a cognitive system as one which produces goal-oriented intelligent action, and operates using knowledge about itself and its environment. The cognitive study of human-machine systems is said to be necessary when studying today's computer applications since through increased automation, the nature of the human operator's task has shifted from an emphasis on perceptual-motor skills to an emphasis on cognitive activities (i.e. problem solving, decision making). Also, through the increasing sophistication of technology, the human-machine interface is said to be gradually becoming the interaction between two cognitive systems. The authors class humans as obviously cognitive systems, machines to be "potentially if not actually, cognitive systems" and a human-machine system as a whole is "definitely a cognitive system" (Hollnagel & Woods, 1983 p. 589).

In more recent work, Hollnagel has also taken an adaptive systems view of cognitive systems (Hollnagel 1999). He argues that a characteristic of a cognitive system is its ability to modify its pattern of behaviour on the basis of past experience so as to achieve specific anti-entropic ends"

(Hollnagel, 1999, p.42). This definition focuses less on the components of the cognitive system and how they operate, and concentrates instead on what a cognitive system does and the purposes it serves. By focusing on the human-machine system as a joint system, he argues that "it thereby becomes important how humans and machines co-operate to achieve the joint system function and how they can complement and support each other" (Hollnagel 1999, p.44). Hollnagel points out that the cognitive system is, more often than not, unable to completely predict how the external world may develop (due to complexity or a probabilistic nature). For this reason, cognitive systems (joint cognitive systems included) must be able to plan ahead, to anticipate the dynamics of the work and respond adaptively towards a new equilibrium once it has been disturbed. Thus the cognitive system must continuously act and observe, plan and re-plan in order to maintain equilibrium in the face of disturbances.

The 'Team Player' Metaphor

The 'team player' metaphor is frequently exercised in research related to the cognitive systems engineering view. In consideration how joint cognitive systems achieve their goals, Hollnagel & Bye (2000) suggest that humans and machines must "collaborate". When working together in a socio-technical system crews carry out operations comprised of sequences of individual tasks which are distributed and co-ordinated among crewmembers. In other words, in order to achieve high level performance objectives, lower level tasks are performed in a distributed way (by both humans and machines). But they argue that for this type of collaboration to work, there needs to be a high degree of visibility of individual actions in order to ensure the on-going synchronisation of system components

Sarter & Woods (1992, 1995, 1997), studying the experiences of civil flightcrew with flight management systems (FMS), describe how cognitive work in today's modern cockpits is "inherently co-operative" between the pilots and automation (Sarter & Woods, 1992). In later studies, the automation is referred to as a "machine agent" in the human-machine system. Human crew members are just one component of a distributed multi-agent architecture. In an observational study they are said to have employed methods to 'communicate' with 'automated partners' in ways not dissimilar to interpersonal communication[9] (Sarter & Woods, 1995).

Researchers have also taken up the team player metaphor to provide guidance for designing 'intelligent' automated systems that are effective team players in flight operations support, for example the NASA study of Malin et al. (1991). These authors adopt a definition of 'intelligent agent' which clearly permits membership by either human or computer based entities ('intelligent' fault management systems, for example). Design issues identified in the NASA study include "collaboration between the operator and the intelligent system". This is where a two-way exchange of information between the operator and automation is required to "develop a shared view of the world".

The way in which these types of research have developed and used the team player metaphor show a bias towards the definition and description of what it means to be an 'intelligent' agent. Of interest, however, is the fact that describing or defining an intelligent agent is not equivalent to describing or defining an intelligent agent who takes part in teamplay. While we do not necessarily dispute the view of a human-machine system as a cognitive system, we have reservations regarding the way in which machines have been classed as team playing agents by some authors. These reservations centre around the issue of agent interaction. Agent interaction is more than communication or information exchange. This would be a simplification of teamplay. Teamplay, we argue, is about more than co-ordination in its simplest sense, the mere alignment of tasks through

---

[9] Though these methods were often inappropriate and resulted in an unsuccessful outcome.

information transfer. Instead it is about joint work and active collaboration. But if we are to defend this position we must be clear about what is distinctive about joint work and active collaboration.

With this in mind, it becomes important to look to literature concerned with human-human co-operation if we are to determine the distinction between human-human and human-machine co-operation and, in turn explore the extent to which we can usefully apply the 'machine as team player' metaphor.

TAKING TEAM PLAYING SERIOUSLY

While there is a great deal of literature which characterises the relationship between automation and humans as teamplay (or the absence of it), there has been relatively little concern within the cognitive system literatures for how teamplay is achieved and sustained. For this we need to turn towards work in the areas of CSCW and language use.

The distribution of tasks within a team and the consequent need for a means of co-ordinating and synchronising such tasks has led some researchers to consider the mechanisms by which such co-ordination is achieved. This management or co-ordination can itself constitute significant work. This has been referred to as articulation work (Strauss 1988) the word 'articulation' being used to refer to the act or mode of joining (see Wright Dearden and Fields 2000 for a more detailed description). Schmidt & Simone (1996) have pointed out the way in which articulation work may be bound up with the settings surrounding collaborative work, co-ordination mechanisms ranging from the availability of overhearing to physical information artefacts such as standard procedures and job sheets. The researchers in this way have described how artefacts can be implicated in collaboration through mediation.

From a superficial reading of this literature, it is tempting to argue that articulation work can be a mechanistic process of planning a sequence of tasks, identifying dependencies and distributing the tasks accordingly. Similarly, that co-ordination can be achieved through the use of job-sheets and other planning artefacts. Teamwork, however, is as open to the vicissitudes of an unpredictable environment as any other kind of activity. In reality, changes to the environment mean that plans have to be modified on the fly and changes to the allocation of tasks in a team have to change in often quite dynamic ways. It is in these disturbances of the joint cognitive system that we gain a greater insight into the nature of teamplay and the complexity of articulation work.

Schmidt Simone (1996) pointed not only to the way in which production line work can be coordinated by a concrete and relatively static planning artefact, the Kanban, but also to the way in which workers deviated from the procedures laid down on the Kanban to deal with unplanned but regular disturbances to the flow of work. They achieved this by, for example, stashing the Kanban card in their back pocket until equilibrium was recovered.

This type of dynamic interplay between the Kanban, as a pre-computed plan for the coordination of work and the workers' adaptation of the plan on a moment-to-moment basis is characterised by Wenger (1998) as an interplay between reification and participation in a community of practice. For Wenger, reification is a process of making an abstraction tangible or concrete in a material artefact. As examples Wenger offers: writing down a law, creating a procedure and producing a tool. Thus the Kanban is reification of a co-ordination plan. However, crucially Wenger goes on to argue that the products of reification can only have meaning in practice through the participation of a community's members. Thus returning to Schmidt and Simone's (1996) example, only because of the history of the community's participation can the Kanban serve both as plan to be followed strictly and as something to be worked around or oriented to as a resource for action (Schmidt 1997).

Wenger is, however, quick to point out that reification without appropriate participation can ossify activity. Formal procedures can blind workers to deeper meanings. Thus teamplay is about the process of construction (or "negotiation") of meaning which occurs when members participate around reifications. The quality of that teamplay is characterised in part by a trade-off between reification and participation. How much of the meaning is distributed in the artefact and how much in the participation? A computer program for instance, according to Wenger, would be the ultimate example of reification. A program is capable of interpretation by a machine that has no access to the meaning of that program. In contrast, a poem is designed to rely heavily on participation for its meaning. Meanwhile conversation constitutes a case where reification and participation are woven so tightly that the distinction becomes blurred, language use can become transparent and conversation seems like pure participation. Wenger reserves participation for human-human interaction. Participation cannot take place except through a process of mutual recognition by which he means that there must exist at least a possibility that what one individual does can contribute to another's experience of meaning and vice versa. Because of this mutuality, participation is not like producing a carving or molding a piece of clay. We would not construe our shaping of these objects as contributing to their experience of meaning. Thus computers and machines cannot participate in a community because they cannot experience this mutuality.

Wenger's claim about the exclusion of machines in teamplay is almost an apriori one. It does not account for why mutuality is so central to teamplay. For this account we need to turn to the area of linguistic pragmatics to find a much more detailed account of the role that mutuality knowledge plays in joint activity.

Common Ground & Grounding

Conversation is one of the finest examples of tightly coordinated participative activity that humans engage in. Elements of Clark's research on conversation and language use have already been brought to bear on the more general study of joint activity or the performance of co-ordinated work (e.g. Brennan 1998; Fairburn et al. 1999). Joint activity is described by Clark as being a goal-defined event which has more than one participant. Joint activities are performed through sub-steps termed 'joint actions' and these are characterised by the co-ordination of the participants involved.

For our current purposes it is important to note the way in which Clark describes participation in relation to joint work. Participation involves the construction and maintenance of 'common ground'. This is a representation of what information, beliefs and knowledge is not only shared between participants, but mutually known by them too. By mutually known, Clark means that a piece of information is known by each participant and, furthermore, each participant knows that the other participant knows this piece of information. For Clark, the concept of common ground and this 'grounding' process are central to the performance of joint work. They are directly related to the jointness and mutuality associated with human co-operation.

In Clark's terms, joint information requires participants to know the information and also to know that each participant knows the information. Actors can hold mutual knowledge regarding aspects of their environment, and they can hold mutual knowledge regarding one another's goals and intentions.

For Clark, computers may in some sense have access to aspects of common ground - they may be able to fulfil the first clause of jointness through holding pieces of information or data contained within common ground, but they are incapable of fulfilling the second clause. Computers cannot recognise nor construct mutuality. Consequently, when humans and machines are involved in the performance of work, there is an asymmetry between them. In particular, computers cannot gain insight regarding other actors' intentions and goals. Furthermore, though attempts have been made to view human-machine interaction in light of Clark's principles, such work has found it necessary to distinguish between the jointness of knowledge inferred by Clark's definition of common ground, and the notion of 'shared context' that may be said to be present between human beings and computer systems (Brennan 1998).

Given that we now have some idea of the nature of true teamplay we will move on to discuss two types of real-life scenario that embody instances of human-human teamplay and human-automation interaction. Comparing the nature of these patterns of interaction in the light of the literature we have just reviewed will allow us to pinpoint the differences between the teamplay of humans and the performance of work by humans and machines in conjunction. We look first to illustrate the richness and complexity of true teamplay using the domain of Air Traffic Control (ATC) as an example.

JOINT WORK IN AIR TRAFFIC CONTROL

The domain of ATC provides an excellent illustration of the way in which teamplay generates adaptive behaviour in response to unpredictable and evolving situations. Furthermore, the two following examples show how the mutuality described by Wenger and Clark is central to the articulation work required when a system moves out of equilibrium.

An Attempted Shortcut

In a recent study Fairburn, Wright and Fields (1999) reported the roles that common ground and overhearing played in dealing with an unexpected ATC routing problem :

An air traffic controller using broadcast VHF radio had given a standard approach route to an aircraft. This required the pilot to fly a pre-planned path down to the airfield. Shortly after this had been agreed, the aircraft was seen to depart from this standard approach route. Negotiations over VHF eventually resulted in the airliner resuming its standard approach. Subsequently however another airliner, having overheard the previous communications, contacted the controller to query the standard approach route to the airport. Shortly after this query, the controller contacted the first aircraft and cleared it to take the shortcut that had been in dispute moments before (see Fairburn, Wright and Fields (1999) for much more detail).

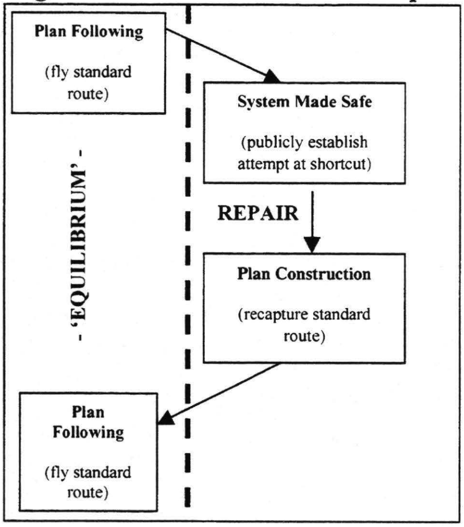
Mutuality played quite a complex part in this scenario. With respect to the initiation of the attempted shortcut, it seems that mutual knowledge had a role to play regarding the way in which implicit expectations on behalf of aircrew could interfere with explicitly agreed upon plans of action. With respect to the repair of the incident, at a basic level mutual knowledge regarding the names of the waypoints on the approach route allowed a mispronunciation to go unrepaired without a negative impact on work[10]. At a more complex level, mutually shared knowledge or common ground served as a resource for action - a means for detecting deviations from planned behaviour. Because the controller knew that the standard routing through the sector had been accepted and he also knew that the pilot knew the routing had been accepted (i.e. the plan to fly the standard approach route was mutually held) the controller was able to infer an attempt to renege on the agreed upon plan of action. Furthermore, it was clear that the deviation from the agreed upon flightpath could be understood as an attempted shortcut. Mutuality played a key role in returning the ATC system back to a state of equilibrium (Figure 1). Shortly after this the air traffic controller re-routed the aircraft along the shortcut route it had tried to take earlier. What this suggests is that it was not the shortcut per-se that was at issue for the controller, but the fact that he

---

[10] The articulated name of 'Practica' was close enough to the true name of the waypoint 'Pratica' to avoid confusion, the controller knew or could correctly deduce to which waypoint the pilot was referring.

had not explicitly sanctioned it as a collaborative plan. Intervening in the attempted shortcut the air traffic controller's intention was to ensure that the articulation work was under his control and that it was mutually known by all parties that such work was his responsibility.

Figure 1: The Process of Shortcut Repair

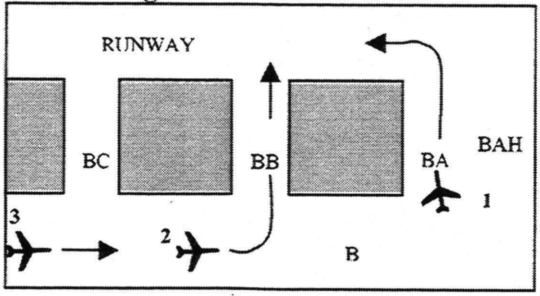


A Runway Incursion

Our second example scenario comes from a ground control case study conducted at a major European airport and is based on observational, ethnographical and informant data obtained as part of the MEFISTO research project.

The agents central to the scenario are distributed across the ATC system. The aircrew of three aircraft (referred to here as 1, 2 and 3) are located on the airport's taxiways and departure runway. In the control tower there is the Ground Controller (responsible for controlling all aircraft during taxiing), the Tower Controller (responsible for co-ordinating and controlling aircraft on the departure and arrival runways) and the Departures Co-ordinator (responsible for ordering and controlling the push-back and initiation of taxiing of all aircraft). Meanwhile, located in an en-route control centre removed from the airport the En-route Controller orders and controls aircraft subsequent to take-off.

Figure 2: An Incident Occurs

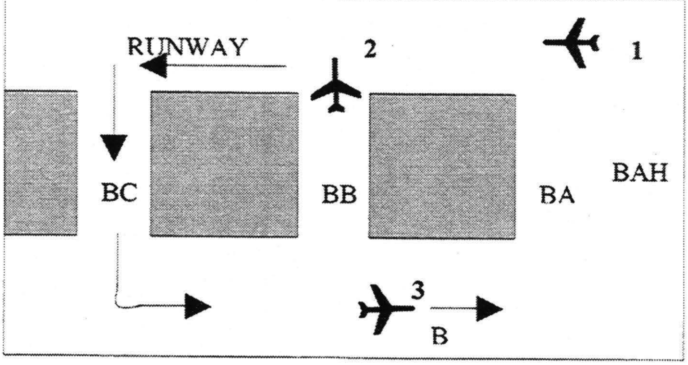


The scenario takes place on a foggy day where visibility is poor (both for aircraft and controllers who in normal conditions can view the taxiways and runways through the tower windows). The three aircraft are in planned departure order, preparing to depart on the appropriate runway (see figure 2). Aircraft 1 is under control of the Tower Controller and has just been cleared to enter the runway in preparation for take-off. Aircraft 2 and 3 are under the Ground Controller's authority. The planned (and cleared) taxi route for all aircraft is to arrive at the holding position of the runway (BAH) via taxiway B and then BA. In the low visibility conditions however, aircraft 2 mistakenly takes exit BB from taxiway B and begins to enter the runway. The scenario evolves as shown in Table 1. The outcome is that aircraft 1 remains stationary on the runway while 3 moves up in the departure sequence and 2 will complete a circuit to re-join the departure queue behind 3 (Figure 3).

Table 1: Runway Incursion Scenario

ACTION SEQUENCE
Pilot of 2 notices the mistake (upon entering the runway) and calls the Ground Controller.
Ground Controller instructs 2 to stop immediately (2 has probably already stopped).
Ground Controller informs the Tower Controller of the situation (though Tower may have overheard).
Tower Controller instructs 1 (at this point on the runway) to stop immediately.
Ground Controller makes a decision about how to resolve the situation. S/he instructs 2 to enter the runway and leave by BC to rejoin the departure queue. Ground Controller instructs 2 accordingly informing the crew that they are now 3rd in the departure queue.
Ground informs 3 that they are now 2nd in the queue.
Ground informs departure co-ordinator of the order change.
Departure co-ordinator informs en-route centre of the change to the departure sequence.
Ground Controller re-computes the departure sequence swapping order of 2 & 3 (reflected in flight strip re-ordering).

Figure 3: The Incident is Resolved

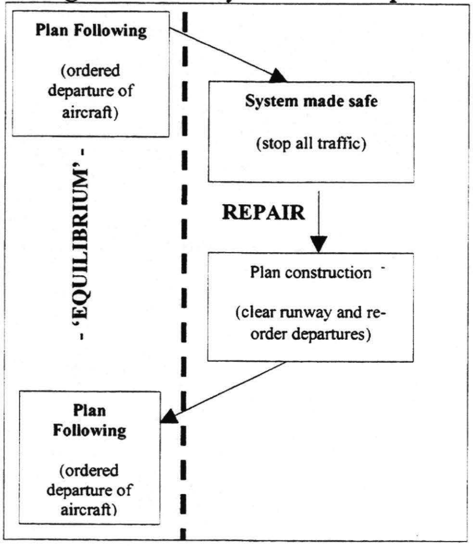


The scenario is of course very complex. The way in which the Ground Controller chooses to resolve the situation is linked with aircraft type, direction of intended travel and slot times for example. The course of action chosen in this case is just one possibility (in a different situation it may have been more critical to preserve the initial departure order for example). We can still, however, use this high level description of the scenario and the work domain structure to illustrate how mutuality plays a key role in real-time teamplay.

The way in which work takes place in the scenario, as in the shortcut case, allows a clearly discernable process of repair to emerge. This process of repair has two sub-components, both of which (making the system safe and regaining equilibrium) are co-operative in nature (Figure 4).

The problem (the wrong turn) is first identified by the crew of aircraft 2. In effect the high level task of making the system safe is initiated co-operatively by aircraft 2. The flightcrew do not feel that they can solve the problem themselves and so they actively seek co-operation from the controller, 'look, we've made a mistake - what shall we do now, given our situation?'. It is important to note that, due to the broadcast nature of VHF radio, aircraft 3 will have overheard this alert also, but aircraft 1 is monitoring a different frequency under authority of the Tower Controller. Therefore, once the incursion has been identified, flagged up for the Ground Controller's attention, s/he alerts the Tower Controller who in turn passes on information to aircraft 1.

Figure 4: Runway Incursion Repair



Once the system has been made safe, the Ground Controller must devise or plan a solution - a method whereby the aircraft can get moving again. In Hollnagel's terms the equilibrium must be restored. When a strategy has been decided upon, the plan following steps are co-operative. The Ground Controller negotiates sub-plans with each aircraft under his/her control, the structure of the domain ensuring that these negotiations are often shared by other aircrew (via VHF), and more likely than not by the Tower Controller also (located beside the Ground Controller in the tower cab).

Common ground plays a pivotal role in the resolution of the runway incursion. What we see in this scenario is an actor who is in overall charge of managing the articulation work required for solving the problem: The Ground Controller agrees upon plans of action with each aircraft managing and co-operatively delegating work to them, while ensuring that these sub-plans all align with the overall purpose of restoring equilibrium to the airport (see Figure 5). At the same time, however, the way in which this work is organised maintains a high level of mutuality across actors. This mutuality is supported by the direct interaction of actors, the structure of the work domain (which allows, in particular for overhearing and verbal negotiation) and the way in which actors already bring with them common ground which pre-dates and transcends this particular scenario (e.g., original departure plan, ways in which ATC system functions etc.).

The Ground Controller agrees upon sub-plans with other actors and other agents are often privy to the sub-plans which are being followed at any one time. In effect, there is a high degree of overlap between agreed upon actions (Figure 6). This translates into a high level of safety. Mutuality, as in the case of the shortcut, makes deviations from planned behaviour easier to spot. Also, actors are able to maintain some level of awareness regarding the current state of play from a systemic point of view. They understand their position in the high level plan to return the system to equilibrium and they can trace the progression of this high level plan through its various substeps.

Figure 5: Ground Controller as Co-ordinator

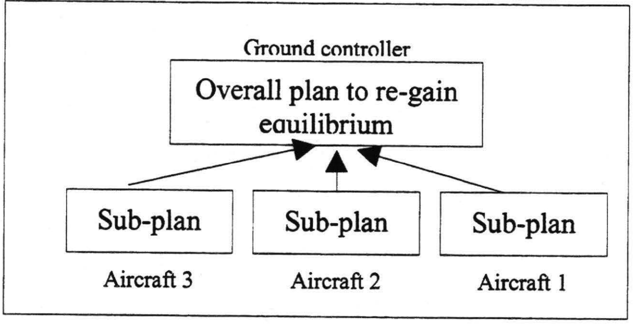
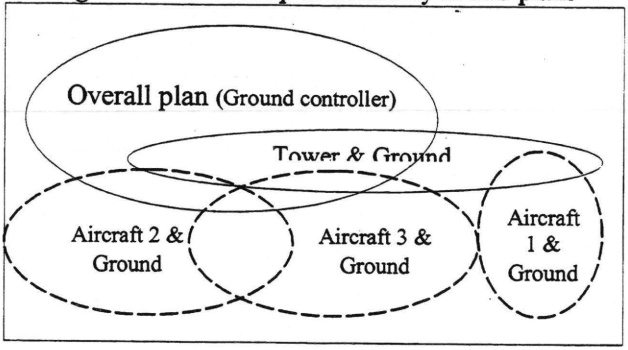


Figure 6: The Overlap & Mutuality of Sub-plans



The ATC examples demonstrate the often complex articulation work that is possible when humans co-operate to achieve mutually known goals. In both cases we see examples of a disturbance to the system causing what Wenger refers to as an interactional misalignment : meanings, actions and plans temporally fall out of alignment and work is required to bring them back into alignment. These examples provide insight into what it means for teamplay to be successful.

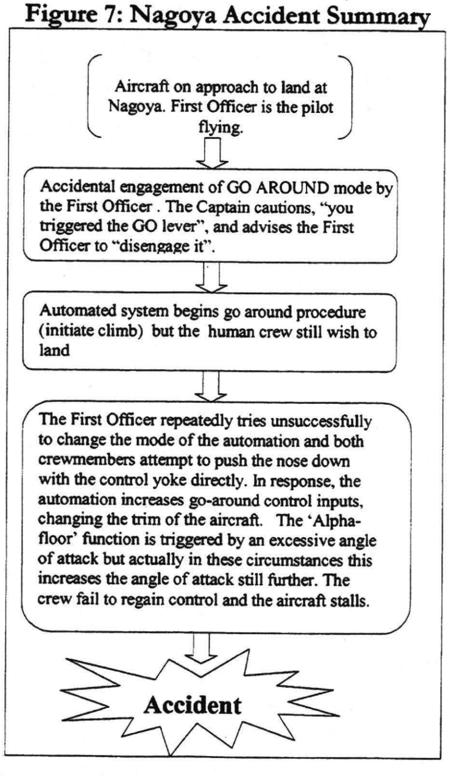
Now that we have looked at teamplay and how truly adaptive behaviour is exhibited when work is assigned across humans, we move on to focus on how the delegation of tasks to automated systems affects the nature of articulation work and common ground.

INTERACTIONAL MISALIGNMENT IN AN ACCIDENT

The aircraft accident at Nagoya (see AAIC 1996) is taken here as an example of an increasingly common form of accident in today's complex distributed domains of work. In a 'going sour' incident (see Cook, Woods & McDonald, 1991; Dekker & Woods, 1999) some kind of potentially recoverable event is allowed to evolve, through a process of commissions and omissions, failed collaboration and communication, into a serious or disastrous outcome. In this type of situation perfectly functioning aircraft can be flown into the ground as the combined effect of humans and automated systems gradually manage a situation into an accident. Our concern here is not to provide another account of this particular accident, nor to offer an explanation of why the accident happened. Rather it is to focus on what would have been required of the automation to deal with the interactional misalignment and bring the system back into equilibrium, given that the pilots were unable to do so unaided.

In the case of Nagoya, the first officer appears to have accidentally triggered the automated 'go around' function while on approach to land (see Figure 7). Activation of this facility resulted in the automation aborting the approach and initiating a climb. It is unclear exactly why the crew were unable to recover control of the aircraft, or indeed why they did not 'give in' to the automation and try another approach. However, for our purposes the initiation of this accident and its evolution towards failure provide a powerful and rich illustration of failed human-automation interaction. The evolving, going sour pattern of Nagoya contains within it automation surprises exemplifying the potential effects of strong but silent automation. More importantly, we can use the accident as a basis for the discussion of human-automation teamplay. The 'go-around' and 'Alpha-floor'[11] functions can be used to give us insight into the way in which work is changed when automated systems are assigned tasks, as opposed to how work is performed during teamplay.

