EACE



Eighth European Conference
on
Cognitive Ergonomics

Cognition and the Worksystem

Granada, Spain, 10-13 September 1996

Proceedings

Edited by:
T. R. G. Green, J. J. Cañas, and C. P. Warren

Technical Chair:
E. Hollnagel
Conference Co-Chairs:
C. P. Warren
F. Détienne
J. J. Cañas
N. Bonnardel

EACE
European Association of Cognitive Ergonomics

ECCE – 8

Proceedings of the Eighth European Conference on Cognitive Ergonomics

Granada, Spain, 10 – 13 September 1996

Editors
Thomas Green
José Cañas
Clive Warren

I.S.B.N. 2-9510412-0-9
Depósito Legal GR - 814-96
Impreso en: Copistería La Gioconda S. L.

EUROPEAN ASOCIATION OF COGNITIVE ERGONOMICS
EACE
UNITE DE RECHERCHE INRIA ROQUENCOURT
78153 ROQUENCOURT, LE CHESNAY CEDEX
FRANCE

ORGANISATION OF ECCE-8

GENERAL CONFERENCE CO-CHAIRS:
José Cañas, Françoise Détienne, Clive Warren

TECHNICAL PROGRAMME CHAIR:
Erik Hollnagel

POSTERS / PANELS CHAIR:
Nathalie Bonnardel

LOCAL ORGANISING COMMITTEE:
Maria Bajo, José Cañas, Thomas Green

REFEREES
Salvador Algarabe
Sebastiano Bagnara
David Benyon
Nathalie Bonnardel
Guy Boy
Pedro Cacciabue
José Cañas
Françoise Darses
Françoise Détienne
Michael Frese
Thomas Green
Jean-Michel Hoc
Erik Hollnagel
Alfred Kobsa
Leena Norros
Reinhard Opperman
Janine Rogalski
Gerrit van der Veer
Willemien Visser
Yvonne Waern
Clive Warren

PREFACE

The goal of the European Association of Cognitive Ergonomics (EACE) is to bring together European researchers working in the domain of cognitive ergonomics, with the aim of developing and enhancing the relationship between cognitive sciences and the techniques of presenting and processing information.

Needless to say, a substantial proportion of research deals with the impact of information technology, whether in work situations such as office automation and process control or in everyday life such as presenting paintings via the Internet. Yet there are wider vistas than that of 'human-computer interaction' (HCI); the problems of interaction between people, across task roles and specialisms, have received more treatment lately than ever before.

The present volume contains the proceedings of ECCE-8, the Eighth European Conference on Cognitive Ergonomics, hosted by the University of Granada, Spain from 10-13 September 1996. The main theme of ECCE-8 was "Cognition and the Worksystem". Contributions that explored the cognitive or social aspects of cognitive ergonomics and human-computer interaction were invited.

The following pages contain the text of 34 papers (or in a few cases extended abstracts, due to local difficulties) and 12 poster presentations. Their order follows the order of presentation at the conference, dealing in turn with system development; the problems of co-operative work; user interface design; modelling; problem solving; aviation research; and learning.

The ECCE conferences are organised so as to enable optimal discussions between leading researchers from different disciplines, in comfortable surroundings. Regrettably there will be no archival record of the discussions.

For the first time, it has been feasible to ask authors to submit papers in electronic form so that they could be presented in a common format. This has also made it possible for non-native speakers to have their English usage checked by a native speaker. The editor in question (TG) has tried to pick up significant departures from conventional usage while not getting too fussy; he apologises in advance not only for anything he missed, but also for any changes that have distorted the meaning of the original. He also apologises for any mangling (especially of all those beautiful diagrams) that may have occurred in the transition across platforms and word-processors.

The editors would like to thank Iain Morrison for prompt and effective software support, Ruth Kearns for help with technical terms and with French words, Alan Blackwell for proofreading, and all the authors whom we forced to answer questions, redraw diagrams, reword their prose, or submit to having their papers changed without consultation and who without exception bore it patiently.