Figure 7: Nagoya Accident Summary



Aircraft on approach to land at Nagoya. First Officer is the pilot flying.

Accidental engagement of GO AROUND mode by the First Officer . The Captain cautions, "you triggered the GO lever", and advises the First Officer to "disengage it".

Automated system begins go around procedure (initiate climb) but the human crew still wish to land

The First Officer repeatedly tries unsuccessfully to change the mode of the automation and both crewmembers attempt to push the nose down with the control yoke directly. In response, the automation increases go-around control inputs, changing the trim of the aircraft. The 'Alpha-floor' function is triggered by an excessive angle of attack but actually in these circumstances this increases the angle of attack still further. The crew fail to regain control and the aircraft stalls.

When assigning functions like the go-around to automation, flightcrew are required to pre-specify variables (e.g. in this case intended altitude). In this way, when a the automated function is desired all that is required is a trigger (the operation of a lever or button). In effect, delegating work such as

---

[11] Designed to achieve optimal climb configuration for the aircraft.

the go-around to the automation produces a high level unit of action or function : to go around, climbing to a certain height. The way in which the automation actually operates has been decided in advance by designers - In effect, the function of going around is rendered an atomic task when it is handed over to the automated system.

The particular crew involved in this accident had insufficient knowledge of the way in which the automation functions to be able to split this atom into its constituent parts. They do not manage to escape this mode and regain their descent. Once the human-automation battle had commenced (aircraft climbing, the crew wanting to descend), the Nagoya accident could have been avoided in 1 of 2 ways - either the human crew could have 'given in' to the automation, or the automated system could have relinquished control back to the human crew. In addition if the aircraft had been slightly higher, pushing forward on the control sticks would have had the desired effect of disengaging the go around function (Ladkin 1996).

From the point of view of articulation work, the crew are clear that the go-around manouevre has been allocated to the automation (albeit accidentally), but they do not have the knowledge to over-ride that allocation and take back control. At the same time, the automation is unaware that that there is anything accidental about the articulation. From the point of view of common ground, in this part of the joint activity, the cognitive system does not even achieve the first of Clark's requirements. The information required to repair the misalignment is not known by the crew so they misunderstand what is required of them to repair. They continue to push the control stick forward. Thus it would seem that the only hope of repair here lies with the automation. What is required is for the automation through the mutuality of teamplay to recognise a misunderstanding and to initiate a repair sequence on its own initiative. What would be required of the automation to do this?

The fundamental problem for the automation is to able to interpret the nose down inputs as a request from the crew to abandon the go-around. Such a request is one at the level of articulation work. In order to have understood the meaning of the crew's inputs in this way the automation would have needed to know that the articulation of the go-around was not intended and was thus not consistent with the crew's intentions; their true intent was to continue the descent but they lacked the ability express this intent given the current state of the environment (see Table 2). None of this set of highly embedded propositions could have been known by the automation. There is evidence (See Ladkin, 1996; AAIC, 1996) to suggest that the crew's goal of aborting the go-around was not feasible for the latter part of this accident scenario. So, in their own turn, the crew through teamplay, should have been made aware that their plan was infeasible. Thus, they would have needed to understand the aircraft's continued descent and increasing pitch angle as an attempt by the automation to bring its own goals and those of the crew into alignment. For them to understand the aircraft's actions at this level, the crew would have had to assume that the automation was aware of their intent (to abort the go-around). Such an assumption of mutual knowledge, had it been made, would of course have been false (the automation does not have this knowledge) but ironically, it might have saved the aircraft.

Lack of common ground between crew and aircraft in the Nagoya accident meant that the joint cognitive system was anything but adaptive. Instead, with the aircraft under its control the automation responded in the only way it was capable, by following a pre-determined, rule based (if … then) program in response to the contingencies that arose. The triggering of the Alpha-floor function exemplifies this strategy. Responding to contingencies is not the same as being adaptive - adaptivity implies the ability, if required, to change the whole set of criteria by which one operates. The automation's inability to act adaptively shifted this responsibility to the humans, who in this case were also unable to successfully resolve the situation.

Table 2: Lack of Common Ground Between Crew & Automation. Table inspired by Suchman (1987)

The Crew		The automation	
Not available to automation	Available to automation	Available to crew	Not available to crew
Accidental nature of activiation	Go-around initiated	In Go Around mode	
Go-around is unintended			
Intentions to abort go around	Nose-down inputs opposing climb	Aircraft continues to climb and increase angle of attack	Counteract opposing forces by adjustment of trim to maintain go-around
			Pitch trim control switch disabled when autopilots are engaged

DISCUSSION

As Hollnagel points out, cognitive systems strive to maintain some type of equilibrium within their environment. Our case studies illustrate this point well, the repair processes in the ATC scenarios clearly being angled towards returning the safety critical environment of ATC back to a safe and expeditious condition. Our example of the Nagoya accident showed a complex system thrown into disequilibrium through a lack of the kinds of mutuality that characterised the recovery in the ATC examples. In addition, Hollnagel describes how complex environments are not always predictable (ATC again exemplifies this). Adaptivity is thus at the heart of joint cognitive systems, but can we really say that human-machine systems are truly adaptive in the same way that human team players are ? Research within human-automation interaction has tended to take a simplified, information transfer viewpoint of teamplay thus far. What is not evident from descriptions of distributed work (see Hollnagel & Bye, 2000) is the way in which automation has an inherent disability when it comes to teamplay.

When humans work together task information is gained through explicit information transfer (e.g. communication, the recording of data on artefacts), but may also be held or deduced implicitly via common ground. The ability of humans to infer the plans and goals of others rests heavily on the ability to understand information transfer in the context of common ground (i.e. mutually held knowledge). As we have discussed, automation cannot fulfil the second clause of common ground, it is incapable of constructing or achieving mutuality. Without common ground, automation cannot infer the intentions or goals of others, nor can it realise meaning from action[12].

Our research would suggest in particular, that the 'automation as team player' metaphor cannot be stretched to encompass the process of repair, as it is during this type of work that mutuality provides great benefit and the type of asymmetry exhibited by human-machine interaction becomes

---

[12] Note : even if the automation in the Nagoya accident had discontinued the go-around due to the pilots pushing their control yokes forward, though lives might have been saved the computer would still not have deduced meaning from the humans' actions.

most detrimental to performance. Cases of failed human-machine interaction prominent in 'going sour' incidents represent situations in which environmental and design factors have caused the characteristically asymmetrical nature of normal human-machine interaction to be extended beyond the bounds of safety. The direct human-automation conflict sometimes observed in cases such as Nagoya is a powerful illustration and ultimate extension of this asymmetry.

In order to explore our viewpoint in more detail we can turn to the description of the Cali accident (see Aeronautica Civil, 1996) provided by Dekker & Orasanu (1999). Prior to the accident (another example of a 'going sour' scenario) while on approach to Cali the crew were given a runway change. The captain obtained permission and then attempted to go direct to a waypoint ("Rozo") to capture the newly assigned approach track to the airport. This action was a desirable alternative to having to turn left and effectively extend the flightpath to meet the route closer to its origination at the "Tulua" beacon. The automation, however, interpreted the captain's 'R' input not to refer to the Rozo beacon, but instead to a beacon 132 miles away near Bogota. The captain's instruction lead the aircraft to turn left towards high terrain. The airliner crashed minutes later while the crew were attempting to return to the airport at Cali.

In their analysis, the authors describe how careful consideration of the accident reveals that its underlying causes are not related to a loss of situation awareness on the part of a human crew unit in interaction with the automation, but it is the result of a more complex interaction of two (captain and first officer) different mindsets and sets of expectations. The computer controlled turn was consistent with the first officer's expectations of what had to be done to capture the approach route to the runway (from Tulua), whatever detailed computer input made the aircraft turn was irrelevant. At the same time, underlying this confirmed expectation and subsequent silence was the assumption that the captain shared the expectation of the turn. Only when the aircraft's turn became too pronounced to reasonably be considered as a change of course to capture the approach route did the first officer begin to comment and the human crewmembers began to realise all was not well.

The authors describe how situation awareness would appear in part to be a function of collaboration. They call the Cali accident a three-way breakdown in awareness. Each human crewmember has his/her own private access to the third "automated crewmember". The three-way breakdown evolved through an interplay between the two crewmembers and the automation. In Cali the two human crewmembers "develop and entertain different models about the world based on the cues they receive, pick out as relevant, and endow with meaning. Automation can interact with these different models and propagate them all the way to fully fledged and unrecoverable confusion" (Dekker & Orasanu, 1999 p. 76).

We agree with the essence of the authors' point. The human-automation divide is indeed much more intricate than first glance may suggest. However, in the light of our discussion, it may be more accurate to view the 'three-way' breakdown in awareness described by Dekker & Orasanu (1999) as a two-way breakdown in mutuality between human teamplayers. The crucial role that the automation played in this accident was to mediate a breakdown in mutuality between captain and first officer, and to exacerbate its effects, make more difficult its recovery through performing work in a manner inconsistent with that of a teamplayer.

CONCLUSION

The argument we have made here in some sense is not new. Norman (1990) compared interaction with automation to the fluidity of human conversation, but he did not analyse the nature of that difference in intentional and linguistic terms. Hutchins and Klausen (1991), on the other hand, describe the role that mutual knowledge and shared information (what they refer to has intersubjectivity) plays in the repair of conversations in the aircraft cockpit. The analysis we have presented here is, to our minds, entirely consistent with both of these pieces of work. What we have done however is to try to use these ideas to look at successful and unsuccessful interaction. Furthermore, we hope by doing so to draw attention to the limits of the automation as team player metaphor as a goal of automation design. Our discussion of mutuality and jointness has shown that the design question surrounding automated systems should not be 'how can we make automation a better teamplayer ?' since clearly with current levels of technology this is impossible. Instead the question should be 'how can we make automation compatible with the teamplaying context in which it will be deployed ?'.

ACKNOWLEDGEMENTS

Research was supported by the Esprit MEFISTO project, see
http://giove.cnuce.cnr.it/mefisto.html.

REFERENCES

AAIC (1996) Aircraft Accident Investigation Report 96-5: China Airlines Airbus Industrie A300B4-622R, B1816 Nagoya Airport, April 26, 1994, Aircraft Accident Investigation Commission, Ministry of Transport, Japan.

Brennan, S. E. (1998). The Grounding Problem in Conversations With and Through Computers. Social and Cognitive Approaches to Interpersonal Communication. Fussell, S.R and Kreuz, R.J. (Eds.). Mahwah, NJ USA, Lawrence Erlbaum Associates, Inc.

Clark, H. H. (1996). Using Language. Cambridge, England UK, Cambridge University Press.

Cook, R.I., Woods, D.D. & McDonald, J.S. (1991). Human Performance in Anesthesia : A corpus of cases. Cognitive Systems Engineering Laboratory Report, prepared for Anesthesia Patient Safety Foundation, April 1991 as referred to in Dekker & Woods, 1999.

Dekker, S.W.A. & Orasanu, J. (1999). Automation and Situation Awareness - Pushing the research frontier. In Coping with Computers in the Cockpit. Dekker, S.W.A. & Hollnagel, E. (Eds.) Brookfield, Vermont USA, Ashgate Publishing.

Fairburn, C., Wright, P.C., and Fields, R.E. (1999). Air Traffic Control as Distributed Joint Activity : Using Clark's theory of language to understand collaborative working in ATC. Proceedings of the European Conference on Cognitive Science (ECCS) Siena, Italy 1999.

Hollnagel, E. (1999). From Function Allocation to Function Congruence. In Coping with Computers in the Cockpit. Dekker, S.W.A. & Hollnagel, E. (Eds.) Brookfield, Vermont USA, Ashgate Publishing.

Hollnagel, E. & Bye, A. (2000) Principles for Modeling Function Allocation. International Journal of Human-Computer Studies, 52, 253-265.

Hollnagel, E. and Woods, D.D. (1983). Cognitive Systems Engineering : New wine in new bottles. International Journal of Man-Machine Studies, 18, 583-600.

Hughes, D. and Dornheim, M.A. (1995) Accidents Direct Focus on Cockpit Automation, Automated Cockpits Special Report Aviation Week and Space Technology, January 30 1995.

Hutchins E. and Klausen, T. (1991) Distributed cognition in the cockpit. In Engestrom, Y. and Middleton, D. Cognition and communication at work. Cambridge University Press.

Ladkin, P.B. (1996) Resume of the Final Report of the Aircraft Accident Investigation Committee into the 26th April 1994 crash of a China Air A300B4-622R at Nagoya Airport, Japan. See http://mek.ml.org/publications/Incidents/DOCS/ComAndRep/Nagoya/nagoya.html.

Malin, J.T, Schreckenghost, D.L., Woods, D.D., Scott, S., Potter, J.L., Hollaway, M. and Forus, K.D. (1991). Making Intelligent Systems Team Players : Case studies and design issues. NASA Technical Memorandum 104738.

Norman, D.A., (1990) The "problem" with automation: Inappropriate feedback and interaction, not "over-automation". In Broadbent, D.E., Baddeley, A.D., and Reason, J.; Human Factors in Hazardous situations. Philosophical Transactions of the Royal Society, London. pp. 585-593.

Sarter, N.B. and Woods, D.D. (1992). Pilot Interaction with Cockpit Automation: Operational experiences with the flight management system. International Journal of Aviation Psychology, 2 (4), 303-321.

Sarter, N.B. and Woods, D.D (1995). How In The World did we Ever get Into That Mode? Mode error awareness in supervisory control. Human Factors, 37 (1), 5-19.

Sarter, N.B. and Woods, D.D. (1997). Team Play With a Powerful and Independent Agent: operational experiences and automation surprises on the Airbus A-320. Human Factors, 39 (4), 553-569.

Schmidt, K. (1997). Of maps and scripts: The status of formal constructs in cooperative work. Proceedings of the Group'97 International Conference on Supporting Group Work, 138-147. New York: ACM.

Schmidt, K. and Simone, C. (1996). Coordination Mechanisms : Towards a conceptual foundation of CSCW systems design. Computer Supported Cooperative Work : the Journal of Collaborative Computing, 5, 155-200.

Strauss, A. (1988). The Articulation of Project Work: An organisational process. The Sociological Quarterly, 29 (2), 163-178.

Suchman, L.A. (1987). Plans and Situated Actions. Cambridge, England UK, Cambridge University Press.

Wenger, E. (1998). Communities of Practice : Learning, meaning and identity. Cambridge, England UK, Cambridge University Press.

Wright, P.C., Dearden, A. & Fields, B (In press) Function allocation: A perspective from studies of work practice. International Journal of Human-Computer Studies.

Integrated Representations for Task Modeling

Martijn van Welie, Gerrit van der Veer, Adriaan Koster

Division of Mathematics and Computer Science, FEW
Vrije UniversityAmsterdam
1081 HV Amsterdam
{martijn,gerrit}@cs.vu.nl
http://www.cs.vu.nl/~martijn/

ABSTRACT

During a task analysis much data is generated. Interpretation of the data and the development of knowledge are crucial to a successful task analysis. Task models are used to document and communicate the knowledge gained during the task analysis process. In order to describe all relevant aspects of the context of use, several representations are needed. This paper examines the requirements for such representations and proposes a set of complementary representations.

Keywords

Task modeling, task analysis, representations, diagrams, UML, contextual design.

INTRODUCTION

Knowledge elicitation and ethnographic workplace studies usually generate a lot of data and this data needs to be captured and well understood. If this data is not captured, the knowledge cannot be communicated to other members of the design team and consequently gets lost. Task modeling is the activity of transforming raw task and user related data or envisioning ideas into structured pieces of task knowledge. This knowledge is usually documented in a specification that uses several different representations. Each representation is intended to emphasize a certain aspect of this knowledge. Considering the complexity of the task world and the various possible views, it is clear that several different representations are needed. Ideally, the analysts have a collection of representations at hand that covers all aspects and views. At the same time, it is preferable that such a collection is kept small so that the designer does not drown in a plethora of overlapping representations. Representations must be useful and usable for designers. This paper discusses requirements for task modeling representations and defines a collection of representations that together cover most aspects of the knowledge that need to be documented, while minimizing the overlap.

CHOOSING REPRESENTATIONS

Task modeling for large cooperative systems can not be adequately handled by the classic task modeling techniques. Representations such as task trees are not well suited for describing the complex dynamics typically found in complex environments such as groupware systems. Additional representations are needed that can deal with aspects such as communication, coordination, social structures and work flow. However, many representations already exist for task modeling as well as other related modeling activities. Not all of them are useful in practice and the question is what makes a representation useful and usable. One aspect of a representation is that it should be effective. In (Macinlay 1986) Macinlay defines the effectiveness of a visual language as "whether a language exploits the capabilities of the output medium and the human visual system". This notion can be expanded to include purpose and audience i.e. what is the representation intended for and who is going to use it, because "visualizations are not useful without insight about their use, about their significance and limitations" (Petre, Blackwell, and Green 1997). Developing usable diagram techniques is difficult and requires insight in all of these aspects. In fact, one could say that usability is just as important for graphical representations as it is for user interfaces, both depending strongly on the context of use. Most of the research in this area is the field of information visualization (Card, Macinlay, and Shneiderman 1999) or visual design (Tufte 1990,Tufte 1983).

If we wish to choose representations, we must first distinguish several purposes for which they can be used and by whom (Britton and Jones 1999). Within task analysis the purposes of representations typically include:

1. To document and to communicate knowledge between designers.
2. To analyze work and to find bottlenecks and opportunities.
3. To structure thinking for individual designers.
4. To discuss aspects of the task world within the design team.
5. To propose changes or additions within the design team.
6. To compare alternatives in the design team or with a client.

Additionally we need some aspects that help discussing and comparing representations. For this discussion we will take the position that a representation essentially is a mapping of concepts and relationships (and possibly attributes) to the visual domain. Some aspects may concern the concepts and relationships while others concern the appropriateness of the mapping in relation to the purpose and audience. The following aspects are important to consider for representations:

Intended Purpose. For what purpose is the representation intended? Certain representations work well for communicating with clients while others only help structure a single designer's thought. Similarly, certain representations focus on time while others focus on structure. Effectiveness is reduced when the representation does not support the purpose.

Coverage. What concepts and relationships are involved? What information is shown and what is not? Is the information suitable for task analysis purposes? The information covered by a representation determines the view it supports.

Complexity. What is the complexity of the representations in terms of the number of concepts and relationships that are shown? If the complexity is high the understandability is usually low, which is probably not desirable.

Understandability. How well can the representation be understood? Understandability concerns how successful the concepts and relationships have been mapped to a graphical representation. Representations should be easy to understand for the intended audience/users. If not, they will not be used. Stakeholders may come from different disciplines which makes a common understanding more difficult to reach. Other aspects such as visibility also play a large role i.e. how easy parts of the representations can be distinguished or what kind of first impression a representation gives.

Intended Audience. Who is going to use the representation? Certain representations are more familiar to designers from different disciplines than others. Also clients and other stakeholders may be familiar with certain representations. For example, UML is familiar to most Software Engineers while unfamiliar to ethnographers.

MULTIPLE VIEWPOINT

Since multiple representations are needed, it first needs to be clear which viewpoints need to be covered. In our task analysis method GTA (van der Veer, Lenting, and Bergevoet 1996), several views are defined. Constructing a set of representations means selecting a set that covers most aspects of the three views in GTA. When the GTA viewpoints are translated we can distinguish several requirements for the whole set:

- Should describe the structure of the work including the tasks, goals and roles;
- Should describe the dynamics of the work in time;
- Should describe the use of objects, tools and other artifacts;
- Should describe the physical environment of the work.

This does not mean however, that for each aspect a single representation is needed or sufficient. Moreover, representations are preferably easy to understand by engineers, ergonomists, psychologists, requiring a trade-off between representational power and comprehensiveness. These representations are targeted for the analysts and designers and not for communicating with the client which requires other representations such as scenarios and use cases. In (Killich, Luczak, Schlick, Weissenbach, Wiedenmaier, and Ziegler 1999) another set of requirements for task models is given. On the basis of their requirements, they conclude that UML (Rumbaugh, Jacobson, and Booch 1997) contains the most complete set of representations that meet the requirements. However, they do not provide a complete set that meets all requirements. Another well known set of representations is given by means of Contextual Design's work models (Beyer and Holtzblatt 1998).

Representing Work Structure

Work structure is concerned with the structure of the work and how it is allocated to people. Work structure is usually represented using task trees that show a hierarchical decomposition of the work. Often some timing information is added using constructors such as SEQ, LOOP, PAR and OR. The constructors cannot always be used especially when the task sequence uses a combination of sequential and optional tasks (van Welie, van der Veer, and Eliëns 1998b). Details of the task can effectively be described using templates. Details include the state changes, frequency and duration, triggering and start/stop conditions. Tasks also need to be explicitly related to the goals and the roles that perform the tasks. When the task structure is viewed in relation to roles the UML Collaboration diagram is useful, because task trees cannot cope well with different roles and each role consequently gets its own tree.

Representing Work Dynamics

Especially when multiple roles are involved in a certain task, timing and changes in control are essential. A task tree cannot represent this well. Workflow or activity models are needed to capture this aspect. Work dynamics involve the sequence in which tasks are performed in relation to the roles that perform them. Additionally, parallel and optional tasks should be modeled, especially when the sequence depends on decisions. Cooperation and communication can partially be described using activity models. Usually scenarios (or use cases) are described in such workflow diagrams. A scenario is triggered by some event and starts with some goal being activated. The scenario usually ends when the goal is achieved but other goals may have been activated in the course of tasks and may not be reached yet. In case studies such as (van Loo, van der Veer, and van Welie 1999) it turned out that this event driven dynamic aspect of cooperative work can be very important.

Representing Tools and Artifacts

The work environment itself usually contains many objects (a hundred or more is not unusual) some of which are used directly in tasks and other that may be "just lying around". The objects can be tools that people use either in software or in hardware but other objects may be directly manipulated in tasks. For some of these objects it may be relevant to describe them in detail. Details may include their structure, their type, and object specific attributes.

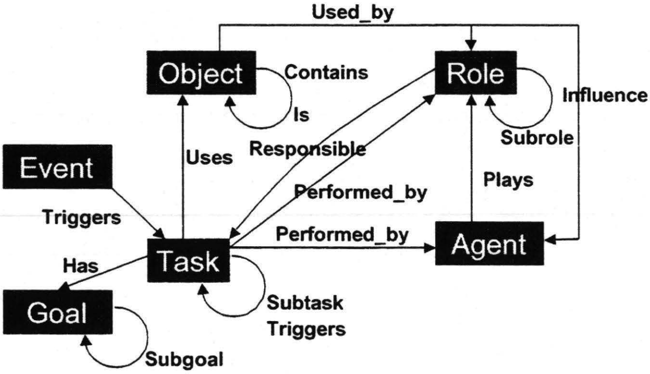
Representing the Work Environment

Often it is important to know what the physical layout of the work environment is. Pictures, drawings, maps and video clips can capture parts of this information. Usually maps or drawings are annotated with comments relating to their impact on the work such as reachability of objects. Most objects that appear in such representations are also represented when modeling Tools and Artifacts. Additionally, the social structure or the culture of the work environment needs to be represented. People rarely perform their work in solitude and social relationships influence work. Typically, roles influence other roles with certain strengths and there are certainly attitudes to be found between them.

INTEGRATING REPRESENTATIONS

The aspects of the different viewpoints and representations need to be semantically integrated. In any set of representations, it holds that each representation is a view of the same task world model where the same elements may appear in different representations. Our task world ontology (van Welie, van der Veer, and Eliëns 1998b), see Figure 1, captures most of the fundamental concepts and relationship that form the basis of all the representations.

Figure 1: The Task World Ontology



However, some representations such as videos are not very clearly structured and hence, are not covered by the ontology. For other representations such as the physical layout, this is theoretically possible but in practice this is not useful to do. In the next sections we will use the concepts and relationships to define what is shown in the representations.

A COLLECTION OF ONTOLOGY BASED REPRESENTATIONS

It is clear that some are more useful/usable than others and that improvements can be made. In this section, we will define a collection of representations that cover the views as defined in the previous sections. This collection of coherent representations is an attempt to provide a more useful collection of representations for practitioners. For each of the views we will define one or more representations that form a useful "package" for that view. Together, the representations can form a practical tool set for the designer. The representations are based on existing representations but include some additions or modifications to make them more usable and useful for task modeling.

Constructing a set of Representations

The collection of representations that is discussed in the next sections combines several existing representations. Additionally, some modifications have been made. Compared to Contextual Design's (Beyer and Holtzblatt 1998) work models (CWM), the main differences are:

- The CWM sequence model is replaced by a workflow model similar to the UML Activity diagram;
- The CWM sequence and CWM flow model are combined into one representation;
- Decomposition trees are added;
- The CWM cultural model has been redesigned;
- The number of concepts is larger than in CWM.

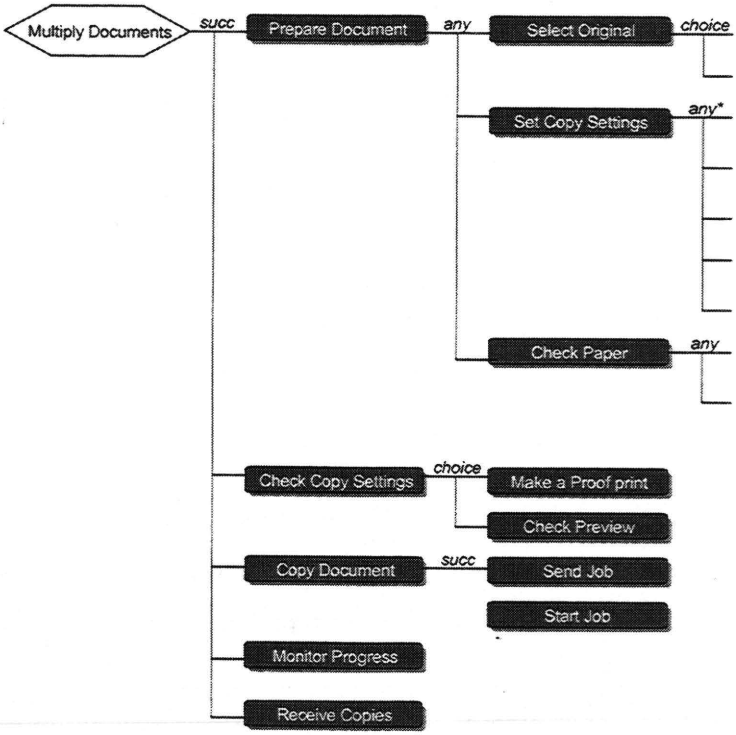
Compared to UML (Rumbaugh, Jacobson, and Booch 1997), we use a modification of the Activity Model. We have added and event and goals lane as well as changed representations for parallelism and choice.

Modeling the Work Structure

The purpose of the work structure model is to represent how people divide their work into smaller meaningful pieces in order to achieve certain goals. Knowing the structure of work allows the designers to understand how people think about their work, to see where problems arise and how tasks are related to the user's goals. The relation between tasks and goals helps the designers to choose which tasks need to be supported by the system and why i.e. which user goals are independent of the technology used.

For modeling work structure the task decomposition tree has proven to be useful and usable in practice. The tree is essentially based on the subtask relationship between tasks. Besides tasks, goals can also be incorporated. At the highest level a tree can start with a goal and sub-goals and then proceed with tasks and subtasks. In that case the sub-goal and has relationship are also used. A task decomposition is modeled from the viewpoint of one role or goal. If complex systems are modeled, several task trees are needed to describe the work for all the roles. It then becomes difficult to see how work is interleaved. Trees normally contain a time ordering using constructors from top to bottom or left to right, depending on the way the tree is drawn. The inclusion of time information can be insightful but it is often also problematic. ConcurTaskTrees (Paterno, Mancini, and Meniconi 1997) use LOTOS operators, which are probably the best-defined time operators. On the other hand, it is not always necessary to be very precise in everything that is modeled. Designers will typically model that certain tasks occur sometimes or almost never. In our opinion, including some time information is useful but this kind of information is better represented in a work flow model if precision is required.

Figure 2: Representing work structure



If time is included, then a number of time operators are plausible. In our experience, it is useful to have a set of standard operators while also allowing designers to create their own operators when needed. For the average usage, the following time relationships have proven sufficient:

- Concurrent. The tasks occur concurrently;
- Choice. One out of a set of tasks is done;
- AnyOrder. All tasks of a set of tasks are done in no fixed order;
- Successive. One task is followed by another.;
- Any. Zero or more tasks of a set of tasks are done in no fixed order;
- combined with other constructor. Used to express iteration.

In the work structure model, the root of the tree is a goal with possibly some sub-goals. Connected to goals are tasks that are represented as rounded rectangles. The tree is drawn from left to right instead of top-to-bottom for more economical use of space, especially when trees become large, see Figure 2. Other aspects of work structure include role structures and the relationships with tasks. For role structures trees can also be used. When used to show goal or role hierarchies the time constructors are not used.

Modeling the Work Dynamics

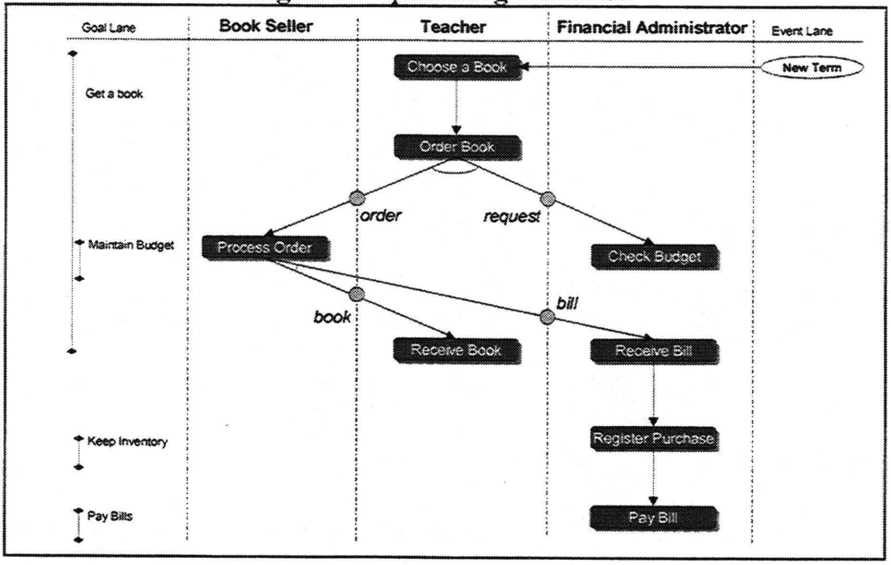
The purpose of the work dynamics model is to show work in relation to time and roles. The model gives the designer insight in the order in which tasks are performed and how different people are involved in them. Additionally, it can show how people work together and communicate by exchanging objects or messages. Typically, such a flow model describes a small scenario involving one or more roles. This way, it shows how work is interleaved.

The flow model specified here is a variation on the UML Activity graph. We included events and goals to make it more suitable for task analysis. Additionally, the representations of the time operators have been modified to be more appealing. This way the collaboration diagram (or Contextual Design's Flow Model) is not needed anymore since the information has been combined in one representation. Each flow model describes a scenario that is triggered by an event. Work usually does not start by itself but instead is often highly event driven (van der Veer, van Welie, and Thorborg 1999). The event is represented by an oval which is connected to the first task. The sequence of tasks is given using a Concurrent operator or a Choice operator and not any of the other operators as suggested for the structure model. Tasks can optionally be arranged in swim lanes, one for each role. Objects can be passed between tasks that have different roles and are drawn on the border of the adjacent swim lanes. When needed, goals can also be added to this representation. With a certain task a new goal can get "activated" until it is "reached" in a later task. The goals are written in the first column with vertical lines to show how long they are activated.

The flow model does not show hierarchical relationships between tasks and a flow model can only use tasks that are hierarchically on the same level. For subtasks, a new flow model needs to be specified. The addition of the goal lane can show many useful aspects when analyzing the workflow. For example, Figure 3 shows that once the "teacher" has received the book his goal is achieved but the scenario is not finished yet.

In terms of the ontology, the flow model is based on the concepts Event, Task, Object and Role. The relationships used are triggers, responsible, and uses. The operators are Concurrent, Choice, and Successive which are parameters in the triggers relationship. The AnyOrder and Any constructors are not valid in this representation and the Successive operator is implicit in the direction of the arrows. For objects that are being passed between roles it holds that each object must be associated to both tasks with the uses relationship. Note that the objects that are used in one task are not shown in the representations.

Figure 3: Representing Work Flow



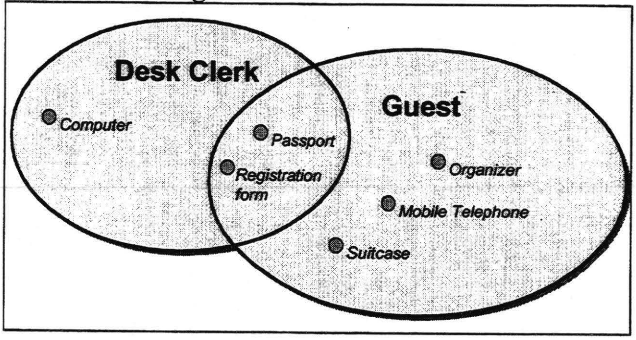
Iteration is not specified in the flow diagram. If a task is done several times, an asterisk can be used to indicate that the task and it subtasks are done several times. However, usually iterations are specified in the Work structure model. Iteration is specified on subtasks and not tasks on the same level, which are shown in the Work Flow model.

Modeling the Work Artifacts

The artifact model shows two relationships between objects: the containment and type relationship. For both a tree diagram is used. The objects themselves can be annotated with their attributes or their visual appearance. In order to express containment and type, the UML class diagram notation can be used. However, it is important to remember that we are only modeling objects that are relevant to the user and not any irrelevant internal system objects. To some this may suggest that the task model describes an object oriented system model, which is not the case.

The use of objects in tasks is partly covered by the Work Flow model. The Work Artifacts model the structural aspects of the objects. However, objects may also be connected to their users i.e. roles or agents, instead of to the tasks where they are used. In such a diagram, the users are represented as ovals and the objects are labeled dots within the ovals. The ovals may overlap if more than one user uses the object (see Figure 4).

Figure 4 Artifacts and roles



Modeling the Work Environment

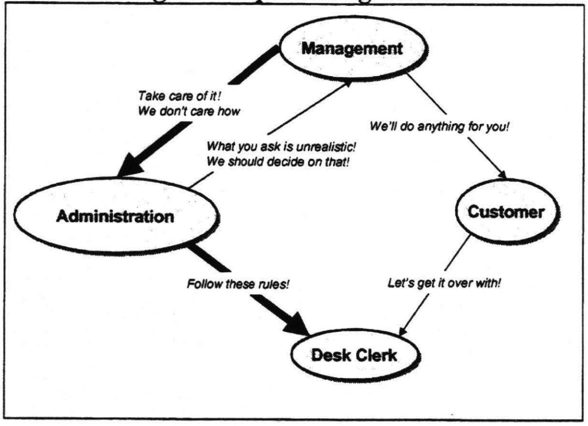
The environment model describes two aspects of the environment. Firstly, the physical layout of the environment and secondly, the culture within the environment. The physical model is simply described by one or more annotated "maps" of the environment. The purpose is to show where objects are located in relation to each other. The objects are those that are relevant for the work and also those that are in the same space. Such layout diagrams can easily be drawn using commercial drawing software such as Visio.

The other model is the Culture Model. The culture model we describe here is an adaptation of the culture model from Contextual Design. In Contextual Design the roles are represented in overlapping circles. However, overlapping of circles does not have any meaning although it suggests that there is one. Hence, we adapted the model (see Figure 5). We define the culture model as follows:

- Roles are represented as ovals;
- The ovals are connected by arrows if there is a force between roles. The relative strength of the force is depicted in the width of the arrow and forces can be bi-directional;
- Forces are annotated with attitudes of the force relationship.

In some cases, a force applies to more than one role. By drawing an extra circle around roles, a force can indicate one-to-many forces which can typically be used to describe "corporate culture".

Figure 5 Representing Culture



STATIC VS. DYNAMIC REPRESENTATIONS

All of the representations discussed in the previous sections are static. However, representations can also be more dynamic. Traditionally, a representation is static i.e. it does not change after it is drawn and is designed for use on paper. However, it is often convenient to emphasize a certain aspect in the representation. When software is used to draw the representations, the representations can be changed dynamically. In (Card, Macinlay, and Shneiderman 1999) they are called active diagrams. For example, one could easily switch between a flow model with or without swim lanes. Alternatively, it could be possible to add some extra information by marking tasks as "problematic" or "uses object X", see (van Welie, van der Veer, and Eliëns 1998a). Such annotations are often done by designers to explain certain aspects to others during a presentation or in documentation. In software we are already very much used to active diagrams and they occur in scrolling, zooming and syntax highlighting. This asks for a more flexible view on what constitutes a representation and when a representation can be modified. The dynamic aspects could be controlled manually by the viewer but could also be pre-specified using a function in which case we usually speak of animation. Now that it becomes increasingly easier to create dynamic representations it is important to understand when and how they could be applied usefully in design.

In task modeling, animation is a way to create more dynamic representations. Animation can be used in simulations of scenarios or task models (Biere, Bomsdorf, and Szwillus). Using simulations an analyst can step through a scenario and get a different feel for what goes on. Other purposes might be to "debug" a task model which is particularly useful for envisioned task models.

CONCLUSIONS

In this paper, the need for a set of integrated representations has been discussed. Such a set has been defined and shown with examples. Most of the representations can be integrated using a task world ontology. The task world ontology serves as the underlying model for the representations. Together, the representations cover several important viewpoints for task modeling.

REFERENCES

Beyer, H. and Holtzblatt, K. (1998), Contextual Design, Morgan Kaufmann Publishers, Biere, M., Bomsdorf, B., and Szwillus, G. The Visual Task Model Builder, Proceedings of CADUI'99, Louvain-la-Neuve, Belgium.

Britton, C. and Jones, S. (1999) The Untrained Eye: How Language for Software Specification Support Understanding in Untrained Users, Human-Computer Interaction 14, 191-244.

Card, S., Macinlay, J. and Shneiderman, B. (1999), Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann Publishers,

Killich, S., Luczak, H., Schlick, C., Weissenbach, M., Wiedenmaier, S. and Ziegler, J. (1999) Task Modelling for cooperative work, Behaviour & Information Technology 18, 325-338.

Macinlay, J. (1986) Automating the Design of Graphical Presentations of Relational Information, ACM Transactions on Graphics 5, 110-141.

Paterno, F. D., Mancini, C., and Meniconi, S. (1997), ConcurTaskTrees: A Diagrammatic Notation for Specifying Task Models, Proceedings of Interact '97, Sydney, Chapman & Hall.

Petre, M., Blackwell, A. and Green, T. (1997), Cognitive Questions in Software Visualisation, in: Stasko, J., Domingue, J., Price, B., and Brown, M., Software Visualization: Programming As a Multi-Media Experience , MIT Press.

Rumbaugh, J., Jacobson, I. and Booch, G. (1997), Unified Modeling Language Reference Manual, Addison Wesley,

Tufte, E. (1983), The Visual Display of Quantative Information, Graphics Press, Connecticut

Tufte, E. (1990), Envisioning Information, Graphics Press, Connecticut

van der Veer, G.C., Lenting, B.F. and Bergevoet, B.A.J. (1996), GTA: Groupware Task Analysis - Modeling Complexity, Acta Psychologica, no. 91, pp.297-322.

van der Veer, G.C., van Welie, M. and Thorborg, D. (1999) Modeling Complex Processes in GTA, Ergonomics 42, 1572-1587.

van Loo, Reinard, van der Veer, G. C., and van Welie, M. (1999), Groupware Task Analysis in Practice: a scientific approach meets security problems, 7th European Conference on Cognitive Science Approaches to Process Control, Villeneuve d'Ascq, France .

van Welie, M., van der Veer, G. C., and Eliëns, A. (1998a), Euterpe - Tool support for analyzing cooperative environments, Ninth European Conference on Cognitive Ergonomics, Limerick, Ireland.

van Welie, M., van der Veer, G. C., and Eliëns, A. (1998b), An Ontology for Task World Models, Proceedings of DSV-IS98, Abingdon UK, Springer-Verlag, Wien.

Coping with Uncertainty in Temporal Planning and Scheduling

Toni Wäfler

Institute of Work Psychology
Swiss Federal Institute of Technology (ETH)
CH-8092 Zurich, Switzerland
waefler@ifap.bepr.ethz.ch
http://www.ifap.bepr.ethz.ch

ABSTRACT

Temporal planning and scheduling in industrial organisations is understood as a task performed within sociotechnical systems. Nevertheless it is argued that the 'classical' sociotechnical systems approach fails in supporting design of temporal planning and scheduling systems, because it mainly focuses on demarcating tasks and allocating them to primary work systems and to individuals in a meaningful way. It gives little advice on how to design the co-ordination of such demarcated tasks, which is the aim of temporal planning and scheduling. To compensate for that shortage the concept of secondary work systems is proposed together with design criteria. The concept and the criteria are confronted with reality in an exploratory case study.

Keywords

Planning, scheduling, job design, sociotechnical system design, uncertainty.

INTRODUCTION

In a world of increasingly turbulent markets and growing technical complexity of products and production processes it is a core problem of work systems to cope with uncertainty. Many different activities are performed within work systems to deal with this problem. Temporal planning and scheduling are part of these activities, meant to cope with uncertainties that arise in matching dynamic production demands with unstable production resources.

Identifying planning and scheduling

Key characteristics of temporal planning and scheduling are (cf. e.g. McKay, 1987, 1992; Schüpbach, 1996; Wiers, 1997):

- Information to be processed is incomplete, ambiguous, dynamic and of stochastic nature;
- information flow follows feed forward and feed back structures as well as formal and informal structures;
- decisions to be taken are highly interrelated not only in content but also time wise;
- decisions have to be taken in a situated way;
- goals to be followed are - even if set clearly - not independent: minimizing lead times as well as machine idle time are contradictory;
- information processing and decision-making is distributed among many different (human and non-human) actors;
- result oriented performance measurement and even more process oriented evaluation of planning and scheduling practices are constrained due to temporal delays between actions and effects as well as unclear mutual relations.

These characteristics indicate that temporal planning and scheduling cannot be considered as a task that can be isolated and allocated to a clearly defined agent be it an individual (e.g. a human scheduler), an organizational unit (e.g. a scheduling department), or a sophisticated technology (e.g. an ERP-system). It must rather be considered as a process that takes place within a complex system consisting of humans, organizational structures and technology that are highly interrelated. The system as a whole is a network of information flow and decision-making. Hence it can be considered a distributed human-computer system. In this system human-computer-interaction as well as human-human-interaction occurs. Both are subject of planning and scheduling research.

On the level of human-computer-interaction it was found that human-computer systems often outperform human or computer individually (e.g. Sanderson, 1989). Therefore it is pleaded for hybrid scheduling systems. On the level of human-human-interaction McKay (1987, p133f) states: "The informal scheduling system is the 'realtime' scheduling system ... The scheduling hierarchy is not straightforward and contains many interdepartmental communications, which are affected by the departments, the individuals, the current world, and the work involved. ... Scheduling duties are not clear; overlaps occur and organizational positions do not necessarily reflect duties, responsibility, and authority. ... Scheduling decision points are not isolated as various levels of scheduling are distributed throughout the organisation."

So far, it can be concluded that temporal planning and scheduling systems can be considered as sociotechnical systems (cf. Rice, 1958; Emery, 1959) performing a cognitive task.

The design of planning and scheduling

A direct adoption of 'classical' sociotechnical design criteria (cf. e.g. Strohm, 1998) to temporal planning and scheduling systems turned out to be difficult (Wäfler, in press). This is because the 'classical' sociotechnical design criteria (mainly) aim at clustering functions to meaningful tasks for individuals and organisational units respectively, on the base of normative considerations. Accordingly, individuals should e.g. be provided with complete work tasks (Ulich, 1998) in order to allow for task orientation (Emery, 1959), whereas organisational units are e.g. to be designed in a way making them as independent of each other as possible (Strohm, 1998), in order to allow for local regulation of variances and disturbances (Susman, 1976). Such criteria for task design require demarcatable tasks on the level of the individual as well as on the level of the organisational unit. Temporal planning and scheduling due to its characteristics (see above) does not meet this precondition. Consequently it is difficult to provide planners as well as executors with complete tasks or to make planning and executing units independent of each other. Therefore it is proposed to supplement the 'classical' approach to sociotechnical systems design with the concept of secondary work systems as well as with specific design criteria.

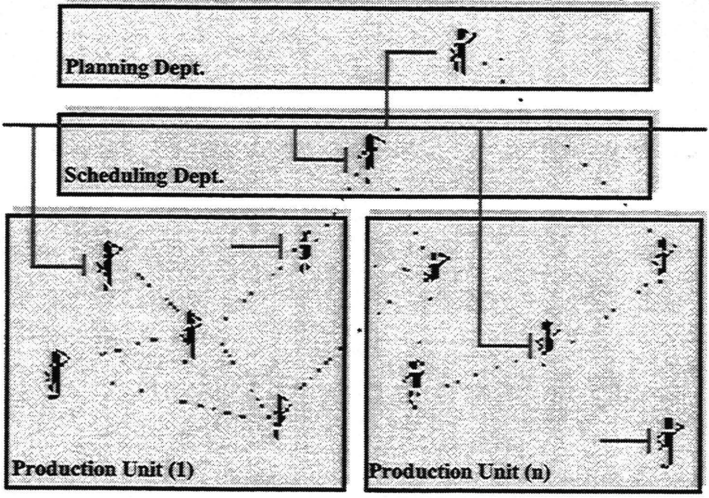
Secondary work systems

In the sociotechnical systems approach organisational units are named 'primary work systems' (Trist, 1981; Strohm, 1996). Primary work systems are organisational units. Each primary work system has originally been created to fulfil a certain task. This is its primary task (Rice, 1958). Moreover, each work system has to perform secondary tasks. These comprise everything that must additionally be performed in order to be able to fulfil the primary task. Production units as well as scheduling departments have a primary and a secondary task (Strohm, 1996). Whereas the primary task of a production unit is the manufacturing of certain workpieces, its secondary task includes securing the system's preservation (e.g.: maintenance, training etc.) as well as regulation (e.g.: input control, internal co-ordination). The primary task of a scheduling department might be the elaboration of 'good' schedules. Its secondary task includes training of its people, maintenance of its tools, planning and co-ordination of its own resources etc.

There is one major difference between these two types of primary systems. The primary task of production units stands on its own while the primary task of the scheduling department relates to the production units' secondary task (i.e. order flow co-ordination, boundary regulation). This difference is important, because primary work tasks related to the secondary work tasks of other primary work tasks cannot easily be demarcated. Criteria for organisational design must take this difference into account. As seen above, classical sociotechnical system design does not make this difference.

In order to avoid the necessity to demarcate planning and scheduling tasks the concept of secondary work systems is proposed (Wäfler, in press). A secondary work system overlaps and penetrates primary work systems. Figure 1 illustrates the relation between primary and secondary work systems. It shows four primary work systems that are interrelated by a network of people and technical devices. All these people are members of one of the primary work systems. Within the network information flow and decision-making takes place in technical (full lines) as well as in social (dashed lines) relations. Planning and scheduling processes take place within such sociotechnical networks that consist of all humans and technical devices that perform any activity in the planning and scheduling process. It comprises far more than the formal planning and scheduling departments. It is a distributed joint cognitive system (cf. Hollnagel & Cacciabue, 1999).

Figure 1: A secondary work system that interrelates four primary work systems



This kind of network is named secondary work system because it performs the (secondary work) task of co-ordinating the order flow through the production unit as well as the boundary regulation between production units. Thinking in secondary work systems has the advantage that planning and scheduling must not be demarcated and allocated to separated primary work systems.

Design criteria

The aim of temporal planning and scheduling is to deal with uncertainties that arise in matching dynamic production demands with unstable production resources. Two fundamentally different strategies to handle uncertainty can be distinguished: minimising uncertainty and coping with uncertainty (Grote, 1997). Basic principles of minimising uncertainty are:

- Complex, centralised planning and scheduling systems;

- reduction of scope of operative action by means of regulation and automation;
- disturbances seen as preventable symptoms of inefficient system design

Basic principles of coping with uncertainty are:

- Planning as resource for situated action;
- maximising operative degrees of freedom through complete tasks and lateral co-operation;
- disturbances seen as opportunity for use and development of competencies and for system change.

The present work follows the assumption that an attempt to minimise uncertainty makes a system inflexible (cf. e.g. Schüpbach, 1996) thereby causing new uncertainty (cf. e.g. Wiers, 1997). Consequently, the strategy of coping with uncertainty will be focused. Following, criteria for designing temporal planning and scheduling as a secondary work system will be proposed. These criteria refer to the three levels (a) organising planning and scheduling, (b) individual planning and scheduling task, and (c) human-computer-interaction in planning and scheduling (cf. Wäfler (in press) for a detailed reasoning).

Organising planning and scheduling

With reference to the organisation of temporal planning and scheduling processes main criticisms stated by many authors concern the classical tayloristic structure that causes a separation of thinking and doing. Suggestions on how to overcome organisational deficiencies related to the hierarchical, control-oriented approach mainly concern the assignment of autonomy and control (Grote, 1997; Grote et al., 2000; McKay, 1992; Schüpbach, 1996; Wiers, 1997), the organisation of the information flow (Grote, 1997; Hoc, 1988; McKay, 1992; Schüpbach, 1996) and the simultaneity of decision-making (McKay, 1992; Schüpbach, 1996; Suchman, 1987):

- Autonomy and control must be distributed throughout the manufacturing process;
- information flow needs to be multidirectional in both senses, vertical as well as horizontal;
- decisions should be taken simultaneously at the meaningful points of decision.

Individual planning and scheduling task

The opportunity to plan one's own actions regarding both aspects, temporal as well as procedural planning, is a significant aspect of task completeness. People who plan their actions reach higher performance with reduced work load and emotional load than people who just act reactively (Hacker, 1998). The opportunity to act on the basis of self-planning has a higher effect on performance than intelligence (Battmann, 1984) or memory capacity (Wiesner, 1995). However, it is in the nature of sociotechnical planning and scheduling systems that total completeness of individual tasks is not possible. Nevertheless, in order to make possible opportunistic planning (Hayes-Roth & Hayes-Roth, 1979) and situated acting (Suchman, 1987) as well as to compensate for limits in planning for others (Dahme 1997; Resch, 1988; Zölch, 1997) two aspects must be considered carefully:

- Adequate autonomy and opportunities of control must be allocated to the executors as well as to the planners;
- mutual knowledge must be guaranteed in order to overcome hindrances in co-ordination.

HCI in planning and scheduling

Temporal planning and scheduling be it considered as an individual or as a multi-person process is subject to automation also. A lot of effort is put into the design of sophisticated information processing technology that is meant to either support or to replace the human planner or scheduler.

The implementation of systems of both approaches mostly does not lead to the expected results (Ulich, 1998). As a consequence many systems are not really used by the human planners and schedulers (Wiers, 1997). Support systems fail to support co-ordination of jobs because they do not provide the users with complete information, with clarity of temporal requirements, and with actual representations of process states (Zölch, 1997). Systems that aim at replacing the human are criticised more fundamentally. Sanderson (1989) has described the deficiencies of approaches to scheduling automation. Her criticism mainly concerns the simplifying character of models and the incompleteness of knowledge bases and inference engines.