Cambridge, 28 June 1996

Thomas Green

José Cañas

Clive Warren

ECCE 8
1996

Table of Contents

Part 1: System Development

1 Strategies for knowledge-based support in graphical user-interface design: A Wizard-of-Oz study
M. Ericsson, M. Baurén, J. Löwgren and Y. Wærn

5 Distributed information resources: a new approach to interaction modelling
P. C. Wright, B. Fields and M. D. Harrison

11 Usability and database search at the Swedish Employment Service
C. M. Allwood and S. Thomée

17 Utilizing a concept model in user interface design
A. A. Clarke, B. Soufi, J. R. Tyrer & L. H. Vassie

23 The diversity of user participation in system development: a video analysis
C. Axtell, P. Waterson and C. Clegg

Part 2: Cooperative Work

29 Coordination training for distributed worksystems in emergency management
J. Dowell and W. Smith

35 Reshaping task contents: operators' contribution to systems reliability
E. Morvan, L. Rognin and J.-C. Sperandio

41 Creation and loss of cognitive empathy at an emergency control centre
H. Artman and Y. Wærn

47 Construction of shared representation through cooperative dialogues of shift changeover
C. Grusenmeyer

53 Computerised ward monitoring systems: understanding the stakeholders
M.A.R. Kirby, R. Kane, K.D.C. Stoodley and J. Bates

57 A preliminary model of the planning and control of the combined response to disaster
B. Hill & John Long

63 Artefact usage & conversation patterns at meetings
B.P. Bright & N. A. M. Maiden

Part 3: User Interface Design

69 Interruptions in computerised mental work: user interface design issues
I. Burmistrov and A. Leonova

71 Individual differences and navigation in hypermedia
K. Höök, N. Dahlbäck and M. Sjölinder

77 Pictorial aids in computer use
R. Navarro, J. J. Cañas and M. T. Bajo

83 Ergonomics in the wild: the practitioner's perspective
Á. Cabrera, M. I Rincón and J. González

89 Studying evaluative knowledge used by designers: towards improving critiquing systems
N. Bonnardel & L. Joblet

95 Cognitive function analysis
G. A. Boy

Part 4: Modelling

97 SimulACRUM: a cognitive architecture for modelling complex task performance and error
S. Grant

103 Cognition in the workplace: integrating experientialism into activity theory
M. Imaz and D. Benyon

109 Goal-oriented categorisation: the role of task, daily activity, expertise and problem attributes
W. Visser & L. Perron

115 Modeling complex work systems - method meets reality
G. C. van der Veer, M. Hoeve and B. F. Lenting

121 Quality models in usability evaluation of business management software
M. Sikorski

127 Cognitive and organizational dimensions of task appropriateness
C. Stary and A. Totter

Part 5: Problem Solving

133 Practical insights from studies of operator diagnosis
E. Hollnagel and B. Kirwan

139 Reusing processes and documenting processes: toward an integrated framework
F. Détienne, J.-F. Rouet, J.-M. Burkhardt and C. Deleuze-Dordron

145 The power of habit in the failure to diagnose unusual faults
J. Patrick

147 Adapting models to cases: a study of collective problem solving
C. Sauvagnac & P. Falzon

Part 6: Aviation Research

149 The issue of ground/cockpit integration: results from field studies
A.-L. Amat and A. Bellorini

155 Multi-level human-machine cooperation in air traffic control : an experimental evaluation
M.-P. Lemoine & J.-M. Hoc

161 Cognitive aspects of ATC: experience from the CAER and PHARE simulations
C. J. Kelly and P. J. Goillau

Part 7: Learning

163 Supporting Continuous Learning
R. Oppermann

167 What's the story? Narrative and the comprehension of educational interactive media
L. Plowman

173 Helping to acquire the perceptuomotor skills underlying the learning of prosody in a foreign language
B. Denoeud, M. Raquillet and A. Foti

Poster abstracts

179 The application of a combined strategy for the interaction between a memory impaired person and a wordprocessor
J. A Bajer

181 Evaluation of iconic representations of patient records in a medical workstation
L. Pellegrin, V. Méjean, N. Bonnardel and H. Chaudet

183 Applying a user-centered design approach to an information kiosk for Spanish social security
J. Merino and M. Quesada