In the design of the technical part of planning and scheduling system two aspects must be considered:

- Technology supports efficient job co-ordination only if it provides the user with complete and actual information (e.g. Zölch, 1997);
- human-computer interaction must be designed following a complementary approach in order to make possible human control over automated planning and scheduling processes (Grote et al., 2000).

CASE STUDY

The case study reported below has been carried out in a Swiss SME. The objective of the case study is firstly to identify and to describe the company's temporal planning and scheduling system in terms of a secondary work system. Secondly, the company's temporal planning and scheduling system is investigated with reference to the requirements of an autonomy-oriented design as elaborated in the previous sections.

Method

An explorative case study has been chosen for several reasons. These mainly concern the nature of information that is processed in a sociotechnical planning and scheduling system as well as the yet unknown relations between causes and effects. Semi-structured interviews, whole-shift observations, and written examination tools have been employed to investigate the company's planning and scheduling system. A total of twenty-two people from different hierarchical levels and with different functions have been investigated in that way.

Results

The secondary work system of temporal planning and scheduling

The main task of the planners is to adjust the ERP-software's parameters in order to reach certain goals (e.g. low stock, service level). Their main information sources are the salespeople and the historical data. The ERP-software makes job proposals to the schedulers, whose task it is to check the proposals' plausibility, to check the readiness to fulfil the proposed job (e.g. availability of material and production resources), and to release it (including make-or-buy-decisions). The manufacturing units are controlled centrally by the scheduler, in a top-down cascaded way via supervisors and foremen down to the workers. The assembling units get their jobs directly from the ERP-software. The foremen do the scheduling for their work groups. However, assembling is highly standardised such that the foremen require no critical decision-making.

Organisational aspects

Temporal planning and scheduling is performed in close co-operation of planners and schedulers who are located in the same office. Whereas planners care more about long- and middle-term planning, schedulers concentrate on the scheduling. Short-term planning is performed jointly. Multi-directionality of information flow as well as mutual constraining is guaranteed by permanent communication and joint decision-making regarding short-term planning. By checking the plausibility of automatically generated order proposals the schedulers implicitly check the usefulness of the system's parameters that are fixed by the planners. The opportunity to request alterations of parameters and the readiness of the planners to do so in mutual agreement as well as the inclusion of the schedulers in feasibility-decisions regarding special orders allows for factual bottom-up constraining. The schedulers' rather broad scope of decision-making allows for simultaneous decision-making.

The interaction between the schedulers and the shop floor is quite different. The schedulers strongly constrain the manufacturing units. They try to optimise the schedules for those resources they consider to be the bottlenecks. For these resources only the jobs of the actual day are released with detailed prescriptions regarding sequencing. Although the ERP-system provides the scheduler with order proposals for several weeks, the schedulers hold them back, even if they are plausible. They do this in order to keep flexibility, as they never know whether rush-orders will make their schedules obsolete. But the following statements of schedulers indicate also a need for central control:

- "... to have control that they really do the work according to the dates we have set."
- "I think there will be a problem with a certain machine because I know what might come next."

As a consequence, lateral co-ordination between the units does not take place. On the one hand the schedulers criticise this situation. On the other hand the shop floor supervisors complain that they do not have the information base to actively co-ordinate with their colleagues. Schedulers state:

- "I must do the co-ordination between the units."
- "Although they sit next door, he calls me and asks me to look after parts he needs from his colleague."

At the same time the shop floor supervisors' statements reflect poor lateral co-ordination opportunities:

- "The due dates on the schedules correspond poorly with the due dates of the next unit."
- "Where the next operations will be performed is not really clear from the documents."
- "I don't know about the others' problems."

The centralised decision-making leads clearly to the fact that everybody delegates responsibility to the schedulers. Bottom-up information flows as formally prescribed, i.e. the shop floor (mostly) gives feedback on job completion through the ERP-system. This information does not nearly cover the information need of the schedulers. They are permanently gathering information by the means of walking trough the factory and making intensive use of their telephone. In order to get an up-dated mental representation of the current state of production there are many hints they pay attention to. Remaining up-dated is a rather hard job, as can be illustrated by an incident that occurred during the observation of a scheduler. Confronted with a large rush order to be delivered as soon as possible he was asked by a planner to determine a realistic delivery-date that could be communicated to the customer. The scheduler set a date five weeks ahead. Afterwards he told the observer that the actual production would take only one week: "But I can only rush up to three orders through the organisation at once. Otherwise I would be permanently phoning and have no time to do anything else."

Top-down information flow from the schedulers to the shop floor is restricted to the communication of decisions. Consequently, the shop floor's comprehension of goals and working of the scheduling departments is low as statements show:

"I have no idea of the procedures in the planning and scheduling departments, I could only guess from statements they make on how they might work."
"I have never been up there."

The interaction between the schedulers and the assembling units is slightly different. As mentioned above the assembling units directly get the orders out of the ERP-system. They schedule their orders themselves. However, there is not really a need to schedule the orders. Priorities are pre-determined by delivering dates. The whole order is completely assembled in one unit. Hence (almost) no co-ordination with other units is needed. Set-up times are little and consequently do not make a co-ordination of jobs necessary. Components are available on site, controlled by a KANBAN-system. If there are any disturbances the assembling units' foremen delegate the problem to the schedulers or to the head of the interim stock, who himself chases for jobs in the manufacturing units or asks the schedulers for help.

In the relation between the scheduling departments and the shop floor, autonomy is centralised, there is no lateral information flow (e.g. there is no direct mutual co-ordination between production and assembling units), constraining only goes top-down, and simultaneous decision-making has not been observed.

Individual planning tasks

The execution of planning has been observed with planners and schedulers as well as with some shop floor supervisors. All of them take a lot of planning-related decisions intuitively based on their experience. At the same time the following statements show, that none of them really have the opportunity to assess the quality of decisions:

- "The quality of single decisions cannot be seen." (planner)
- "As long as nobody complains." (shop floor supervisor)
- "You get a 'Thank you' when it works." (shop floor supervisor)
- "If I don't know what I don't know, I cannot ask." (shop floor supervisor)

However, planners seem to plan most systematically. By the means of spread-sheets which they have programmed themselves they try to compute optimal parameter adjustments on the base of historical data and forecasts. Due to the nature of their task immediate feedback is poor. Conclusions regarding the quality of decisions can only indirectly be drawn from the observation of trends in the development of stocks, service levels, and the like.

Sandwiched between planning and executing the schedulers have to plan opportunistically because it is they who (try to) match the plan with reality. Scheduling is a multidimensional problem: "We don't want to produce too early. But we always want to have enough stock of every product. We don't want to produce anything we don't need afterwards. We want to have it when we need it. But we also want, that there is always something to do on the shop floor. We don't want to force them to work like crazy in spring and to have nothing to do in autumn. So ... there are many different things we have to consider."

To reduce the scheduling problem's complexity and to keep in (imaginary) control, the schedulers are forced to simplify the problem by reducing the variables that are considered, and - of course - by constraining the executing units on the shop floor. Simplification could for instance be reached by non-consideration of capacities: "... we just don't look at capacities. If a capacity problem occurs we look for other solutions: activate other machines, try to get it produced externally." Constraining goes as far as scheduling is clearly seen as task of the schedulers, by the schedulers themselves as well as by the shop floor supervisors

The shop floor supervisors react in different ways to the fact that scheduling is clearly assigned to the schedulers. Some of them are just reactive, refusing any responsibility. They heavily rely on the instructions they get directly from the schedulers. They do not (even try to) schedule according to the information they get from the documents, because they made the experience that this information is not reliable, and that - in case of lack of clarity - it is better to ask the schedulers. On the other side, some (but not many) of the supervisors also try to be more active.

All in all expected symptoms of centralised scheduling could be observed. Planners and schedulers are (at least partly) overloaded when attempting to control the whole manufacturing process. Hence they mostly decide reactively. Proactive deciding could be observed - if at all - in attempts to gain time by introducing buffers. The shop floor supervisors on the other side show reactive behaviour too. Most of them do not even see it as their task to schedule but only to execute prescribed schedules. Those who try to act situated, lack the needed information.

Human-computer interaction

On the level of human-computer interaction it must be stated that the implemented ERP-system is mainly used as an information system. There are no algorithms for planning and scheduling. In fact, the system does make proposals on jobs to be released. But this process is under human control, due to the planners' competence to freely set the corresponding parameters. The control is restricted not by the technology but by the fact that the interrelations of the parameters and the variables are not really clear to anybody and the system does not explicitly help to make it clearer. However, planners as well as schedulers say that the system provides them with a better information base than they had before. Anyway, the schedulers have pointed out two main problems regarding information presentation. First it is quite difficult to identify the real-world relations in the system's abstract representations. One must have a deep knowledge of the products in order to be able to interpret the abstract representations. This causes problems for the schedulers: "It is an unpleasant situation for me to release two identical jobs within short time, or to slow down a running job in order to let a similar job catch up. Sometimes I let the first job go through and delay the second one in order to prevent bad blood." The second deficiency refers to the clearness of the information presented. One often has to navigate through many windows in order to gather the information required to take a certain decision. Therefore the schedulers use a lot of additional instruments where they collect information in an arranged way (e.g. tables). They even still use the planning-board they planned with before they had the system.

The shop floor supervisors use the system for both, to feedback states of jobs (e.g. in progression, completed) and as a source of information. The type of information they look for in the system mainly concerns the availability of material as well as job states in preceding units. But this information is not really used to plan and schedule proactively, its rather used to check whether any disturbances in the schedulers' plans are to be expected. However, that kind of active information seeking has been observed rather seldom. Many of the shop floor supervisors when asked, did not really know how to find the relevant information in the system.

To summarise it can be stated that the ERP-system provides the planners and supervisors with information and does not restrict human control over planning and scheduling processes. However, information presentation needs improvement. On the shop floor level the system's information potential is not really used.

CONCLUSIONS

The findings of the case study clearly show a (conceptual) lack regarding autonomy-oriented design of temporal planning and scheduling systems understood as secondary work systems. The organisation of the planning and scheduling process, the assignment of planning tasks to individuals as well as the human-computer interaction (still) mainly follows the goal of putting single persons into control. Distribution of autonomy and control that is considered to increase an organisation's competence in coping with uncertainty is still not envisioned. What is required are concepts for organisation as well as for technical support of planning and scheduling processes that locally empower situated acting. The secondary work system for planning and scheduling should be considered as a network of brains that has to be promoted in its capability to flexibly adapt to dynamic situations. Organisation as well as technical support of information flow and decision-making should aim at interconnecting these brains' creativity. Instead, organisation on the one hand kills creativity by centralising autonomy and control, and technical support on the other hand by the means of clearly structured abstract databases does not really interconnect brains, but terminals. To elaborate technical as well as organisational design solutions that aim at realising autonomy-oriented planning and scheduling systems is a work that still must be tackled.

REFERENCES

Battmann, W. (1984). Regulation und Fehlregulation im Verhalten IX: Entlastung und Belastung durch Planung. Psychologische Beiträge, 26, 672-691.

Dahme, C. (1997). Systemanalyse menschlichen Handelns. Opladen: Westdeutscher Verlag.

Emery, F. (1959). Characteristics of socio-technical systems. London: Tavistock Document No. 527.

Grote, G. (1997). Autonomie und Kontrolle. Zur Gestaltung automatisierter und risikoreicher Systeme. Zürich: vdf Hochschulverlag.

Grote, G., Ryser, C., Wäfler, T., Windischer, A. & Weik, S. (2000). KOMPASS: a method for complementary function allocation in automated work systems. International Journal of Human-Computer Studies, 52, 267-287.

Hacker, W. (1998). Allgemeine Arbeitspsychologie. Bern: Huber.

Hayes-Roth, B. & Hayes-Roth, F. (1979). A cognitive model of planning. Cognitive Science 3 (4), 295-310.

Hoc, J.-M. (1988) Cognitive psychology of planning. London: Academic press.

Hollnagel, E. & Cacciabue, P.C. (1999). Cognition, technology and work: An introduction. Cognition Technology & Work 1(1), pp. 1-6.

McKay, K.N. (1987). Conceptual framework for job shop scheduling. Master Thesis. Waterloo Canada: University of Waterloo.

McKay, K.N. (1992). Production planning and scheduling: A model for manufacturing decisions requirement judgement. PhD Thesis. Waterloo Canada: University of Waterloo.

Resch, M. (1988). Die Handlungsregulation geistiger Arbeit. Bern: Hans Huber.

Rice, A.K. (1958). Productivity and social organization: The Ahmedabad experiment. London: Tavistock.

Sanderson, P.M. (1989). The human planning ans scheduling role in advanced manufacturing systems: An emerging human factors domain. Human Factors, 31(6), pp. 635-666.

Schüpbach, H. (1996). From central planning and control to self-regulation on the shop floor. In: Eric Scherer (Ed.). Shop floor Control – a systems perspective. Berlin: Springer.

Strohm, O. (1996). Produktionsplanung und-steuerung im Industrieunternehmen aus arbeitspsychologischer Sicht. Zürich: vdf Hochschulverlag.

Strohm, O. (1998). Integrated design according to the concept of people, technology and organization. In: P. Vink, E.A.P. Koningsveld & S. Dhondt (Eds.). Human factors in organizational design and management VI. Amsterdam: Elsevier.

Suchman, L.A. (1987). Plans and situated action. The problem of human-machine communication. Cambridge: Cambridge University Press.

Susman, G.I. (1976). Autonomy at work: A sociotechnical analysis of participative management. New York: Praeger.

Trist. E.L. (1981). The Evolution of Socio-Technical systems. In: A.H. Van de Ven & W.F. Joyce (Eds.). Perspectives on Organization Design and Behavior. New York: John Wiley.

Ulich, E. (1998). Arbeitspsychologie (4. Aufl.). Stuttgart: Poeschel.

Wäfler, T. (in press). A systems approach to human planning and scheduling. In: B.L. MacCarthy & J.R. Wilson (Eds.). Human performance in planning and scheduling: Fieldwork studies, methodologies, research issues. London: Taylor & Francis.

Wiers, V. (1997). Human computer interaction in production scheduling. Eindhoven: Technisch Universiteit Eindhoven.

Wiesner, B. (1995). Diagnostik individueller Planungsprozesse. Regensburg: Roderer.

Zölch M. (1997). Aktivitäten der Handlungsverschränkung. Dissertation: Universität Potsdam.

Supporting Temporality and Synchronization

Rachel Israel
EURISCO, European Institute of Cognitive Sciences and Engineering
4, avenue Edouard Belin, 31400 Toulouse
Rachel.Israel@onecert.fr
http://www-eurisco.onecert.fr
2UTC-COSTECH, Université de Technologie de Compiègne, 60200 Compiègne

ABSTRACT

Concurrent Engineering and the related use of Digital Mock-Ups (DMU) have strong effects on the temporal conditions for joint activities. Based on sociological observations of two international cooperative projects, we propose a qualitative typology of temporal constraints to be included in a project schedule in order to support co-design and synchronization, even under high temporal pressure.

Keywords

Cooperative work, intermediary objects, temporal conditions of joint activities, instrumentation of cooperation, time-pressure.

JOINT ACTIVITY SYNCHRONIZATION

The temporal constraints of collective action over-determine individual and collective activities adjustment. It is necessary to know what is at stake in the time line of a project, and how this time is defined. One of the dominant approaches of the temporality of co-operative action in ergonomics is that of cognitive synchronization: one seeks to elucidate the mechanisms allowing situated decisions and coordination scheduling within the groups at work, mainly by the construction of shared representations, mutual awareness or common spaces of information (Schmidt and Simone 2000). The situated cognition paradigm (see Suchman 1987, Heath and Luff 1991, Hutchins 1995) have pointed out the role of objects as symbols and support of fine tunings and synchronization for the construction of intersubjectivity. The sociology of networks, with Latour (1994), also highlighted that the meaning of an innovating object is constructed starting from alliances which are carried out around and thanks to the object. When all the entities of a network can produce an agreement on the nature and qualities of the object, this object then exists as "a black box" which federates the network at the same time as the network constitutes the object: it is a circular causality. The stability of both the objects and the networks ensues from this synchronization loop as a co-production.

From the same point of view, Jeantet and Boujut have studied the role of graffiti, schemas, planning etc., in the design process. As intermediate objects, they condition the emergence of the end-product, making it possible for the group to gradually elaborate an agreement on the problem and its design solution. The intermediate objects are places of coordination, with functions of translation, mediation and representation. They are also milestones of the design process. In a more general way, the measure of time comes from technical objects, which affect our perception and pre-structure our actions Rabardel (1995) as a tertiary, external memory (Husserl 1999). As Stiegler (1994) puts it, "technique is time". We affect these technical objects in return by our uses and our anticipations.

The question is thus that of the shift between social and technical evolution rhythms (Stiegler 94). Moreover such a shift is central for authors like Schmidt and Simone (1996) or Jeantet and Boujut (2000). They point out the need for questioning the initial conditions of emergent synchronization, so specific to cooperative activities. These conditions bring into play the social construction of a project temporality which exceed the decision and action perimeter of the actors of the co-operative project. Thus, in their paper on the instrumentation of co-operation, Boujut and Jeantet (2000) pointed out the importance of reflexive agreements concerning the functioning of a co-operative group. This approach is shared by Terssac (1996), or Zarifian (1996) for instance.

Having defined co-operation as an activity whose operating mode is closely related to the contents of the task, the quoted paper continues to sketch such initial conditions: a specific operating mode, founded on organizational meta-rules, empty frameworks which will be cooperatively made meaningful by the actors as their activities are evolving. The temporal framework of co-operation is part of the problems raised, because "the usual tools for project management badly handle the unstructured periods of negotiation of collective action" Boujut and Jeantet, 2000, p102) A sociological point of view on collective action addresses as well the essential concept of synchronization. To cognitive synchronization, we will add synchronization of affects, i.e. motivation, disengagement and other emotional schemas (Salvador 1995), and sub-symbolic synchronization, i.e. the construction of sensori-motor schemas of action (Piaget 1974, Rabardel 1995), shared motor imagery and practices, which are also significant although they are not based on "mental representations" (Varela 1993, Printz 1997, Salvador 1995). We will add that any synchronization has a mimetic base, since it is a question of acting, of thinking and of feeling like others (Tarde 1903, Salvador 1995). The concept of temporality points to each skill "social time", i.e. structured according to their specific constraints, process, tools and culture. It should evolve to meet the cooperative activity needs.

The temporal framework of the multi-skills project results too from a complex agreement, but takes place outside the project border. It is then imposed as a top-down constraint to the project actors. It then comes in the form of a counterproductive tension with a local, chaotic temporality, resulting from situated and underground arrangements, more or less translated in the planning. As a result, organizational tools, planning and DMU, supposed to ensure the synchronization of the project, fail to become intermediate objects of the project itself. And when by need (breakdowns, blocking etc) the actors of the project assert themselves in the loop, it is by the back door, at an exaggerated emotional and energy cost, and following specific and precarious processes.

We will approach the temporal framework of the project by two questions:

- How these temporalities affect the construction of the DMU as an intermediate object of the process design for the members of a cooperative project;
- how the dynamics of DMU affects existing temporalities.

We will illustrate this loop using several situated examples, to highlight the temporal resources implemented to facilitate synchronization. We will propose a qualitative typology of temporal constraints which must be institutionalized as initial conditions of a project to support synchronization. If the principle of adaptable meta-rules tuned by the co-operative group is recognized, this paper attempts to propose such types of meta-rules for the temporal framework of cooperative action.

STUDY CONTEXT AND METHODOLOGY

These reflections are grounded within cooperative activities in aeronautics design process. We spent two years performing successively ethno-methodological participative observations within two international cooperative teams (60 people, German and French). The projects were innovative in two ways. They used for the first time, tools specified for long aircraft design projects:

1. Organization: creation of an Integrated Team for aircraft re-design, i.e. short-term (one year) and high time-pressure projects;
2. technology: use of a digital mock-up (DMU), and of Concurrent Engineering design methods. The French Design teams, representing 5 different skills (Mock-up Integrators, Structure and Systems design, Stress, Manufacturing) were more or less co-located on the same "platform". The production teams, working within 3 different specialized manufacturing units, sent representatives to the platform two days a week; the German partners had sent a co-located design team (4 to 10 people from different skills) for project A, and stayed in Germany for project B. All the members were supposed to use the DMU. In a general way, the different skills and partners are inter-dependant and need to exchange information about their part of the worksharing according to joint activities dynamics.

These projects' characteristics may seem to be too specific, but they are aimed to develop for they respond to industrial needs for reactivity.

Data, information and knowledge about the use and the management of time have been built out of the transcriptions of direct observations of articulation work (Strauss 1992), of verbal and non-verbal communication and synchronization (Heath and Luff 1991) at work during breaks, of interviews and of exhaustive note-taking during project coordination meetings. These elements have been classified into physico-behavioral units such as "coffee pause", "design review" or "drawing set control". Thus, the same categorization can be used by both the actors and the observer to perform analysis. After a content analysis, a qualitative grid analysis was used to compare the two cooperative projects, according to a threefold approach:

1. The management of norms/practices gap;

2. the technical and organizational changes related to cooperation tools, especially concurrent engineering and digital mock-ups;

3. the construction of project shared values.

SOCIO-TECHNICAL EVOLUTION OF THE WORKPLACE

We will focus on DMU because they are initially designed to shorten industrial cycles according to "concurrent engineering", i.e. cognitive synchronization and activity parallelisation, Darses, (1997). Thus, they strongly affect temporalities and synchronization conditions.

Working with a physical mock-up

After the design definition by the Design Office, the production department builds physical mock-ups at various scales to test technical solutions. There is a significant period of time before the mock-up is tested, possibly requiring significant modifications. But the longer the delay, the more the definition still in progress becomes a stabilized state of fact which sometimes puts a limit to required modifications.

Working with a digital mock-up (DMU)

The baseline drawings corresponding to agreed technical solutions are carried out in 3D on CAD workstations by Structure and System Designers. They are continuously integrated by Mock-up Integrators (MI) into 3D virtual mock-ups which represent the project (all the impacted plane parts), i.e. the project progress in quasi real-time. The various skills can refer to it in parallel. Certain functions allow the MI to carry out consistency checks, interference checks, or automatic calculations. With the DMU, a stabilized technical choice can be seen by all the different points of view. The drawing sets zoning of the aircraft reflects the production process organization. Design and tests are iterative, and modifications performed within a few hours/days. People now cooperate either in a distributed, parallel way, or in a co-design way; they are no longer only confronted to a "problem-solving" activity as the problem has to be set too (Darses 1997).

Mock-ups and temporality

Using digital mock-ups implies significant changes related to time segmentation, both on individual and collective levels. This is mainly due to the fact that a mock-up, as a technical object, is a time generator, a pace setter. Thus, the severance of time embedded in a tool dynamic can affect the social relation to time (Stiegler 1994). We will now account for examples showing the way the DMU in use differentially affects the temporal frames of design and production, and of French and German partners.

Situated approach of the DMU

Before illustrating the changes induced by the DMU on the inter-skills synchronization by some "breakdowns", we need to provide some social context. The DMU, in support of the Concurrent Engineering process, finalizes a long-term company strategy: it represents the will to organize all the industrial activities according to the production organization. In the 1950's, the logic was the opposite: activities were structured according to the design needs. But the values which were attached to the primacy of design did not disappear, and the new values were superimposed on it (Terssac 2000).

DMU: towards iterative temporalities

On project A, it was discovered at a quasi-finalized level that pre-drillings for the electric systems supports of the cross-beams were 2 mm too short of the existing holes, which is an unacceptable defect.

People only exceptionally re-design a grounded-aircraft, i.e. an assembled and fitted out aircraft. Contrary to an aircraft on the assembly line, it already has holes which means customizing the design standard. That was not done. However, all the project members knew that: how could that have been forgotten? The result of a crossed investigation shows it was not a question of forgetting, but on the contrary, respect for a guiding principle of design: the system adapts. Explanation: with a "physical mock-up" structure, design validation was a long process. So the more flexible systems were the ones modified. But with the DMU, iteration becomes the rule since the structure is virtual. The "misunderstanding" came simply from the application of a gold rule made obsolete within a few weeks. This rule controlled not only the relations of synchronization between two types of technical interventions, but also the relation of force between Structure and System designers: the energetics of the social metaphor "the System comes after" lasts well beyond the technological changes. The point is that in spite of the DMU and the cognitive synchronization grounded into the new process (its comprehension was perfectly shared by the designers), the leading image of the social Structure primacy remains stronger and guides anticipations. Cognitive synchronization are necessary here, but are not sufficient for an effective tuning.

This superposition of temporal patterns is sometimes also found "within" the artifacts. In spite of the informational blur of co-operative projects and the conjectures which it implies, a CAD module for electric systems design requires information at a level of granularity incompatible with this situation. A "system" team had to do their first months' work again: in order to use the module, they had to anticipate non-available information, but corresponding well to the principle already quoted: the systems adapt, i.e. they start to work after the structure, and therefore with fine level information. As a result, the level of action design requested by the CAD artifact from the electricians does not correspond to their perception on the project, leading them to chimerical anticipations that can not be shared. There is a break in their anticipation/perception/action loops (A/P/A loops) resulting in a loss of control. The technical system CAD-DMU, which implies a refunding of the skills temporalities, has reproduced ghost-temporalities, which haunt the iterative present of cooperative work.

A new structure-system articulation will occur only by the negotiation of temporal frameworks and consistent technical mediations, project by project, and phase by phase. Then intersubjective synchronization will be easier. But in the case of time-pressure constraints often associated to cooperative projects, the question remains open.

Views on an intertwined process

Designers working with the DMU follow a specific process:

- They agree on study plans, which thus become intermediate design objects allowing and symbolizing the group tuning. (Boujut Jeantet 2000) ;
- "design solutions" are validated in a design review; the MI (Mock-up Integrator) integrates the corresponding drawings and checks their consistency on its workstation, where he shows the possible problems to the designers. The drawings'status is modified ;
- Highly codified information, corresponding to manufacturing parameters are added. The drawings become production drawings; their status in the DMU is modified ;
- the drawing sets are checked by he design teams; then checked by the MI. Their status in the DMU is modified. The MI sends the production drawings to the Design Check Office for a formal control. Their status in the DMU is modified ;
- Drawings evolve according to check results:
- If OK, the status is declared "official", and drawings can be sent to manufacturing ;
- If not, the drawing is modified and goes through the control loop again ;
- After an "official" production drawings modification, its issue number is modified, but the coding alpha-numerical structure does not say if the change was design or formalism oriented.

What are the results from this first experiment with the DMU?

For the designers, whose flow of experience is in direct contact with the flow of the drawings, who work close to the MI and benefit from a situated support, it takes a few months to master the process logic. They do not know when information on standby will be available, but they manage to evaluate its impact and to juggle with the planning. But the process is less clear for the production members: they are far away, their organization does not correspond to aircraft zoning, but to manufacturing specialization, they know the DMU only from a few days of off-site training and they do not have a situated support. Moreover, the relationship between Manufacturing and the Design Office remain marked by the dual design/execution tradition. It is thus not easy for the production teams to say: it is not clear to us. When they do, the answer is: you were trained with the DMU. The time for tool-training is confused with the time for appropriation of the new process; the significant changes which affect synchronization of cycles and mutual anticipation are not taken into account in the preparation of the project. The production members encounter/experiment serious breaks according to their A/P/A loops.

On top of the above mentioned issues is the temporal stress of the project: the global planning is too short to respect this standard process. This is understandable: the manufacturing teams can go on with their preparatory work using a production drawing which is technically OK, but will require 5 days of formal correction process. The definitive drawing will arrive later. Then come the negotiations on what a "usable" drawing is for the production teams. However, on the design side, the technical information is not stable, and its fluctuation is not very foreseeable, mainly for two reasons:

1. Differences in cycle between design and stress specialists, whose tools, methods and organization are time-consuming, and poorly adapted to very iterative synchronization;

2. differences in temporal management of the project by the German partner. In the German way, a project starts slowly because of the initial negotiation of its bases; the German organization is more partitioned and the transverse mechanisms of coordination Schmidt and Simone 1996) are limited. For German, the French quick and strongly implicit way to negotiate and propagate information is puzzling. Due to the lack of a common intercultural training, mutual anticipations are poor, and information flows are poorly synchronized.

Drawing set modifications can thus be requested very short notice. The factories can be working on a production drawing in technical conformity, remaining on standby for formal correction, but which has since, became technically obsolete. When browsed the DMU has no trace of the modification as it is still in progress. In spite of telephone calls, e-mails and everyone's goodwill, we can imagine the difficulty the production teams encounter in making sense of these flows and counter-flows of which they have no situated practice. But when the manufacturing teams announce that from now on, they will accept only "official" production drawings, sticking to the standards, they blame the shared practices which grounded the community (Lave and Wenger 1991).

Thus, the resulting emotional load becomes significant; the postulate of sincerity on which Goffman (1970) grounds any successful interaction breaks down, and the cognitive conditions of synchronization deteriorates. Despite everybody's goodwill to make a success of the challenge and countless energy, the points of view cannot converge. The level of stress increases.

The collective management of stress

Hence, the collective management of stress is a significant factor of regulation within the integrated team, which results in the exchange of reflexive narratives. Narratives reorganize common interpretation schemas, and thus, the situation itself. They worsen it in a kind of emotional catharsis grounded in the imitation of affects, then reduce it with humor, relativize it, helping those who are made late by the situation to feel guiltless. This time for construction of collective reflexive narratives is invaluable for it allows to step back and provides a new framing for both the problems, the roles and the affects. At the same time, it allows subtle systemic analysis of the problems: why the partners, the production teams, in the end, do not really use the DMU; why the latter role in the design process design is not well-defined, etc. Surprisingly, these positions, very critical, do not lead to demotivation. My hypothesis is that it is precisely the collective and reflexive descriptions of the intertwining elements of the situation which makes it possible to relativize: the construction of interpretative fictions in term of affects (we'll show them all, we're above that); the reactivation of the myth of the pioneer which marked the history of aeronautics (it is hard, but we will do); and the a-tuned sociological and organizational analysis. Moreover, this situation of cooperative reflexivity directly feed a collective experience feedback which can be appropriated by everyone; it is occurring in the activity, and is thus more efficient (Moisdon 1997). However, it is the non-standard situation of the integrated team that allows this intertwining to be collective and distributed. Under normative conditions, exchanging jokes, laughter and criticism so loudly and openly would pose problems, and would exceed spatially and temporally marked statutory limits. But these essential moments are not institutionalized. They are integrated in no planning, and they are not compatible with usual statutory limitation.

This cooperative management of stress reinforces the feeling of solidarity of the co-located teams, and facilitates sub-symbolic, emotional and cognitive synchronization. A team leader talking of the efficient distribution of work in his team will say: "you know... it was just done like that. When one of the guys was rushed, another went to help him; work was distributed according to feeling, we never really talked about it". When such a level of sub-symbolic and emotional synchronization has been reached, they are few problems on the level of cognitive synchronization: success is always anticipated. The time of commitment in common actions, reshaping the flow of events in a shared and foreseeing memory plays a crucial role in these constructions. However, within the group, the parallelization of the design activities, as much as the strong individualization of the tasks carried out by the CAD (Béguin 1994), is obviously worsened by the temporal, spatial and organizational structures whose genesis is grounded in parceled structures of work and experience.

TEMPORAL FRAME: A TYPOLOGY

We can now propose a qualitative, non-exhaustive typology of temporal constraints needed to improve the temporal conditions for individual/collective synchronization and reflexivity.

- Time is needed to make a quick situated inventory of practices and uses within the project, in order to ground emergency in habits, or to support the development of new habits in activity. This avoids time-consuming tools/methods misuse or non-use;
- time for the collective appropriation of a tool/user expertise corresponds to the time it takes for each skill to reestablish the A/P/A loops that were disrupted by the introduction of a new tool, and to develop shared schemas and habits;
- time for the appropriation of the work process is the time needed to build an image of one's position and of his/her colleagues positions within the new process. It has nothing to do with tool training, and aims to synchronize mutual consistent A/P/A loops whereas certain tools tend to split or over-develop cross-perception and anticipation;
- time for quality, progress status and technical consistency checks is the time it takes both to design an effective control process, and to negotiate the conditions of a global cross-control. This means a negotiation of the main evaluation criteria and processes, which traditionally belongs to a panoptical and top-down hierarchical tradition;
- time for acquiring a social expertise in coordination and organization is the time needed by the various skills to achieve a realistic and situated understanding of their own and others' behavior and social practices. It means building shared affects as interpretative schemas. It aims to build situated rules and anticipate others' needs and constraints. This implies sequences of reflexive evaluation and a specific categorization of the social context evolution into the organizational memory;
- time to "wait for" information is the time for storytelling as ego- and ethno-genesis (Rastier 1999), and for constructing fictions. It corresponds to the continuous time during which the various skills attempt to fill in missing information and anticipate on the future of the project. This allows the collective management of the stress related to deadlines, to decision struggles and to role/ leadership redefinition within the project.

A cooperative temporal frame taking such a typology into account is missing in two ways:

1. It is not planned;

2. it creates a gap. On the one hand, those who can take this time for they are in constant interaction. The articulation work (Strauss) allows them to reshape the project time, through narratives, interpretative fictions. So they can manage the informational uncertainty and

negotiate a locally adapted but too precarious management of resources. For others, the missing time remains a burden all along the project.

Our issue was to stress the role of time constraints on the construction of intersubjectivity. We found problems when the time needed to allow deep cognition/habits/motivation co-construction and tool/organization temporal consistency is not included in project management. This balance is part of the ecology of social activities. It must then be partly planned, partly negotiated with the members of the project, bridging the gap between improvisation based on mutual awareness versus coordination in terms of predefined work schedule (Schmidt and Simone 2000). Only then can the consensus within the community of users emerge thereby leading to synchronized practices that must also be brought to a standardization phase.

CONCLUSION

Using the DMU brings an increasing need for cooperation and a constant difficulty to create conditions for cooperation to take place within traditional structures. This tends to thwart the project teams' expected reactivity. This appropriation leads to the emergence of a co-operative organization, which questions the concept of a hierarchy of position, to the benefit of a local, competence-based hierarchy. The genesis and the role of the latter hierarchy is evolving according to situated needs, and redefined in relation to a dynamic and collective building of the project. But this local hierarchy has no power related to the initial temporal frame of the project: it is not constituted then, and no memory is kept of "social" management of the temporal troubles. This social construction of emergency is reproduced, and thus considered as "normal".

The DMU provides an identical vision to all, but this vision is not common. This is why the DMU as such will become a project's intermediary object when temporal conditions for synchronization are scheduled, even within high time-pressure projects. The DMU is much more than a place for storage and improvement checks, or a provider of up-to-date representation of the design process in progress. But this means organizing conditions for a inter-skills process of fine tuning. In order to share a common perspective, DMU users must fight to become cooperation actors, and to redefine cooperation rules by entering the organizational and technical design loop (Israel 1999). This is what they do, despite standards and work schedule. The synchronization of practices and affects plays an essential role there: the same information, the same facts will become a very different event according to the interpretative schema. Taken in a positive loop, obstacles will become momentary obstacle and reason for the group to surpass itself; taken in a negative loop, the same obstacle will be seen only as an illustration of the organizational breakdowns and a reason for wait-and-see policy. This dominating schema, collectively built, plays a crucial role on individual behaviors and on the project progress: an affect is thus a project intermediary object.

This is why taking into account the initial temporal conditions of a project is essential. The projects we observed revealed the pleasure of building a co-operative opera which allowed people to take initiatives and to exercise an efficient collective management of problems and stress. The DMU is already a tool in the process of appropriation; organization patterns more consistent with its technical potentialities are beginning to emerge. Due to the tight overlapping of synchronization modes and to people implication, it does work. So why change anything about the temporal conditions for cooperative activities? Because shortening the time it takes for human synchronization has its limits, as shown by Vaughan's work on Challenger (Vaughan 1996).

REFERENCES

Béguin P.(1994) Travail avec la CAO en engénierie industrielle. De l'individuel au collectif dans les activités avec instruments. Thèse d'ergonomie, Paris, CNAM.

Darses, F.(1997) L'ingénierie concourrante, un modèle en meilleure adéquation avec avec les processus cognitifs de conception, Ingénierie concourrante : de la technique au social. Paris : Economica.

Goffman E.(1970) Strategic Interaction. Oxford. Basil Blackwell.

Heath, C., Luff P. (1991): Disembodied conduct: Communication through video in a multi-media office environment,' in CHI '91.ACM SIGCHI Conference on Human Factors in Computing Systems, New Orleans, 28 April-2May.

Husserl E.(1999) L'origine de la géométrie, Paris, PUF.

Hutchins E.(1995) Cognition in the Wild. Cambridge. MIT Press.

Israel R.(1999) Une mémoire de l'émergence : vers un outillage socio-technique de la coopération, Solaris n°5, http://www.info.unicaen.fr/bnum/jelec/Solaris/d0/index.html.

Jeantet A et Boujut J.F., L'instrumentation de la coopération . Toulouse Journées Prosper/CNRS June 7-8.

Latour B. (1994). La science en action. Paris. La Découverte.

Lave, J., Wenger E.(1991) Situated Learning. Legitimate Peripherical Participation, Cambridge, Cambridge University Press.

Moisdon JC ed (1997). Du mode d'existence des objets techniques. Arslan.

Piaget, J.(1974) Adaptation vitale et psychologie de l'intelligence. Hermann, Paris.

Prinz, W. (1997). Perception and Action Planning. European Journal of Cognitive Psychology, 5.

Rabardel, P. (1995) Les hommes et les technologies. Approche cognitive des instruments contemporains. Paris. A.Colin.

Rastier F. (1999) Action et Récit, in La logique des situations, sous la direction de M. de Fornel et L. Quéré. Paris. Editions EHESS.

Salvador, L.-L. (1995). Swarm intelligence, collective action and imitation. Proceedings of the First European Conference on Cognitive Science, 267-275. Paris : ARC & INRIA.

Schmidt, K. Simone C.(1996): 'Coordination mechanisms: Towards a conceptual foundation of CSCW systems design,' Computer Supported Cooperative Work. The Journal of Collaborative Computing, vol. 5, no. 2-3, pp. 155-200.

Schmitd K and Simone C. (2000) Mind the gap! in Coop 2000, Sophia Antipolis, France 23-26 may.

Stiegler B.(1994) La faute d'Epiméthée, La technique et le temps Galilée, Paris.

Strauss A. (1992) La trame de la négociation, Paris L'Harmattan.

Suchman L (1987) Plans and Situated Actions. Cambridge. Cambridge University Press.

Tarde G(1903) The Laws of Imitation. New York Holt.

Terssac G de. (1996) Coopération et Conception. Toulouse. Octarès.

Varela F (1993) L'inscription corporelle de l'esprit, Paris, Seuil.

Vaughan, D. (1996) The Challenger Launch Decision. Chicago IL: University of Chicago Press

Zarifian P. (1996) Travail et Communication. Essai sociologique sur le travail dans la grande entreprise industrielle Paris PUF.

Plans versus Outcomes: Establishing the cost of planning

Peter Timmer
Ergonomics & HCI Unit
University College London
26 Bedford Way
London, WC1H 0AP
p.timmer@ucl.ac.uk
http://www.ergohci.ucl.ac.uk/people/peter.html

John Long
Ergonomics & HCI Unit
University College London
26 Bedford Way
London, WC1H 0AP
j.long@ucl.ac.uk
http://www.ergohci.ucl.ac.uk/people/john.html

ABSTRACT

This paper proposes a taxonomy of planning outcomes, and then applies it to data from a simulation of Air Traffic Management. The paper distinguishes interventions that are specified and immediately executed (instant execution) from interventions specified, and then executed after some temporal delay (planned interventions). Plans that are formed, but not executed, are considered a cost to the controller. The taxonomy contains three types of plan outcome, two of which constitute different types of cost: plans that are discarded as inappropriate (due to anticipation failures of controllers); and plans that 'decay' before execution.

Keywords

Planning, instant execution, plan-and-execute, plan-and-discard, plan-and-decay.

INTRODUCTION

When assessing the performance of an operator-technology worksystem, a number of high-level criteria are frequently employed (usability, quality, costs), criteria that can be decomposed into measurable metrics. For example, the Software Usability Measurement Inventory (SUMI) (Porteous, Kirakowski & Corbett, 1993), decomposes usability into: efficiency; affect; helpfulness; control; and learnability, and provides fifty questions and an agreement scale as a mechanism for measurement using such metrics, and thereby establishing overall criteria compliance.

In this paper, the measurement of operator costs (incurred during interaction) is considered, rather than usability. Historically, such costs have been measured with respect to three distinct metrics: time to complete a task (Gray, John, Stuart, Lawrence & Atwood, 1990); mental workload (Tattersall & Foord, 1996); and errors (Wioland & Amalberti, 1998). Here, operator costs are treated in a slightly different manner, and a metric is proposed that focuses upon the relationship between a) the mental planning behaviour, invested by an operator in trying to manage a process, and its relationship with b) observed management interventions with that process. Costs are considered to have been incurred when planning behaviour is directed at specifying process interventions, but observation shows that these planned interventions are not executed. The assumption that lies behind such a conception of costs is one of waste, and its minimisation – planning interventions that are not executed is a wasteful cost to be minimised, whereas planned interventions that are executed is a benefit to be maximised. While it is clear that costs may be associated with any mental and physical behaviour undertaken by an operator, wasted or unwasted, in this paper the focus is only upon costs, associated with planning behaviours that fail to bring about interventions as specified in a plan.

After introducing the context within which this concept of costs arises, the paper presents a taxonomy of planning outcomes (different types of cost), and illustrates that taxonomy with the measurement of costs for different operators, when carrying out an Air Traffic Management-like (ATM) task, in an ATM microworld. Of particular concern to this research is (1) planning behaviour, (2) subsequent physical action (intervention), and (3) an analysis of the relationship between planned actions and actual interventions. Finally, it is proposed that by measuring costs in this way, support is offered in characterising the severity of planning problems with an existing worksystem. The paper's purpose is therefore to illustrate cost measurement, rather than to say anything substantive about ATM per se.

NEED FOR PLANNING SUPPORT

Planning is a ubiquitous mental behaviour, observed in the management of most, if not all, dynamic processes and many static processes. In consequence, the design of effective planning support tools has a long history in the Cognitive Ergonomics literature (Hoc, 1988). This research is motivated by a comparatively recent development in transport management domains, especially Railway Signal Management (RSM) and Air Traffic Management (ATM), that has placed a focus upon developing support for more effective planning, further ahead in time (David, 1997; Miailler, 1998). As demand within the UK, for rail and aircraft movements, outstrips current abilities to supply such movements, this problem may be broadly couched as 'how can more traffic (aircraft or trains) be managed over a limited infrastructure (sectors/flight levels or railway lines/signals)'. In ATM, the concept of 'gate-to-gate' management has been proposed as a solution, whereby future ATM worksystems will be expected to plan as many interventions as possible (with aircraft), from departure gate to arrival gate. The goal of this operational concept is to optimise the capacity for traffic over limited infrastructure. Likewise, for rail transport, if the paths of classes of trains could be better planned, for example, to avoid slow freight trains delaying high speed commuter trains, more rail traffic could be accommodated by extant infrastructure. From the gate-to-gate operational concept, it is clear that there is a need for planning technologies that will enable management worksystems to plan interventions at an early stage, far into the future, and that those plans be a) of sufficient quality to support effective management, and b) retrieved when appropriate.

Gate-to-gate management is a solution to the growing problem of managing greater numbers of aircraft over a fixed infrastructure, and it is taken as a context for developing planning systems. However, the present work attempts to support the development of planning systems somewhat earlier in the development process, at the problem formulation stage (Rasmussen, 1992). While gate-to-gate management places the requirement for effective long-range planning at the heart of system development, it prompts the question of how existing levels of planning effectiveness may be characterised. In this paper, particular emphasis will be placed on a method that enables such a characterisation to take place, especially with respect to expressing the extent to which an existing worksystem's planning behaviour is effective or wasteful – how much of what is planned is executed? If plans are not executed, what happened to them? It is from this context that the concept of costs as wasteful planning behaviour arose. While time, workload and error are all metrics that will support the development of planning systems, such metrics do not necessarily capture planning outcomes. It is proposed that this problem needs to be characterised, by some method, to enable reasoning in design about (gate-to-gate type) solutions.

PLANS AND THEIR OUTCOMES

The planning literature distinguishes 'planning time', or the moment of plan formation, from 'execution time', or when a planned intervention is to be executed. Frequently, (as in ATM at present), the time between plan formation and execution is very short. Gate-to-gate management concepts would place a greater temporal distance between planning and execution. However, the design problem facing the development of planning systems for ATM and RSM is not merely to support more detailed planning earlier. The design challenge is to develop plans, at an appropriate level of detail, such that they can be executed as specified, and that those plans support work (train or aircraft management) of a desirable quality (safe and expeditious). In contrast to planned interventions, many observed interventions are specified immediately prior to execution, i.e. without a time delay between specification and action. In this paper, such interventions are termed 'instant execution', and are contrasted with planned interventions. Amalberti & Deblon (1992) make a similar distinction. It is important to state that the scope of this paper's concern is for 'action planning', i.e. the planning of operator actions to intervene with the managed traffic. Plans for traffic management may occur at several levels of abstraction, before the detail of a particular altitude, heading or speed are specified, for example, plans for handling pairs or groups of aircraft, or aircraft arriving or exiting a sector at particular beacons. Such plans are beyond the scope of this paper.

The simple dichotomy presented above, between interventions that are planned, and those that are not, may be further refined, by examining more closely actual plan outcomes. When contrasting the plans that an operator forms, with observed outcomes (interventions), it is clear that a simple relationship between plans and interventions does not exist. Rather, plans may have a variety of outcomes. First, plans may decay (i.e. be forgotten), later to be reformed, either with the same details for action as earlier specified or different details for action. Second, plans may be discarded. The future state of the managed domain, for which the plan was formed, may have been inaccurately anticipated. So, at execution time, the plan may not ensure that management goals of safety and expedition are attained.

In this paper, a simple taxonomy will be used for categorising interventions, and relationships between interventions and any preceding planning activities. First, interventions may not be planned at all (instant execution). Data revealing a high proportion of such management would be symptomatic of an existing worksystem operating in a manner far removed from a gate-to-gate operational concept – largely reactive. Second, interventions may be well related to earlier planning, and executed as specified. Such an outcome shall be termed plan-and-execute. Third, interventions may have been planned, but are not executed, due to problems of memory retrieval (the simulation is similar to operational ATM in that no support is offered for documenting plans once specified), i.e. the plan was forgotten. When plans decay, replanning takes place (possibly in an identical manner to the previous planning behaviour), rendering the earlier planning wasteful. Such an outcome shall be termed plan-and-decay. Finally, interventions may have been planned, but at some point between planning time and execution time they are rendered inappropriate, possibly due to a failure to completely and accurately anticipate the future state of the domain. Such planning is also judged wasteful, and a cost to the worksystem, as it does not lead to the intended intervention. Such an outcome shall be termed plan-and-discard.

In the following sections, the simulated task will be outlined, and a model of one operator's plans and outcomes presented. Summary data for five trained controllers are presented, using the taxonomy, and emergent patterns of plan outcomes for these controllers discussed.

SIMULATION TASK DEMANDS

The data that follow were obtained by trained novice operators, interacting with a re-construction of the ATM domain and worksystem, as observed at the Ringway Control Centre in Manchester. The task involved an operator interacting with a radar screen and Flight Progress Strips (FPSs), issuing interventions to aircraft by a point-and-click method through the radar display. Scenarios involved the management of between 8 and 10 aircraft across a sector, which comprises two intersecting airways marked by five beacons. Aircraft entered the sector with an entry altitude and speed, route, and exit altitude (as specified on the FPSs). The operator's task involved managing aircraft across the sector, while maintaining aircraft safety (safe separation) and optimising expeditiousness (fuel consumption, progress to plan, number of manœuvres). These goals were attained by interventions to aircraft altitude and speed (for more details see Dowell (1998)). While the managed process was dynamic, and imposed a significant planning burden upon the operator, this simulation is of a relatively low fidelity with respect to the operational environment. Nevertheless, the simulation is at the very least an adequate cover-story for the exploration of cost measurement, as outlined above.

For each management scenario, operator behaviour was videoed. As interventions were issued by direct manipulation, the operator's voice channel was free, and concurrent verbal protocols were used to capture planning behaviour and plan specifications. At the end of each scenario, the computer-based model of the domain generated a record of all interventions with all aircraft, and an analysis of all aircraft states as a consequence of the intervention – safety and expeditiousness values (Dowell, Salter & Zekrullahi, 1994). From video data of each interactive scenario, operator-technology behavioural models were constructed (Timmer & Long, 1997). From such models, planning behaviour could be selectively extracted for more specific analysis, in terms of outcomes (instantiating the taxonomy) and thereby assessing costs.

MODEL OF PLAN OUTCOMES

A model of an operator's plans and actual interventions, for a single aircraft TAW, is detailed in Table 1.

Table 1: Comparison of interventions planned against interventions executed

	Intervention Specification	*Intervention*
1	Give TAW altitude 125	TAW to 125
2	Give TAW speed 720 when legally separated with SAM	
3	Give TAW altitude 135 when legally separated with SAM	
4	Plan (2) decays	
5	Leave TAW at 125 for present	
6	Give TAW speed 720 when legally separated from SAM	
7		TAW to 720
8	Give TAW Altitude 135	TAW to 135
9	Give TAW altitude 130 when near exit beacon	
10	Plan (6) decays	
11	Give TAW altitude 130 later	
12	Plan (8) decays	
13	Give TAW altitude 130 later	
14		TAW to 130
15	Leave TAW	

Where specification occurs immediately prior to intervention (instant execution), data are shown in bold italics in the 'Intervention' column. The model shows that four interventions were made with TAW, two instances of instant execution and two planned interventions. The table also shows that seven plans were formed (at Lines 2, 3, 5, 6, 9, 11 & 13), only two of which were executed as specified (at Lines 7 & 14). At Line 15, an explicit plan to leave the aircraft in its goal state may also be considered to have been executed.

From detailed models of worksystem behaviours (Timmer & Long, 1997), the outcomes for unexecuted plans can be categorised within the taxonomy. At Line 3, a future change of TAW's altitude is specified, for when TAW is safe with respect to another aircraft SAM. This plan is discarded (plan-and-discard) at Line 5 as inappropriate (TAW is to be left at its current altitude). Referring to such a discard plan (and the mental effort expended in its specification) as wasteful is reinforced by the fact that this intervention is made later by instant execution (Line 8). At Lines 9, 11 and 13 the same intervention is specified, yet is forgotten twice (plan-and-decay), before finally being executed at Line 14. From such historical data, for a single operator managing a single aircraft (for a scenario between eight and ten historical models can be constructed, one per aircraft) the taxonomy may be used to characterise plans, outcomes and where costs were incurred, as follows. In summary, this operator's management of TAW shows that 50% of interventions were planned, and therefore 50% were implemented by instant execution. When looking at all the plans for interventions, 50% were of the plan-and- execute variety (including plans at Lines 5 and 15), 38% were plan-and-decay and 12% are judged inappropriate between planning time and execution time (plan-and-discard). The taxonomy seems to show a considerable amount of unexecuted planning (costs), much failure to implement being due to plan retrieval (decay).

SUMMARY DATA

Five management scenarios were observed in total, each scenario consisting of between 8 and 10 aircraft (shown in brackets next to the scenario number). Table 2 shows summary data for plan outcomes for all these observed scenarios.

Table 2: Summary data for interventions planned and instant execution

Scenario (a/c no.)	*Planned Interventions*	*Instant Execution*
1 (10)	11	27
	29%	71%
2 (10)	15	11
	58%	42%
3 (8)	12	19
	39%	61%
4 (8)	6	19
	24%	76%
5 (8)	7	21
	25%	75%
Total	51	97
	34%	66%

If we imagine what data for an implemented worksystem managing aircraft gate-to-gate would look like, one would expect to find a very high percentage of interventions planned (near 100%), and very few interventions of the instant execution variety. We would also expect to find those plans being formed at an early stage during a scenario (no data are presented concerning when the plans in the summary table were constructed). Considering characterising the planning activities with the existing system (across operators and scenarios), it is clear that the worksystem is largely managing the traffic by instant execution. On average, two thirds of interventions are unplanned, compared to one third planned. This pattern is fairly stable, with the exception of the operator in Scenario 2, who planned nearly 60% of interventions. The range of planned interventions extends from 24% to 58%, with more operators skewed towards the lower range. Having established this overall picture of the relationship between planning and intervention, it is possible to instantiate the taxonomy and analyse planning outcomes. An examination of plan outcomes makes it possible to establish whether so few planned and executed interventions (approximately one third) were due to little apparent planning, or whether plans were not being executed, due to other reasons. Table 3 shows summary data for plan outcomes using the taxonomy.

Table 3: Summary of plan outcomes for all scenarios

Scenario	Plan-execute	Plan - decay	Plan-discard
1 (10)	11	6	2
	57%	32%	11%
2 (10)	15	4	3
	68%	18%	14%
3 (8)	12	8	5
	48%	32%	20%
4 (8)	6	7	6
	31%	37%	32%
5 (8)	7	5	6
	39%	28%	33%
Total	51	30	22
	50%	29%	21%

Table 3 shows that on average, half the plans formed by operators were executed as specified (plan-and-execute). This figure does, however, vary, from 68% for Scenario 2, to 31% for Scenario 4.

Therefore, nearly half of all planning failed to result in executed interventions, i.e. half observed planning behaviour was a workload cost that yielded no observable benefit (waste). Using the taxonomy of plan outcomes, it appears that the greatest planning costs are associated with the decay of plans (approximately 30% for plan-and-decay), the formation of inappropriate plans that are later discarded accounting for 20% of overall planning behaviours (plan-and-discard).

DISCUSSION

From use of the taxonomy, given the subjects modelled, the following general characterisation of planning outcomes is possible. Levels of plan execution are generally low, around one in three, and from such data it can be inferred that the style of management with the existing worksystem is very much moment-to-moment. Nevertheless, it would seem to be the case that these operators are undertaking a considerable amount of planning. The problem seems to be that plans are not being seen through to execution, 50% being the average. Therefore, half the planning effort is being wasted, an incurred worksystem cost. In extreme cases this figure rises to 69%, but can be as low as 32%. When examining the outcomes for these unexecuted plans, plan-and-decay seems to be the most frequently used category of the taxonomy. It seems that after operators plan, the lack of support for the retrieval of formed plans means that much of their expended effort is wasted. This concept of waste is reinforced when one examines the instances where a plan is formed, decays, and then is re-specified with exactly the same details for action – plan at Line 2 (Table 1) decays at Line 4, re-planned at Line 6; plan at Line 9 decays at Line 10, re-planned at Line 11, decays Line 12, re-planned at Line 13.

With regard to a gate-to-gate operational concept, and the process of early problem formulation, the taxonomy clearly assists characterisation of planning outcomes for the existing worksystem. The current worksystem, across observed scenarios, appears to support reactive management by instant execution, rather than planning. Nevertheless, the data show operators undertaking a considerable amount of planning, only half of which appear to yield a tangible management outcome. A lack of technological support for the retrieval of formed plans appears to be the most significant contributory factor, with some failure to anticipate the future state of the domain also leading to the formation of inappropriate planning.

From detailed models of worksystem behaviour, the taxonomy is therefore of assistance in characterising planning outcomes, and performance of the existing worksystem. However, several limitations of the taxonomy should be taken into account when drawing firm conclusions. First, not all decayed or discarded plans are necessarily wasted. The planning process itself will enhance operator situational awareness of the current and future state of the domain, and therefore the planning process itself is likely to yield beneficial outcomes, for non-planning tasks, in ways difficult to predict. Second, although plan decay appears wasteful, re-planning is likely to be facilitated by partial memories of a plan, and so time and effort will be reduced in planning an identical intervention on the second occasion. Finally, the taxonomy is not a complete set of categories for plan outcomes. For example, plans may not be discarded in their entirety when judged inappropriate, but rather 'repaired' (Woods, 1988), minor details of the specification (an altitude for example) being slightly modified after initial planning. From detailed behavioural models it is difficult to reliably judge when a plan is discarded or repaired, as verbal protocols are not complete in this regard.

To conclude, relationships between plans and outcomes are not as straight forward as they may appear, and in the management of dynamic processes, much effort may be wasted in the formation of plans that are never implemented as specified. Such planning behaviour is considered a cost, particularly a waste of scarce worksystem resources that is not captured by more traditional cost metrics such as time and workload. Such a measure of wasted costs is seen as complimentary to other measurements of cost, and particularly beneficial at early stages of system development, when worksystem performance is not as desired, and the process of problem formulation requires a mechanism that will support problem characterisation.

REFERENCES

Amalberti, R. & Deblon, F. (1992) Cognitive modelling of fighter aircraft process control: a step towards an intelligent on-board assistance system. International Journal of Man-Machine Studies, 36, 639-671.

David, H. (1997) Radical revision of en-route air traffic control. Eurocontrol Experimental Centre Report No.307. Available from the web:
www.eurocontrol.fr/public/reports/eecreports/1997/rep307.htm

Dowell, J. (1998) Formulating the cognitive design problem of air traffic management. International Journal of Human-Computer Studies, 49, 743-766.

Dowell, J., Salter, I., & Zekrullahi, S. (1994) A domain analysis of air traffic management work can be used to rationalise interface design issues, in Cockton, G., Draper, S. & Weir, G. (ed.s) People and Computers IX. Proceedings of HCI'94. Cambridge: Cambridge University Press.

Gray, W. D., John, B. E., Stuart, R., Lawrence, D. & Atwood, M. E. (1990) GOMS meets the phone company: Anlaytic modelling applied to real-world problems, in Diaper, D., Gilmore, D., Cockton, G. & Shackel, B. (ed.s) Human-Computer Interaction-Interact'90. Elsevier Science Publishers B.V. (North Holland).

Hoc, J-M. (1988) Cognitive Psychology of Planning. London. Academic Press.

Miailler, B. (1998) ATM Strategy for 2000+: Volumes 1 and 2. Eurocontrol Report FCO.ET1.ST07.DEL02. Available from the web: www.eurocontrol.be/ded/atmstrat

Porteous, M., Kirakowski, J. & Corbett, M. (1993) SUMI user handbook. Human Factors Research Group, University College Cork, Ireland.

Rasmussen, J. (1992) The ecology of work and interface design, in Monk, A., Diaper, D. & Harrison, M. (ed.s) People and Computers VII. Proceedings of the HCI'92 Conference. Cambridge: Cambridge University Press.

Tattersall, A. J. & Foord, P. S. (1996) An experimental evaluation of instantaneous self-assessment as a measure of workload. Ergonomics, 39, 740-748.

Timmer, P. & Long, J. (1997) Separating user knowledge of domain and device: A framework, in Thimbleby, H., O'Conaill, B. & thomas, P. (ed.s) People and Computers XII. Proceedings of HCI'97. Springer

Wioland, L. & Amalberti, R. (1998) Human error management: towards an ecological safety model. A case study in an air traffic control microworld, in Green, T. R. G., Bannon, L., Warren, C. P. & Buckley, J. (ed.s) Proceedings of ECCE-9, the Ninth European Conference on Cognitive Ergonomics. Limerick, Ireland. European Association for Cognitive Ergonomcs.

Woods, D. D. (1988) Commentary: Cognitive engineering in complex dynamic worlds, in Hollnagel, E., Mancini, G. & Woods, D. D. (ed.s) Cognitive Engineering in Complex Dynamic Worlds. Academic Press.

This page is left intentionally blank.

Session #5:
Flying Machines

Postural instability as indicator of the effectiveness of an artificial horizon with optic flow in aiding spatial orientation during flight

Lars Eriksson

Defence Research Establishment, Division of Human Sciences, Linköping, Sweden

Claes von Hofsten

Uppsala University, Department of Psychology, Uppsala, Sweden

ABSTRACT

Spatial disorientation during flight (Benson, 1978a, b) is a critical cause of aircraft mishaps and even pilot and aircraft losses (Gillingham, 1992; Gillingham & Previc, 1993). Most important in supporting the pilot's spatial orientation are the instruments that in various ways, more or less directly, represent the position of the aircraft relative the ground and its horizon. During low visibility, however, the risk for spatial disorientation increases dramatically even though several visual aids exist in the cockpit. Because of this shortcoming von Hofsten (1997) proposes a visual aid primarily consisting of an artificial horizon enhanced with a textured ground beneath it.

The idea is that the textured ground generates optic flow (Gibson, 1950, 1979; Lee, 1980), an effective visual source of information for maintaining spatial orientation that has been shown to even dominate information input from other senses in more normal circumstances (Lee & Lishman, 1975; Lishman & Lee, 1973), i.e., visual motion causing postural instability. By varying different characteristics of optic flow combined with an artificial horizon and study their effects on postural control, it is possible to get indications of which properties of the visual display that are most effective in affecting spatial orientation. This gives primary indications of how such an artificial presentation should be constructed to have the potential to counteract or reduce spatial disorientation in the flight situation.

One in a series of such experiments is presented here. The experimental setup included three computer monitors connected and positioned so that the displays covered a visual field of approximately 150° horizontally and 40° vertically. A head-tracker system registered the translational and rotational X, Y, Z changes of the participant's head position at 30 Hz. The participant was positioned in the "Sharpened Romberg Stance", erect stance with one foot in front of the other, and fixated the display center during the presentation sessions. A total of 16 participants were used. The design was a 5 x 4 x 2 factorial within subjects; five different display configurations each with four forward motion velocities and two roll velocities of the moving horizon. Every display configuration consisted of white rectangles distributed on a black ground with increasing texture element density towards a clearly defined horizon (texture gradient), and a starry sky. One condition provided a full-view display and the other four included a central field with no presentation (except a fixation cross), the size of which was either 10° x 10°, 20° x 20°, 30° x 30°, or 40° x 40°. Each of these five display configurations was presented in separate sessions, with every presentation session including: beginning and end with no forward motion but with roll motion sequences included, and in between, slow and fast forward motions combined with roll motion sequences.

Overall, the results show (1) greater postural instability with slow and fast forward motions, as compared to the conditions with zero forward velocity, and (2) that an area up to 20° x 20° of the central visual field can be omitted with postural instability as effectively caused as with a full-view display. Hence, not only can this be taken as a confirmation of the effectiveness of an artificial horizon with flight-adapted optic flow in aiding spatial orientation, but also as an indication of its preserved effectiveness while presented primarily in the peripheral visual field.

Performance – pre and post mission ratings

Berggren, P.

Abstract or paper not available at the time of printing.

Psychophysiological assessment of pilot mental workload

Staffan Magnusson

FOA - Defence Research Establishment, Linköping, Sweden

ABSTRACT

Using psychophysiological data to objectively measure pilot mental workload has been of great interest for several years. The present study presents psycho-physiological reactions during a critical phase in a simulated attack mission flown at the Blekinge Wing (171st fighter squadron). Nine pilots flew a predetermined, well-planned attack mission in a high-fidelity simulator. All pilots flew the exact same mission three times each. Each time, psycho-physiological data were collected as well as subjective ratings of mental workload. Heart rate and heart rate variability were measured by ECG, and eye blinks and eye movements were collected using EOG. The physiological data were collected using a digital multi-channel, portable recording device (VITAPORT 2).

In the present study, special interest is placed on the psychophysiological reactions during the most critical phase of the attack mission - the weapons' delivery. The psychophysiological measures display quite a different pattern during this phase than during the rest of the mission. The special characteristics of the psycho-physiological reactions during this phase are discussed, as well the change in reaction from the first to the third mission. The data are also compared to the pilots' subjective ratings of mental workload, as an attempt to validate the psychophysiological measures.

Pilots' understanding of situational awareness

Jens Alfredson
Saab AB
Jens.Alfredson@saab.se
Kjell Ohlsson
Graduate School of Human-Machine interaction
University of Linköping, S-581 83, Linköping, Sweden

ABSTRACT

As part of a European project, VINTHEC (Visual INTeraction and Human Effectiveness in the Cockpit), a questionnaire about situational awareness (SA) was distributed to a number of pilots. The answers were collected on seven point Likert-scale, with the opportunity to complement the answer with free text. The pilots were British, military pilots and Swedish private pilots. The idea was to better understand how pilots understand the concept of situational awareness, and to see if there were differences between how the two groups of pilots interpreted and experienced SA. The results are based on the response from 31 military pilots and 120 private pilots, of which 145 were male and 6 female, with an average flying experience of about 2000 hours.

76% of the subjects answered that they knew SA, and among these the interpretation of the concept was similar. Examples of abilities that were rated high as required for high SA, was the ability of bringing information from various sources together and good crew co-operation. Examples of feelings that the pilots connected to the notion of high SA were the feeling of feeling in control and feeling self-confident. There were no major differences among the pilots on how to interpret the concept of situational awareness. This implies that the concept of situational awareness has got strong face validity among the broad pilot community. This gives us further hope of revealing the common understanding and agreement on the essence of situational awareness.

INTRODUCTION

As part of a project, financed by the European union, VINTHEC (Visual INTeraction and Human Effectiveness in the Cockpit), a questionnaire about situational awareness (SA) was distributed to a number of pilots. The idea was to better understand how pilots understand the concept of situational awareness, and to see if there were differences between how the two groups of pilots interpreted and experienced SA. Examples of prevalent interpretations of SA were explored. Further the study aimed at detecting divergence between state of the art theory, and practitioners understanding.

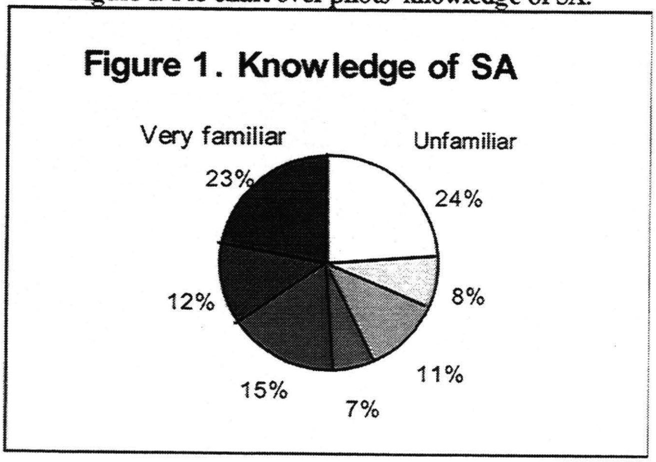
METHOD

The main part of the questionnaire consisted of questions that were to be answered by rating on a seven point Likert-scale. Each of these questions delivers suggested alternatives as well as an opportunity for the respondent to add any other alternative. In addition the respondent were able to make any other comments regarding his/her interpretation of SA, so that nothing important were to be left out. The questionnaires were written in a Swedish and an English version – see Appendix 1 for the English version. The questionnaire consisted of questions about the respondent, as age, sex, and flying experience, and also questions about the respondents understanding of the term SA, for example, abilities required for high SA. The pilots were British, military pilots and Swedish private pilots. The results are based on the response from 31 military pilots and 120 private pilots, of which 145 were male and 6 female, with an average flying experience of about 2000 hours. The evaluation of the questionnaire has, due to its construction, been concentrated to detect clear patterns among the answers, instead of a detailed statistical analysis, that runs the risk of inveigle into a thorough but misleading interpretation of data.

RESULTS

76% of the respondents answered that they knew SA, of which 23% were "very familiar" (see also Figure 1). All but one of the respondents that answered that they were "totally unfamiliar" with the term SA, were from the Swedish group of private pilots. Among the respondents that knew SA the interpretation of the concept was similar. "Literature, journals, and colleagues" were the sources that were most frequent as sources of information of which the respondents had been familiarised with SA from.

Figure 1. Pie chart over pilots' knowledge of SA.



Examples of abilities that were rated high as required for high SA, was the ability of bringing information from various sources together and good crew co-operation (see Table 1). 30.2% of the respondents added own statements to the ones already in the questionnaire, for the question about required abilities. Examples of what they added were (some statements are translated from Swedish), "anticipation of often possible actions", "imaginative ability", "put away disturbing elements", "mental preparation for stress", "the ability of correcting your mistakes", "ability to absorb or to assimilate", and "to update information".

Table 1. For the question: "Which abilities do you think are required for high situation awareness", the statements are listed according to the mean scale (from higher to lower).

Statement	Mean Value
1. Bringing information from various sources together.	6.286
2. Being aware of everything that I think is important.	5.956
3. Good crew co-operation.	5.955
4. An ability to concentrate on certain important aspects.	5.877
5. An ability to handle unexpected events.	5.805
6. Thinking of what could be about to happen.	5.789
7. Feel ready for new unexpected events.	5.619
8. Good co-operation with the air traffic control.	5.566
9. Being aware of that I have control of the situation.	5.303
10. Being able to think about many aspects at the same time.	5.196
11. Being able to relate the present situation to past experiences.	4.955
12. Good communication abilities.	4.779
13. Act without having to consciously think of what I am doing.	4.196
14. Thinking of what has happened.	4.035

Examples of feelings that the pilots connected to the notion of high SA were the feeling of feeling in control and feeling self-confident (see Table 2). 13.4% of the respondents added own statements to the ones already in the questionnaire, for the question about feelings and thoughts. Examples of what they added were (some statements are translated from Swedish), "concentration", "happiness", "thinking or planning of what to do", and "low workload".

Table 2. What the respondents think about or feel when they have high SA. For the question: "What do you think about or feel when you have high situation awareness", the statements are listed according to the mean scale (from higher to lower).

Statement	*mean value*
1. Feeling in control.	5.974
2. Feeling self-confident.	5.781
3. Feeling secure.	5.330
4. Think about what is currently happening.	5.272
5. Feeling calm.	5.202
6. A feeling of unity with the instruments.	4.991
7. Feeling something in particular.	3.633
8. Feeling of being stressed.	2.526
9. Filled with emotions.	2.437
10. Feelings fluctuate from one moment to the next.	2.219
11. Feeling aggressive.	1.947

Table 3 shows that it is situation dependent if the respondent is feeling high SA or not. 18.1% of the respondents added own statements to the ones already in the questionnaire, for the question about when they are experiencing high SA. Examples of what they added were (some statements are translated from Swedish), "When training together with others", "When there is a competition", "when the weather is bad", "when you must solve some problem", "when flying to an unfamiliar place", "At low altitude", and "air to air refueling".

Table 3. When the subjects feel they are experiencing high SA. For the question "When do you feel that you are experiencing high situation awareness", the statements are listed according to the mean scale (from higher to lower).

Statement	Mean Value
1. Depending on the situation.	5.847
2. During landing.	5.823
3. During takeoff.	5.584
4. In familiar situations.	5.380
5. During an emergency.	5.176
6. When the workload is high.	4.301
7. When nothing in particular has happened for a period of time.	2.965

DISCUSSION

There were no major differences among the pilots on how to interpret the concept of situational awareness. This implies that the concept of situational awareness demonstrates strong face validity among the broad pilot community. This gives us further hope of revealing the common understanding and agreement on the essence of situational awareness. The fact that there was not much difference among the pilots in their responses also strengthens the points that are pointed out in the further discussion below.

It is strikingly many abilities of various kind required for high SA that are rated high. This implies that SA is not easily expressed in one, or few, factors, but requires a complex mix of abilities. It is interesting that co-operation abilities such as "good crew co-operation" and "good co-operation with the air traffic control", are rated high, as well as meta cognitive abilities such as "being aware of that I have control of the situation". "Think of what could be about to happen", with a mean of 5.8, is rated higher than "thinking of what has happened", with a mean of 4.0, but both aspects are rated fairly high, indicating a wide time span associated with SA. That the abilities required for high SA are many in a complex interaction implies that SA is a complex construct. What further reinforces that notion, is that the ability that received the highest rating, with a mean of 6.3, was the ability of "bringing information from various sources together". The results imply that the respondents regard SA as more than "perception of the elements of the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" as proposed by Endsley (1988). It is clear that the respondents also relate meta cognitive aspects, collaboration aspects and that the "volume of time" regards past events also, but perhaps even more coming events.

That some feelings, such as "feeling in control", "feeling self-confident", and "feeling secure", are associated with SA is interesting, while for example a "feeling of being stressed" is not. This implies that SA is associated with feelings, by the respondents. Since feelings are subjective experiences this might also be seen as an incentive to regard subjectivity in the study of SA. This, however, does not mean that the person who is experiencing high SA is emotionally excited. On the contrary, the questionnaire shows that statements like "filled with emotions" and "feelings fluctuate from one moment to the next", rather implies that the person that experiences high SA are emotionally calm. The fact that the statement "feeling calm" got high scores, with a mean of 5.2, further reinforces that notion. While many researchers take the stand that SA is "objective", this shows that that point of regard might not be the best to reduce the gap between theoretically used concepts and practitioners notion of the concepts. Since strong face validity is a strong reason to use the concept of SA, the notions of the pilots are important.

It is interesting that the respondents state that they feel that they are experiencing high SA in very different situations, such as both "during an emergency", and "in familiar situations. Situational

Awareness Rating Technique (SART), developed by Taylor (1990), is a rating technique for SA allowing operators to rate various factors. Later Taylor developed Cognitive Compatibility - Situational Awareness Rating Technique (CC-SART) (Taylor, 1995a, 1995b). One of the factors in CC-SART is activation of knowledge, which is the amount of familiarity with the situation. From the results of this study it seems tempting to doubt the relevance in trying to "calculate" SA from "dimensions" such as familiarity, when the feeling of experiencing high SA seem to be directly situation dependent. If SA is measured directly, instead of through the calculation of pre-defined dimensions, this risk will diminish. Further the rating was high for the situations "during landing", and "during takeoff". The wide spectra of situations that the respondents rated high is further underlined by that the statement "depending on the situation" received the highest rating, with a mean of 5.8. The statement "When nothing in particular has happened for a period of time" yielded the lowest rating, with the rating of 3.0, indicating that in really low workload situations, the respondents are not very likely to feel that they are experiencing high SA. This could be interpreted as a consequence of the co-variation of low workload and low SA (e.g. Svensson et al, 1997), but could also be explained by that the sensation of SA is less intense in low workload situations, or a combination of the two possible explanations.

CONCLUSIONS

The study indicates strong face validity for the concept of SA.

The study indicates that the respondents regard SA as a complex construct, with many various requirements.

The study shows that the respondents associate SA with metacognitive aspects, collaboration aspects and a time window including past events and coming events.

The results open up the question if situational awareness is only objective. It is indicated that subjective components are an essential part of the strong notion of SA that the pilots expressed, that even might be related to feelings.

The respondents states that when they feel high SA, is more dependent on situations than dependent on a specific description category of situations, such as the level of familiarity of the situation.

REFERENCES

Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). In Proceedings of the IEEE National Aerospace and Electronics Conference (NAECON). New York, NY. IEEE.

Svensson, E., Angelborg-Thanderz, M., Sjöberg, L., & Olsson, S. 1997, Information complexity - mental workload and performance in combat aircraft, Ergonomics, 40, 362-380.

Taylor, R. M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. In Situational awareness in aerospace operations (AGARD-CP-478; pp. 3/1-3/17). Neuilly-Sur-Seine, France: NATO-Advisory Group for Aerospace Research and Development.

Taylor, R. M. (1995a). Experimental measures: Performance-based self ratings of situational awareness. In International Conference on Experimental Analysis and Measurement of Situation Awareness, Daytona Beach, 1-3 November, 1995.

Taylor, R. M. (1995b). CC-SART: The Development of an experimental measure of cognitive compatibility in systems design, Report to TTCP UTP-7 Human Factors in Aircraft Environments, Annual Meeting, DCIEM, Toronto, Canada, 12-16 June, 1995.

APPENDIX 1

The English version of the questionnaire:

The purpose of the following questions are to help the work on an EC-project concerning flight tasks in civil aviation. We hope that these questions will give us a picture of your understanding of the term "situational awareness", a term that is increasingly being used. We would appreciate your co-operation, and would be very grateful if you could send your answers to:

Jens Alfredson
Department of Human Work Sciences
Lulea University of Technology
S-971 87 Lulea, Sweden

Before the 13th of May 1996.

All data are strictly confidential.

1) What is your current occupation?

2) What flying experience do you have?

Company:

Aircraft type:

Number of hours:

3a) To what extent are you familiar with the term "situational awareness"?

1 = totaly unfamiliar 7 = very familiar
1 2 3 4 5 6 7

3b) If your answer is 2,3,4,5,6 or 7, where are you familiar with the term from?

For example: Friends Colleagues Journals Literature Letters Other:...

If your answer is 1, please go to question 7.

4) Which abilities do you think are required for high "situational awareness"?

Please respond by indicating the extent to which each of the following factors are important:
1 = not necessary at all 7 = absolutely necessary

Being able to think about many
aspects at the same time. 1 2 3 4 5 6 7

Being aware of everything that
I think is important. 1 2 3 4 5 6 7

Being able to relate the present
situation to past experiences. 1 2 3 4 5 6 7

Act without having to consciously think of what
I am doing. 1 2 3 4 5 6 7

An ability to concentrate on
certain important aspects. 1 2 3 4 5 6 7

An ability to handle
unexpected events. 1 2 3 4 5 6 7

Good communication
abilities. 1 2 3 4 5 6 7

Feel ready for new
unexpected events. 1 2 3 4 5 6 7

Being aware of that I have
control of the situation. 1 2 3 4 5 6 7

Thinking of what has happened. 1 2 3 4 5 6 7

Thinking of what could be
about to happen. 1 2 3 4 5 6 7

Good crew cooperation. 1 2 3 4 5 6 7

Good cooperation with the air
trafic control. 1 2 3 4 5 6 7

Bringing information from
various sources together. 1 2 3 4 5 6 7

Is there any other abilities you think are required for high "situational awareness"?

....
.... 1 2 3 4 5 6 7

....
.... 1 2 3 4 5 6 7

....
.... 1 2 3 4 5 6 7

....
.... 1 2 3 4 5 6 7

5) What do you think about or feel when you have high "situational awareness"?

1 = not true 7 = true

Think about what is
currently happening. 1 2 3 4 5 6 7

A feeling of unity with
the instruments. 1 2 3 4 5 6 7

Feeling calm. 1 2 3 4 5 6 7

Feeling something in particular. 1 2 3 4 5 6 7

Filled with emotions. 1 2 3 4 5 6 7

Feeling secure. 1 2 3 4 5 6 7

Feeling in control. 1 2 3 4 5 6 7

Feelings fluctuate from one
moment to the next. 1 2 3 4 5 6 7

Feeling aggressive. 1 2 3 4 5 6 7

Feeling self-confident. 1 2 3 4 5 6 7

Feeling of being stressed. 1 2 3 4 5 6 7

Is there anything else you think about or feel when you have high "situational awareness"?

....
.... 1 2 3 4 5 6 7

....
.... 1 2 3 4 5 6 7

....
.... 1 2 3 4 5 6 7

....
.... 1 2 3 4 5 6 7

6) When do you feel that you are experiencing high "situational awareness"?

1 = not true 7 = true

When the workload is high. 1 2 3 4 5 6 7

In familiar situations. 1 2 3 4 5 6 7

During takeoff. 1 2 3 4 5 6 7

When nothing in particular
has happened for a period of time. 1 2 3 4 5 6 7

During landing. 1 2 3 4 5 6 7

During an emergency. 1 2 3 4 5 6 7

Depending on the situation. 1 2 3 4 5 6 7

Is there any other situation in which you are experiencing high "situational awareness"?

....
.... 1 2 3 4 5 6 7

....
.... 1 2 3 4 5 6 7

....
.... 1 2 3 4 5 6 7

....
.... 1 2 3 4 5 6 7

7) Are there any other comments you would like to make regarding your interpretation of "situational awareness"?

8a) Sex: Male Female

8b) Age group: 17-21 21-30 31-40 41-50 51-60
61-

Thank you for your co-operation!

This page is left intentionally blank

Session #6: Safety and Errors

Is it Safe Enough? Filling the Gaps in Human Factors Certification of Flight Management Systems (FMS)

Gideon Singer

Chief Test Pilot, SAAB Commercial Aircraft, Sweden.

This paper is published, as a chapter in:
Dekker, S. & Hollnagel, E. (Eds.), "Coping with Computers in the Cockpit",
Ashgate, ISBN 0 7546 1147 7

ABSTRACT

In commercial flight, on most modern aircraft today, navigation and flight path are being controlled by a Flight Management System (FMS). This system is the interface between the crew and the aircraft navigation and auto-flight computers. When designing and certifying a new FMS installation in a new or existing aircraft, a successful process of validation is necessary in order to attain the final approval. This validation process consists of two parts, the validation of the technical merits and the validation of the interface with the operator(s) in the integrated cockpit dynamics. The technical requirements and methods for testing system performance, accuracy and fault analysis are well established and are defined in the certification requirements for such systems and aircraft. The methods for technical validation of such systems have been used successfully in many types of aircraft avionics and have shown very high level of reliability. However, validation of the way the system interfaces with the crew, displays information or reacts to crew inputs is not well defined and each manufacturer or vendor is free to adapt its own philosophy and methods of showing compliance.

Previous research has shown that problems exist with many FMS installations and that they provide fertile ground for human. Some reports have pointed out specific modes that are of special concern and might have been a contributing factor in some accidents and incidents.

These reports however, are of limited use for organizations engaged in designing or approval of a new system due to lack of constructive new methods.

In several investigations conducted since the late eighties, and especially in the 1996 FAA Human Factors Study Team Report, it becomes clear that the majority of accidents and incidents were due to Human Error rather than technical faults.
It is also apparent that the FAA report addresses mainly large, modern jet aircraft where the FMS is integrated into the initial cockpit design. The majority of today's system integrations are actually on smaller or older aircraft, where an FMS is not part of the original cockpit concept. Due to cost and schedule constraints, the design and evaluation processes are minimized to the cover mainly functional aspects, leaving Human Factor issues often deficient.

After reviewing the existing rules and guidance material it becomes clear that the existing requirements are general, vague and leave too many questions unanswered. None of the requirements include any method or criteria for compliance regarding Human Error probability and effects. The result is often written compliance statements of a successful review that are just as vague and generic and no not address the main problem we have been faced with.

Clearer methods and criteria are needed in order to capture and begin to quantify the deficiencies in human factors design and be able to mandate an improvement at an acceptable price. As long as such requirements are not available, changes will be deferred to the future and the potential for error will remain.

This paper will try to address some of the deficiencies of FMS interface in a typical "add-on" designs, explain the potential for error and the possible effects of such errors. It will then show the limitations of today's certification requirements in regulating such deficiencies and suggest new requirements for future use. An attempt will be made to give specific guidelines for validation methods with Pass/Fail criteria without mandating a specific system architecture or logic.

Time-To-Collision and Action Sequencing on Aircraft Conflicts in Air Traffic Control

Thierry Morineau
LAMIH – Equipe PERCOTEC
University of Valenciennes and Hainaut-Cambresis
Le Mont Houy – BP311 59304 Valenciennes, France
Thierry.Morineau@univ-valenciennes.fr
http://www.univ-valenciennes.fr/LAMIH

ABSTRACT

An analysis of data from an experiment performed on a simulator by Air-Traffic Control operators is presented. This analysis focused on Time-to-collision (TTC) of each aircraft conflict occurring when the radar controller performs a first action of conflict resolution. Results indicate the importance of this priority criterion in the Radar Controller's action sequencing, with some inter-individual differences. We note that when this external criterion is not respected, frequently the conflict having priority has not been detected at this time and thus was outside of the Situation Awareness of the operator.

Keywords

Air Traffic Control, Time-To-Collision, Sequencing, Context, Situation Awareness, Cognitive control.

INTRODUCTION

"En route" Air Traffic Control presents a task with many functions that must be carried out for the operator. Indeed, the Radar Controller (RC) must perform several tasks successively or in parallel, such as traffic organization in entering or leaving the sector, supervising in order to optimize aircraft trajectories and to avoid conflicts, and finally conflict resolution between aircraft. This work leads the operator to a situation where time-sharing strategies must be engaged. The Radar Controller benefits from the help of the Planning Controller (PC), whose job is mainly to regulate the aircraft flow with the other sectors. Naturally, a task is composed of several subtasks, such as exchange with command-control system, communication with the other operator or with aircraft pilots or sectors. Even though Air Traffic corresponds to a continuous process, some discrete states can be highlighted in Air Traffic Management, which could constrain the controller's activity (Morineau, 2000).

Hopkin (1995) suggests that different tasks and subtasks in Air Traffic Management requires the RC to arrange them. This activity of sequencing would depend on a set of constraints. He highlights 5 kinds of constraint:

- Constraints related to dependent functions: for instance, a conflict has to be detected before it can be resolved;
- constraints related to division of work: the action of one controller can be achieved only after the other controller has completed his own action;
- constraints related to methods of information presentation, which drive the procedure of the action;

- constraints related to priorities: an emergency action has to be performed before routine ones;
- constraints related to tools or tasks already done, which bring availability for doing another action: for instance, the availability of a communication channel with a pilot.

Constraints correspond to a sequence of events or actions which reduce the number of degrees of freedom in the activity of the operator. Hopkin (1995) notes that these main constraints are particularly concerned with the organization of traffic entering or leaving the sector. For this specific task, the timing of controller actions would depend particularly on aircraft positions. Otherwise, the controller would have "a considerable discretion over the sequencing of control actions within the sector." (p.110). Also, we can address the issue of the range of degrees of freedom available for the operator during activity of conflict resolution within his sector.

Indeed, the speed of the Air Traffic process represents an important temporal constraint, which could particularly decrease the range of possibilities in action sequencing; some conflicts having priority. And from a cognitive point of view, we know that the speed of the process is a contextual parameter, which partly determines processes engaged by the operator (Hoc & Amalberti, 1995). Also, we can make the hypothesis that when the RC detected several conflicts, he executes a first action of resolution on the conflict having priority. The criterion by which priority is determined would be the Time-To-Collision (TTC) of the conflict, compared with the other conflicts occurring. We can find the study of this concept of TTC in the domain of vehicle driving, where it is used to evaluate cognitive aspects of traffic accidents (see for instance, Stewart, Lishman, & Cudworth, 1996). In Air Traffic Control (ATC), this criterion could be assumed cognitively by the operator toward an exo-centered spatio-temporal representation of the traffic, allowing prediction on future points of conflicts (Sperandio, 1976). This current mental representation (Hoc & Amalberti, 1995) or Situation Awareness (Endsley, 1995a) about the global situation controlled permits the RC to optimize the timing of his actions on conflicts, allowing globally a time saving and thus a possible regulation of an incorrect action taken on a certain conflict.

From an engineering point of view, the knowledge of such a criterion would allow us to make some recommendations for the design of cooperation modalities between operator and a computer support system.

To test our hypothesis, we have observed the activity of RC during the simulation of an Air Traffic. For each first action of conflict resolution, we measured the Time-To-Collision of each potential current conflict at the moment of the action execution. Note that this analysis had been set up on an experiment designed for another purpose (see Crévits, Debernard & Denecker, 1999).

METHOD

The experimental situation consisted in a simulation of an Air Traffic in the French C1/C2 sector of Bordeaux. This scenario's simulation was fulfilled in the frame of a computer system development called AMANDA, in the LAMIH Laboratory. This project is managed by the Centre d'Etude en Navigation Aérienne (CENA).

6 pairs of controllers participated in the experiment. The Air Traffic simulated corresponded to realistic weekend traffic but particularly charged (70 aircraft per hour). The control activity for a pair of controllers lasted 30 minutes, with 15 experimental interruptions of 2 minutes. During these pauses, the radar screen was dark and the simulation interrupted. Then, the Radar Controller was interviewed about his traffic representation. This method has been assessed as not intrusive (Endsley, 1995b). The content extracted from these interruptions has not been studied in the present research (see Crévits et al, 1999). Each session began with a period of formation in order to

facilitate the use of specific functions defined in the simulator, such as strip displayed on screen, computer mouse, and automation of communication between sectors and with aircraft. This latter function corresponds to a simulation of data link system, allowing direct computer contact between ATC and aircraft

With the help of the simulator system and the use of the MacShapa software package, we recorded and computed data on actions executed by the RC in response to conflicts appearing in the simulation. We took into account exclusively actions, which had been engaged in the following context:

- The action was executed on conflicts occurring on beacon;
- the target conflict embedded aircraft that were at a same level of flight;
- the target conflict was an actual;
- conflict and not a false alarm;
- it was the first action of resolution on the target conflict;
- at least two conflicts occurred during the action.

Note that when a small delay existed between aircraft in conflict on a target beacon, we took into account the timing of the faster aircraft. This information allowed us to calculate a Time-To-Collision index (TTC) concerning each current conflict, when the RC performed a first action on a single conflict.

Moreover, we realized a coding which was based on a predicate-argument formalism, allowing identification of some cognitive aspects of activity. This coding used a coding scheme developed in the LAMIH laboratory (Hoc & Carlier, 1998; Hoc & Lemoine, 1998). Three types of cognitive operation concerning conflict resolution activity have been highlighted: The first one, the conflict identification by the RC is inferred on first elements of verbalization about conflict. The second one concerned decision about first action of conflict resolution. And the third one coding focused on the first action of resolution effectively performed. These cognitive operation codings permitted us to determine the cognitive context in which a single action took place.

RESULTS

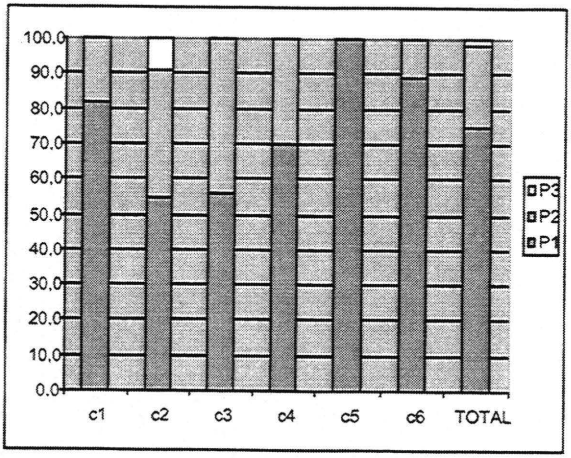
Before examining issues, note that c5 and c6 radar controllers showed some salient difficulties during their activity. Radar Controller called "c5" presented problems of cooperation with his Planning Controller colleague. And RC called "c6" apparently was overloaded and showed difficulties in monitoring the traffic.

We observed the level of respect of the TTC criterion by our sample of controllers in their first actions on aircraft conflicts. An average amount of 75% (± 9.3%)[13] of cases, in which the controller coped with a set of conflicts, acted on the conflict having priority in regard to the TTC criterion. 23% of actions are executed on conflicts with a secondary level of priority and 2% of actions concerned conflicts coming on a third position.

---

[13] Half confidence interval at the .10 level

Figure 1: Pertentages of action on conflicts as a function of the level of Time-To-Collision priority (P1, P2, and P3) for each radar-controller (c1 to c6)



We remark that conflicts having no priority stand mainly at a second level of urgency. Data show only one case of action, with which the level of priority is 3. The respect with which the TTC criterion is held seems to vary as a function of the operator. If we look at individual differences on this criterion, we observe that c5 and c6 controllers, who were concerned respectively with cooperation and traffic monitoring problems, have respected TTC the most frequently.

Action on priority conflict

In an absolute manner, the amplitude of delay before collision on beacon accorded by operators on P1 priority conflict appears as relatively restricted (Table 1).

Table 1: TTC Means and Standard Deviations for action on P1 priority conflicts (minutes)

Controller	N	Means	S.-D.
c1	9	9.5	2.9
c2	6	8.7	3.1
c3	5	6.3	3.6
c4	7	6.3	2.7
c5	9	7.3	2.5
c6	8	6.8	1.0
	44	7.6	2.8

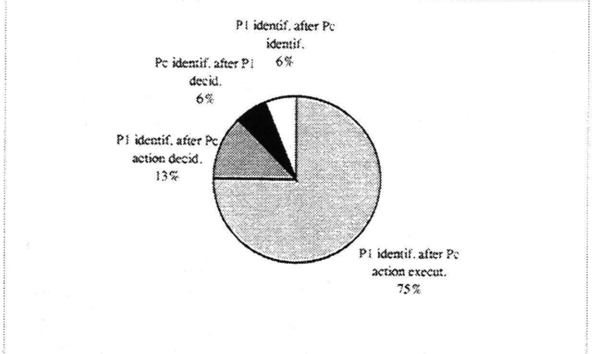
An average time of 7.6 minutes (± 0.7 mn) with a Standard Deviation of 2.8 minutes is available before the situation becomes fatal, if no action is executed. But if we consider this value relatively to the average time necessary for an aircraft to cross the sector, which is about 10 mn, this action execution delay seems quite important. If we compare the percentage of actions executed on P1 priority conflicts with standard deviation of TTC, we obtain a significant negative correlation (Rho =.83; p=.04). The more operators respect frequently the TTC criterion in conflict action ordering, the more the variability of their TTC values is reduced.

It would means that operators, who follow strictly the TTC criterion, are more rigid on the timestamp accorded in conflict resolution engagement. Moreover, it appears that c5 and c6 radar controllers who presented some problems during their activity and respected the TTC criterion the most frequently show the smaller standard-deviations, especially c6 controller who was overloaded (S-D.= 1 mn).

Action on no priority conflict

Actions to solve conflicts, which do not have priority, are not very frequent (25%) but are frequent enough to require a specific analysis. We attempted to put this kind of action back in the cognitive context, in which it took place. To perform it, we took into account cognitive operations about the target conflict of the action and about conflict having priority at the same moment.

Figure 2: Non priority conflicts action as a function of cognitive context (P1: Priority conflict, Pc: Target conflict for the resolution action)



In 75% of cases (12/16), the priority conflict has not been yet detected when the action on the target conflict is achieved. It means that the priority conflict was not included in the traffic current representation or Situation Awareness of the RC. In addition, we note two cases for which the detection of the priority conflict is made after a decision of action on the target conflict is fulfilled.

DISCUSSION

Results show the importance of the Time-To-Collision criterion for conflict resolution actions arranged by the Radar Controller. This flow of action seems to be structured by the degree of emergency involved by each current conflict. Study of the cognitive context in which this criterion is not respected allows us to examine evidence on the subjective aspect of the TTC observance. It depends on the content of the Situation Awareness (Endsley, 1995a) or current representation (Hoc & Amalberti, 1995) held by the operator. The operator regulates action on the basis of conflicts represented in his Situation Awareness. The cases where this criterion is not respected correspond to an incomplete Situation Awareness. In this case, the Radar Controller's decision doesn't fit the traffic pattern because of the inadequate content of the Situation Awareness, even though the decision process is correct (Endsley, 1996).

However, the actual status of this criterion in the ATC activity is not necessary evident with regard to two dimensions of activity analysis: the performance effectiveness and the cognitive level of control on action.

Concerning the performance effectiveness, we observe that two controllers who apparently showed a deficiency in performance, caused respectively by human cooperation interference and difficulties in air-traffic monitoring, were the first to follow the TTC criterion in their action sequencing. Moreover, we find that in general this technique of ordering correlates with a greater rigidity in the average stamp time accorded to priority conflicts. They follow the Time-To-Collision constraint more strictly. According to a reviewer of the present paper, driving studies would show that skilled drivers may not care to optimize Time-To-Collision and can navigate more freely within the time-space envelop. In the same manner, rigidity with respect to the Time-To-Collision criterion could indicate a less skilled Air Traffic Management performance.

Concerning the level of cognitive control in activity, we can examine the proposal that it corresponds to an internal criterion known by the Radar Controller and, which consciously governs modalities of action. A possibility is that it is more of an external temporal constraint, which triggers the selection of management schemas between conflicts, in an automatic manner, and for each context where making an action is necessary (Turner, 1998). An evidence for this explanation may be found in the study of Situation Awareness in ATC. Mogford (1997) studied the explicit recall of information by controllers, which constitutes an important aspect of their activity. It appears that the aircraft speed is not considered as a priority dimension to take into account by operators. Conversely, Spérandio (1976) shows that controllers particularly appreciated spatio-temporal information on a specific strip in their activity. A cognitive processor such as the Contention Scheduling System that Norman and Shallice proposed in their model could assume an automatic selection in activation of conflict processing schemas as the function of the level of conflict urgency (Norman & Shallice, 1986; Shallice, 1982). This cognitive interpretation would permit to explain why radar-controllers, who present a poorer effectiveness in their activity, are those, who respect the TTC criterion in a rigid manner. Their action is mainly based on the triggering of schemas, activated by salient constraints. More efficient operators would take into account the TTC constraint but would engage it with more flexibility as the function of other strategies of sequencing.

Finally, the intrusion of the current representation or Situation Awareness for taking or not into account conflict during sequencing allows us to suppose different levels of cognitive controls in the activity. A cognitive problem of aircraft conflicts sequencing can be approached at different levels of problem solving. At a low-level, the problem involves strictly respecting a Time-To-Collision criterion for short-term emergency management. At a higher level, this sequencing depends on the content of the current process representation, which is frequently updated and depends on more long term planning strategies (Hoc & Amalberti, 1995). We are now attempting to evaluate this hypothesis with several other indicators of medium and short-term cognitive control.

Nevertheless, this first result on TTC criterion in ATC brings us a way to observe the controller activity. In fact, this indicator would serve us to guide the design of an assistance system, which takes into account this criterion to be more cooperatively in its interaction with the operator. In practical terms, we plan to design a hierarchical organization of conflict information as a function of this criterion. But we do not forget that it could represent a particular level of cognitive air-traffic management, which does not interfere with specific mental representations or higher-level sequencing strategies engaged by some operators.

ACKNOWLEDGMENTS

This study was supported by the French research center on air traffic control (CENA). I thank I. Crévits, S. Debernard, P. Denecker, J. M. Hoc, T. Poulain for their collaboration on this study. I also thank the controllers for their participation.

REFERENCES

Crévits, I., Debernard, S. and Denecker P. (1999) DSS Dedicated to Supervision : The Case of Air-Traffic Control. Human Centred Processes, HCP'99, Brest : September 1999.

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

Endsley, M.R. (1995b) Measurement of Situation Awareness in Dynamic Systems. Human Factors, 37, 65-84.

Endsley, M.R. (1996) The Role of Situation Awareness in Naturalistic Decision Making. In C.E. Zsambok & G. Klein (Eds). Naturalistic Decision Making. Mahwah : Lawrence Erlbaum, 269-283.

Hoc, J.M. and Amalberti, R. (1995) Diagnosis: Some theoretical questions raised by applied research. Current Psychology of Cognition, 14, 73-101

Hoc, J.-M. and Carlier, X. (1998) A method to describe human diagnosis strategies from a cognitive point of view. 9th European Conference on Cognitive Ergonomics, ECCE-9, Limerik, Ireland : August 1998.

Hoc, J.-M. and Lemoine, M.-P. (1998) Cognitive Evaluation of Human-Human and Human-Machine Cooperation Modes in Air Traffic Control. The International Journal of Aviation Psychology, 8, 1-32.

Hopkin, V. D. (1995) Human Factors in Air Traffic Control. London : Taylor & Francis.

Mogford, R.H. (1997) Mental Models and Situation Awareness in Air Traffic Control. The International Journal of Aviation Psychology, 7, 331-341.

Morineau, T. (2000) Approche de la discontinuité dans l'activité de contrôle d'un processus continu: Illustration dans le domaine du contrôle de trafic aérien. Short paper for Ergo-IHM'2000, Biarritz, F, Oct.

Norman, D.A., and Shallice, T. (1986) Attention to action: willed and automatic control of behavior. In G.E. Schwartz and D. Shapiro (Eds.) Consciousness and self-regulation, vol. 4, New-York: Plenum Press.

Shallice, T. (1982) Specific impairments of planning. Philosophical Transactions of the Royal Society of London, B298, 199-209.

Spérandio, J.-C. (1976) De l'espace-plan à l'espace mobile aérien, comparaison expériementale entre deux supports d'informations spatio-temporelles. Le Travail Humain, 39, 139-154.

Stewart, D.L., Lishman, J.R. and Cudworth, C.J. (1996). An alternative Source of Time-To-Collision, in A.G., Gale, I.D. Brown, C.M., Haslegrave, S.P., Taylor (Eds). Vision in Vehicles – V, Amsterdam : Elsevier Science, 51-60.

Turner R.M. (1998) Context-mediated behavior for intelligent agents. International Journal of Human-Computer Studies, 48, 307-330.

An explanation of human errors based on environmental changes and problem solving strategies

José F. Quesada, José J. Cañas and Adoración Antolí

Department of Experimental Psychology
University of Granada
Campus de Cartuja
18071 Granada, Spain

jose@geneura.ugr.es,
delagado@goliat.ugr.es,
antoli@platon.ugr.es
2http://www.ugr.es/~delagado

ABSTRACT

People that show good performance in problem solving tasks make also errors. Psychological theories of human error predict that those errors are to some extend the consequence of the difficulties that these people have to adapt to new environmental conditions. This paper describes two experiments and a research methodology designed to test this hypothesis.

Keywords

Human error, Complex Problem Solving, Microworlds, Transitions Between Actions, Firechief.

INTRODUCTION

Most psychological theories (i.e. Norman, 1981; Norman & Shallice, 1980; Rasmussen, 1983; Reason, 1990; Hollnagel, 1998) proposed to explain and predict human error share certain characteristics: (1) people loose conscious control when they increase their ability at performing a task; (2) there is a hierarchical structure (schemas, semantic network or control levels) in which higher levels include, organize and control lower levels; (3) practice and elaboration let to a representation which hides the process details that can lead to a lack of flexibility. (4) there is a trade – off between quick, fluid actions and controlled, flexible actions.

These psychological theories seem to agree on the idea that in order to avoid a human error you need to realize that the situation has changed to be able to `log out' the automatic processing mode and come into the controlled processing mode. To detect the situation change and the necessity of a non – routine response, it is necessary to come into a higher level of attentional control, where you access the new situation and plan the action to be taken. You need to perceive environmental cues in a different way, reinterpreting them. How the person represents the task and the set of strategies employed to deal with it determine how easily she/he shifts attention to the new environmental conditions.

For example, Rasmussen (1983) proposed a theory that can be used to frame this idea. He distinguished three levels of processing: (1) skill based level, for activities we do in an automatic way, (2) rule based level, for situations in which our experience gives us a response in a known situation, and (3) knowledge based level, for new situation in which there are no rules and we need to plan a different response. Within this framework, practise in a task leads to more automatic activities at the rule and skill based levels. However, when the task conditions change a person must do activities accomplished at the knowledge based level to adapt to those new conditions. In general, when a person is skillful would have problems in detecting the environmental changes and to change her/his strategy to adapt to these changes.

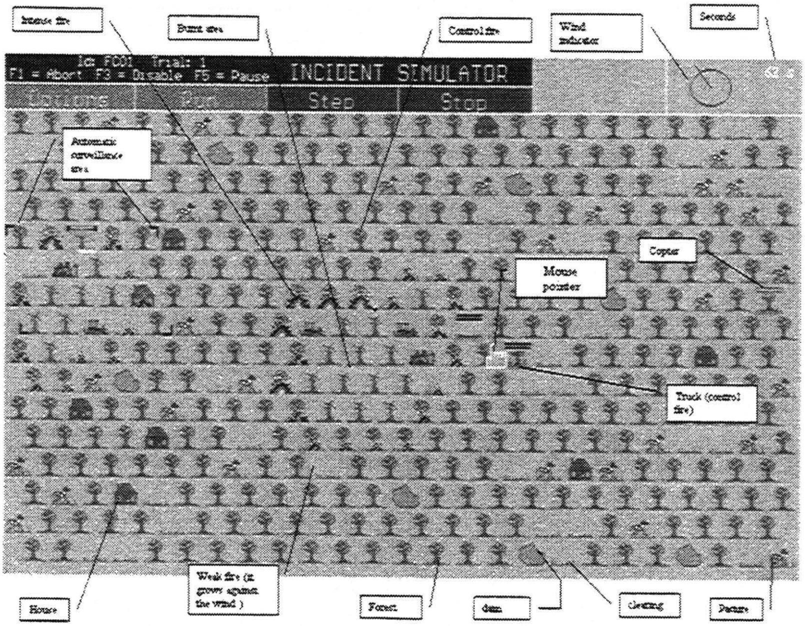
The present research was designed to test and pursue this hypothesis by developing an experimental program in which people learn to perform a complex problem solving task. The conditions in the task remains constant for some time and people have the opportunity to develop strategies to deal with them. At one point, those conditions change and the effect of that change on their performance is observed. We hypothesized that performance would be impaired when the environmental change affects the particular strategy people develop.

To prove this hypothesis it is necessary to elicit planning mistakes or errors related to changes from the inferior levels to knowledge based level. This is only possible in problem solving tasks where planning and decision making abilities are more important for optimum performance than perceptual and motor abilities.

Microworlds are complex problem solving tasks that are the appropriate research environment to test this hypotheses. Microworlds are based on simulation of task changing dynamically and they are prepared to reproduce the important characteristic of the real situations (the state of the problem changes autonomously as a consequence of the actions of the subject and the decisions must be taken in real time) but leaving open the possibility to manipulation and experimental control.

Our work is based on Firechief, a microworld generation program created by Omodei and cols. (Omodei and Wearing, 1995),

Figure 1: Screenshot of Firechief (Omodei and Wearing, 1995)



In Firechief participants find a screen that simulates a forest where the fire is spreading. Their task is to extinguish the fire as soon as possible. In order to do so, they can use helicopters and trucks (each one with particular characteristics) that can be controlled by mouse movements and key presses. The different cells (see figure 1) have different values of flammability and points: houses are more valuable than forests, for example. The participant's mission is to keep save as much forest area as possible, but preserving more the most valuable cells and preventing the trucks to be burnt. Helicopters move faster and drop more water than trucks, but the latter can make control fires. Trucks are unable to extinguish certain fires, and they will be destroyed if they are sent there. The fire is more intense and spreads faster in the direction the wind blows. Participants can see a window with their overall performance score at the end of the trial. This figure is calculated adding every safe cells and subtracting the value of the trucks burnt. The task is complex and participants feels interested from the beginning to the end. At the same time, it is possible to control experimentally every single feature of the system, and to prepare experimental situations for checking a wide variety of hypotheses.

There are four commands that are used to control the movement and functions of the appliances: (1) Drop water on the current landscape; (2) start a control fire (trucks only) on the current landscape segment; (3) move an appliance to a specified landscape segment; (4) search in a specified portion of the total landscape area. Commands are given by first selecting the desired vehicle (by moving the mouse cursor into the landscape segment containing it) and then pressing a key on the keyboard. At the end of each trial, the program saves the command sequence that the participant issues in that trial.

In the two experiments described in this paper participants performed the task in a situation in which the environmental conditions remained constant for some time. The wind was blowing always toward the East in the first sixteen trials. In the last four trials the wind changes direction. If our hypothesis was correct, the participant with better performance would be those that experience a decrease in their performance, while those with worse performance would adapt to the change. However, this effect would depend on the problem solving strategy adopted by each participant. There could be strategies that allow participants to keep a good performance level in spite of the change, while other strategies will prevent them adapting to the new situation. Since the fire spreads faster in the direction the wind blows, the strategies most affected by the change would be those that are based on predictions of where the fire will spread.

EXPERIMENT 1

In the area of traditional Problem Solving, where there is assumed a limited problem space, a usually well-defined goal, and an only way of reaching the goal in the smaller number of steps is relatively easy to identify the strategies that a person adopts. However, in most Complex Problem Solving tasks this identification is more difficult since it does not exist the optimum strategy, and furthermore the protocols of the participant are so wide in data that they have probably been produced by more than one simple strategy at the same time.

For this reason we have devised a method of analysis to identify the strategy that the participants adopted in our experiment. The method consists basically of comparing a participant's behavior with that of a simulated person who would adopt a hypothetical strategy.

The protocol output by Firechief at the end of each trial contained the sequence of commands that a participant issued during that trial (e.g. move a truck, drop water). Those commands could be used to construct a matrix of transitions between actions. Rows and columns in the matrix represent the command and the cells contain the number of times that one command follows another. This matrix contains important information about problem solving strategies since transitions between actions reflect how a person issues the commands (Howie and Vicente, 1998). Therefore, our method for inferring participants' strategies is based on the analysis of matrices of transitions between actions and is done following a number of steps.

1. Construct empirical matrices of transitions between actions

For each trial of each participant, we obtained a protocol file where all the actions were registered, in a temporal sequence. Afterwards, we extracted every transition between two actions, and put them in a symmetrical matrix with size equal to the number of possible actions.

2. Design a set of theoretical, simple strategies based on a task analysis

The Firechief program has a simulation module that allows the implementation of problem solving strategies. This module permits one to introduce code in Pascal, and provides a function library to facilitate the design of those strategies. When recompiled, the simulation module takes the programmed strategy and accomplishes the task as a participant that has adopted it would. Therefore, the program generates equivalent protocol files to those that would be generated when a human participant accomplished the task. That is to say, we could have a protocol from a simulated participant that has performed the task with a single hypothetical strategy. And from this protocol we could also obtain a matrix of transitions between actions.

At this state of our research we have devised three possible strategies participants might be using:

Move and drop water (SLRN): People using this strategy move appliances to the closest unattended fires and drop water there. Trucks are not sent to fires too fierce where they could be destroyed

Control Fires (SCON): This strategy can only be used by trucks. It finds the closest fire and then sends an appliance to deliberately light small fires in a location two segments away from it in a randomly chosen direction. Before that, the algorithm checks that the location is unburned, unoccupied and no burning.

Setting up an automatic vigilance area (MISC): An automatic vigilance area is a rectangular zone that the participant can draw in the map after selecting an appliance. If the appliance has water and a fire is found in an adjacent segment within the area, the appliance moves into that segment and automatically drops water on the fire. If the appliance is empty of water and a dam is found, it automatically fills with water. If none of these conditions apply, the appliance will move one segment in a random direction and the above search process repeated.

These strategies were simple, easily distinguished and orthogonal in the sense that the matrices of transitions generated by them did not correlate.

3. Correlate empirical and theoretical matrices

When we introduced the theoretical matrices as predictors of participants' matrices of transitions between actions it was possible to identify which one of them was used by her/him. Then, to evaluate the possibility that one participant had used one particular strategy we calculated the similarity between her/his empirical matrix and that obtained from simulating the strategy. A significant correlation between those two matrices meant that to some extent that strategy was responsible for her/his performance. Therefore, we performed a multiple regression analysis with the empirical matrix as the dependent variable and the simulated strategies as the predictors. The Betas in the analysis represented the partial correlation between the strategies and the performance of the participant. For example, if one participant adopted the SLRN and SCON strategies in one trial, we should find significant betas (alpha = 0.05) for the matrices representing these strategies.

4. Classify participants according to which strategy they used

We built a rectangular matrix representing each participants' similarities with the theoretical strategies. Rows represented the participants and the three columns represented the strategies. One cell in the matrix had a value of 1 if that participant used the corresponding strategy. Finally, we performed a cluster analysis on that matrix to group participants with similar strategies.

Therefore, what we have after applying this method is a classification of participants into groups based on the similarity of their strategies. This grouping could be used to as a quasi-experimental independent variable in a factorial design to evaluate its interaction with other manipulated independent variables.

METHOD

Participants

Thirty-seven students at the University of Granada participated in the experiment as part of class requirement.

Procedure

Participants were asked to play 20 trials, where 16 of them kept wind direction constant and the last 4 had variable wind conditions (wind changed from east to west slowly). They did not know that beforehand.

RESULTS

We did the analysis using only the 4 trials before the wind change and the 4 trials in which the change was experienced.

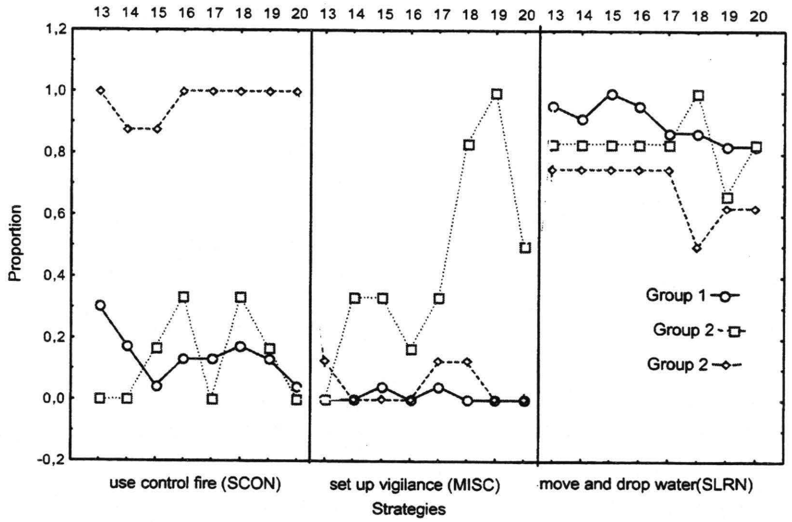
In order to group participants in accordance with their strategies we did a cluster analysis with the data of significant correlation between their strategies and the theoretical ones in each trial, coded as zero and one.

The cluster analysis that gave better discriminant results produced 3 groups (see Figure 2). The groups had an unequal number of participants: 23 participants belonged to group 1, 6 belonged to group 2 and 8 belonged to group 3. The grouping could be interpreted as follows:

Group 1 Participants whose strategies rarely correlate with SCON or MISC. They mainly adopt a strategy similar to SLRN.
Group 2 Participants who use mainly SLRN, but react to the wind change increasing MISC- like strategies.
Group 3 These participants use neatly SCON in every single trial, though they also use SLRN and MISC a little bit.

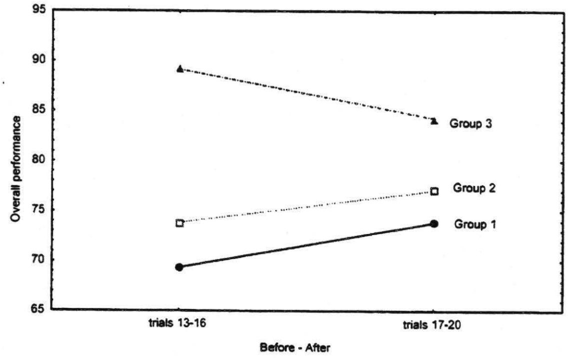
To support this interpretation with statistic arguments, we did three analyses of variance (ANOVA), one for each strategy, with three independent variables, groups, trials and before-after change. The former was a between groups variable and the other two were within-subject variables. In this analysis, the SLRN strategy did not discriminate between groups neither showed significant effects in trials or in wind change. The SCON strategy discriminated between groups, $F(2,39) = 73.32$, $Mse = .2357$, $p<0.01$, and showed an interaction between trials and before-after change, $F(3,102)=13.84$, $Mse = .0719$, $p < 0.05$. This interaction express little variations in groups 1 and 2 that are not relevant to our objectives. Finally, the MISC strategy also discriminated between groups, $F(2,34)= 63.218$, $Mse = .0059$, $p<0.01$, and showed an effect of before-after change, $F(1,34)= 40.075$, Mse=0.357, p<0.01, and trials, F(3,102)=5.192, Mse = .0457, p<0.01. The interaction of groups by before-after was also significant, F(2,34)=28.717, Mse=1.025, p<0.01 and an interaction of groups by trial, F(6,102) = 6.784, p < 0.01. These results can be interpreted as saying that the only group who increase their use of MISC is group 2, and they do this after the change. The LSD tests showed that the interaction was not significant for groups 1 and 3.

Figure 2: Proportion of times correlation between theoretical and empirical strategies becomes significant (trials 13 – 20)



Once we have grouped participants, we used these groups to perform an ANOVA using the overall performance scores[14]as the dependent variable. The independent variables were groups (between groups, 3 levels), trials (within-subject, 4 levels) and before-after change (within-subject, 2 levels). Results showed a significant effect of group, F(2,34) = 6.84, Mse = 786.71, p < 0.01 (observed power: .897) and a significant effect of the interaction between group and before-after change, F(2,34) =3.52, Mse =155.04, p< 0.05 (observed power:.618, see figure 3).

Figure 3: Group by before - after wind change interaction



---

[14] Defined as the sum of all cells that remain unscathed subtracting the value of all burnt trucks. It was expressed as a percentage of the total area.

Thus, the different group formed in accordance with the differential use of strategies, were also different in overall performance. Planned comparisons showed that group 3 had superior overall performance scores than group 1 and 2 .F(1,34)=9.84, mse=786.713 $p< 0.05$. Group 1 and 2 were not significantly different, F<1. The wind change affects the groups differently. The change worsens group 3 performance, who were the best participants, although this difference did not reach significance F (1,34) = 2.58, Mse = 155.036, p=0.116. In group 1 participants gets better, F(1,34)=6.961, Mse = 155.036, $p<0.01$. In group 2 participants did not experience any change F<1.

The variable before-after difference in significance was probably owed to the bigger size of group 1(n=23), which declared significant an increment of the same magnitude as the group 2 experienced (n= 6).

Experiment 2

Experiment 2 was designed to replicate experiment 1, but with an important modification. We decided to eliminate the physical possibility of doing automatic vigilance area, and, thus, of doing a strategy similar to MISC. We had two reasons for doing this: (1) the most interesting results in experiment 1 were regarding group 3, that is, people who used control fires (SCON). This strategy could be responsible for the decrease in performance, so we were interested in forcing participants to use it. A simple, indirect way to do so is by eliminating one of their alternatives (MISC). (2) The MISC strategy gave participants the opportunity of delegating good amount of their work through automatic searching for fires.

METHOD

Participants

Twenty-four students at the University of Granada participated in the experiment as part of class requirement.

Procedure

Like in experiment 1, participants were asked to play 20 trials, where 16 of them kept wind direction constant and the last 4 had variable wind conditions (wind changed from east to west slowly). They did not know that beforehand.

The only difference with experiment 1 was that participants were not allowed to use automatic vigilance areas. That option was unavailable during the trials, and it was omitted in the instructions that participants read.

RESULTS

Again, we did the analysis using only the 4 trials before the wind change and the 4 trials in which the change was experienced. To group participants in accordance with their strategies we did a cluster analysis with the data of significant correlations between their strategies and the theoretical ones in each trial, coded as zero and one.

The cluster analysis that gave better discriminant results produced 2 groups of 12 participants each. The grouping could be interpreted as follows (see figure 4):

Group 1: Participants whose strategies significantly correlated with SLRN, that is, they normally moved their appliances and dropped water.

Group 2: Participants whose strategies were more based on SCON (used control fire) than in SLRN.

To interpret cluster analysis results and get a statistical reinforcement to these statements, we did two ANOVAs,, one for each theoretical strategy. We used three independent variables, groups, trials, and before-after change. The results showed that both SCON and SLRN discriminated between groups, $F(1,23) = 5.117$, $Mse = .715$, $p < 0.05$, and $F(1,23) = 191.07$, $Mse = .1407$, $p < 0.01$ respectively.

Using these 2 groups created in cluster analysis, we did an ANOVA using overall performance scores as dependent variable. The independent variables were
Groups (between groups, 2 levels), Trials (within- subject, 4 levels) and Before-After change (within- subject, 2 levels).

Figure 4: Proportion of times a correlation between theoretical and empirical strategies becomes significant (trials 13 - 20)

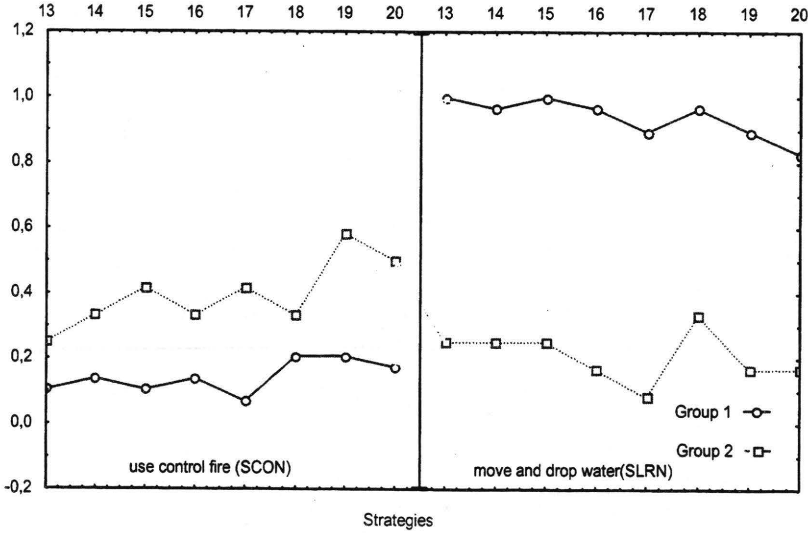
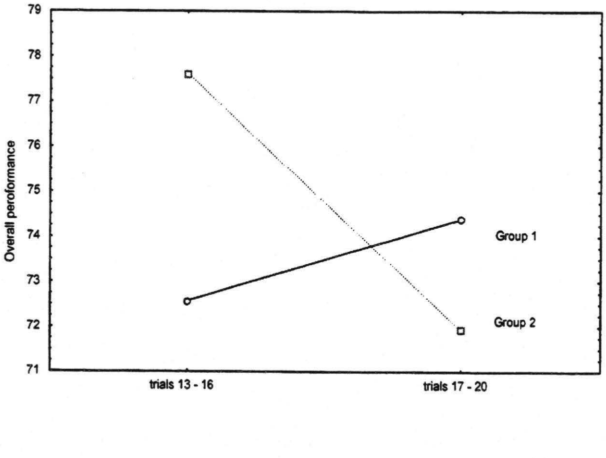


Figure 5: Group by before - after wind change interaction



Results (see Figure 5) showed that group 2 which started with a performance much better than group 1, was affected by the change. They performed worse than group 1 in the second level of the variable Before-After, as showed the significant effect of the interaction between Groups and Before-After Change,, $F(1,22)= 5.5156$, $Mse = 121.4$, $p < 0.05$ (observed power, .612). The interaction between Before-After Change and Trials was also significant $F(3, 66) = 2.96$, $Mse = 57.8$, $p <0.05$, but this was not relevant to our objectives.

This results indicated that, as in experiment 1, the group who used control fire commands was the best before the change and the worst after it. Thus, participants who employed a strategy similar to SCON had a tendency to be affected by the change in the wind direction. Participants in Group 1, whose performance was under group 2 before the change, kept a constant improvement in spite of the change, and finished the experiment with a performance better than group 1.

DISCUSSION

The results from both experiments confirmed our hypothesis. After some time performing a task a person acquired knowledge about the environmental conditions and developed problem solving strategies appropriate to those conditions. When the environment changed some people were affected by that change and showed a decrease in performance. Those people used the SCON strategy and were performing better than the others before the change.

In order to explain these results we need to consider what each of these strategies required from participants' cognitive processing.

The SRLN strategy simply consists on moving appliances to the closest fire and dropping water there. It does not require making predictions on where a new fire will start. Therefore, the direction of the wind did not matter much for reaching a good performance. Something similar happens with the MISC strategy. A person using it defines an area around one already started fire and sends an appliance there to search for it. Therefore, it does not require any prediction at all. However, a person using the SCON strategy needs to predict in which direction the wind will be blowing and issue a control fire according to that prediction. The fire will spread depending on the wind direction. Therefore, the person selects the location to issue a control fires depending on that prediction.

Therefore, it seems to be an easy explanation for the relation between the environmental change and the problem solving strategies. However, it remains to be explained why participants that used the SCON strategy performed better before the change.

The strategies simulated in these two experiments were very simple and need to be elaborated to explore this hypothesis within a general model of decision making in Complex Problem Solving tasks. Therefore, in order to find a complete explanation for our results, our next step in this research program will be to make a more complete model of how people make predictions and design possible strategies based on that model.

REFERENCES

Hollnagel (1998). Cognitive reliability and error analysis method. Elsevier . Tokyo, Japan.

Howie and Vicente (1998). Measures of operator performance in complex, dynamic, microwolds: advancing the state of the art. Ergonomics, 81, 485-500.

Norman, D. A. and Shallice, T. (1980). Attention to action: Willed and automatic Control of behavior,. In R .J. Davidson, G. E. Schwartz and D. Shapiro, Consciousness and self regulation, pg. 1 - 15, New York. Plenum times.

Norman, D. (1981). Categorization of action slips, Psychological review, volume 88, pg. 1 – 15.

Omodei , M.M. and Wearing, A.J. (1995). The FireChief microworld generating program: An ilustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. Behavior Research Methods, Instruments and Computers, 27, 303-316.

Rasmussen, J. (1983) Skills, Rules and Knowledge: Signals, Signs, and symbols and other distinctions IEEE transactions on human, systems and cybernetics volume 3,pg. 653 – 688.

Reason, J. (1990) Human error. Cambridge University Press, New York.

SUBJECT INDEX

Failure	Difficulty and safety during the management of a severe incident management sequence
Fire chief	An explanation of human errors based on environmental changes and problem solving strategies
Free flight	Traffic maneuverability and cockpit display characteristics determine whether commercial airline pilots can maintain self separation in realistic scenarios of en-route flight
Graphical notations	Using cognitive dimensions to analyse graphical notations
Hand-held devices	Interacting with a personal wearable device
Human error	An explanation of human errors based on environmental changes and problem solving strategies
Human factor	Difficulty and safety during the management of a severe incident management sequence
Human reliability analysis	A three-stage information analysis to support predictive human reliability analysis
Human-computer interaction	Exploring the Metaphor of "Automation as a Team Player" : taking team playing seriously
In-car information	Structured Method for Research and Design of an In-Car Information System
Incident	Difficulty and safety during the management of a severe incident management sequence
Intelligent systems	Interacting with a personal wearable device
Job design	Coping with uncertainty in temporal planning and scheduling
Local-area network design	A method for analysing collective design processes
Management process	Difficulty and safety during the management of a severe incident management sequence
Mental load	Studying cognitive effort and processes involved in computer graphics
Mental representations	Degree of automation and its influence on the development of mental representations
Microworlds	An explanation of human errors based on environmental changes and problem solving strategies
Mobile computing	Interacting with a personal wearable device
Performance influencing factors	A three-stage information analysis to support predictive human reliability analysis
Pilot performance	Traffic maneuverability and cockpit display characteristics determine whether commercial airline pilots can maintain self separation in realistic scenarios of en-route flight
Planning	Coping with uncertainty in temporal planning and scheduling
Planning and control	Structured Method for Research and Design of an In-Car Information System
prototyping	Interacting with a personal wearable device
Reliable HCI	Structured Method for Research and Design of an In-Car Information System
Representations	Integrated representations for task modeling
Safety	Difficulty and safety during the management of a severe incident management sequence
Scheduling	Coping with uncertainty in temporal planning and scheduling

Self separation	Traffic maneuverability and cockpit display characteristics determine whether commercial airline pilots can maintain self separation in realistic scenarios of en-route flight
Sequencing	Time-to-collision and action sequencing on aircraft conflicts in air traffic control
Situation awareness	Time-to-collision and action sequencing on aircraft conflicts in air traffic control
Sociotechnical system design	Coping with uncertainty in temporal planning and scheduling
Software design	A method for analysing collective design processes
Software review	A method for analysing collective design processes
Speech and voice	Interacting with a personal wearable device
Structured methods	Structured Method for Research and Design of an In-Car Information System
Task analysis	A three-stage information analysis to support predictive human reliability analysis Integrated representations for task modeling
Task modelling	Integrated representations for task modeling
Teamplay	Exploring the Metaphor of "Automation as a Team Player" : taking team playing seriously
Technical review meeting (TRM)	A method for analysing collective design processes
Time	Difficulty and safety during the management of a severe incident management sequence
Time to collision	Time-to-collision and action sequencing on aircraft conflicts in air traffic control
Transitions between actions	An explanation of human errors based on environmental changes and problem solving strategies
UML	Integrated representations for task modeling
Uncertainty	Coping with uncertainty in temporal planning and scheduling
Usability	Using cognitive dimensions to analyse graphical notations
Usability evaluation	Interacting with a personal wearable device
Verbal interactions	A method for analysing collective design processes
Workload	Studying cognitive effort and processes involved in computer graphics

HMI
GRADUATE SCHOOL FOR HUMAN-MACHINE INTERACTION

HFA
SWEDISH CENTRE FOR HUMAN FACTORS IN AVIATION

ISBN 91-7219-828-1