185 Informal communication is about sharing objects and media
J. Ramsay, A. Barabesi and J. Preece

187 Effects of noise on driving behaviour
M.-T. Bajo, J.-J. Cañas, F. Padilla and M. del C. Puerta

189 Extraction of interface quality measurements from dialogue description notations
P. M. Valero-Mora

191 Painters' attitudes towards internet art presentations
A. E. Voiskounsky

193 Human factors engineering using the PUMA (Performance and Usability Modelling) tool
C. P. Warren, P. Day, and M. Hook

195 Modelling of operator team behavior in nuclear power plants
S. Yoshimura and N. Hasegawa

197 index of keywords

199 author index

Strategies for Knowledge-based Support in Graphical User-Interface Design: A Wizard-of-Oz Study

Mikael Ericsson[1], Magnus Baurén[2], Jonas Löwgren[1] and Yvonne Wærn[2]

[1]. Department of Computer and Information Science
Linköping University, 581 83 Linköping, Sweden
Tel: +46 13 284021 Fax: +46 13 142231
Email: {miker,jlo}@ida.liu.se

[2]. Department of Communication Studies
Linköping University, 581 83 Linköping
Email: {magba,yvowa}@tema.liu.se

ABSTRACT

The use of knowledge-based tools, capable of generating comments on an evolving design, is seen as a promising approach to providing user-interface designers with some of the knowledge they need in their work. However, there is a lack of empirical explorations of the idea. This paper reports a Wizard-of-Oz study, in which the designers' perceptions of the usefulness of a commenting tool were investigated along with the appropriateness of different tool behaviors and forms of comment. The results show that a commenting tool is seen as disturbing but useful, that the comments from an active tool risk being overlooked, and that comments pointing out ways of overcoming identified design problems are the easiest to understand. Implications for the design of commenting support tools are identified.

Keywords

user interface design support, commenting tool, critiquing, user interface guidelines, Wizard-of-Oz, perceived usefulness, mental workload.

INTRODUCTION

User-interface design is an activity with high knowledge requirements, as evidenced by scientific studies, professional practice, and the amounts of textbooks and courses on the subject. A concrete example of the professional need for design knowledge is the increasing tendency of customers in industrial systems development to require style-guide compliance.

There is a rather rich literature on knowledge-based user-interface design support (for a survey, see Nakakoji et al. (1996)). To illustrate the different approaches taken to tool behavior, consider a few concrete examples all involving the use of lexical and syntactic design knowledge. Sierra (Vanderdonckt, 1995) is based on lexical and syntactic design guidelines and allows the designer to browse and search for relevant information. KRI/AG (Löwgren and Nordqvist, 1992) uses the same kinds of knowledge to evaluate a user-interface design upon demand, pointing out deviations from the rules stored in the knowledge base. Framer (Lemke and Fischer, 1990) is more active in that it autonomously generates comments. Don (Kim and Foley, 1993)) uses fairly sophisticated lexical and syntactic rules to automate the design of dialog boxes, based on a model of the application's data and control structures. Finally, Mimir (Henninger et al., 1995) illustrates an attempt to build an organizational memory in which lexical and syntactical user-interface design knowledge (among other aspects) is accumulated and enhanced over time.

This paper is aimed at exploring in more detail the notion of an evaluative support tool that provides the designer with user-interface design knowledge in the context of the design task at hand. The technology we address is known in the AI community as "critiquing", where it is typically defined as a knowledge-based system that comments upon a proposed solution to a problem rather than solving it autonomously. The general case in favor of using this kind of commenting tools as knowledge support in a design environment has been addressed elsewhere (e.g., Fischer et al., 1991, 1993; Löwgren and Laurén, 1993; Silverman and Mezher, 1992); our aim here is to concentrate on more detailed questions pertaining to the use of commenting tools to support user-interface design. Specifically, the questions we ask are the following.

Are comments from a support tool useful in a user-interface design setting?

The term "useful" needs further qualification. One interpretation is, of course, that the tool enhances the quality of the finished design or of the design process. We refrain from this interpretation due to the detailed level of investigation; rather, we concentrate on the usefulness as perceived by the designer. The importance of this aspect of software design support tools has been addressed by, e.g. Carey (1988) and Morris (1987), As we shall see below, our study contains several ways to approach perceived usefulness.

What is the most appropriate form of the comments?

We distinguish between declarative comments, that merely point out the existence of a possible design flaw, and imperative comments that also suggest to the designer how the flaw can be remedied. This dimension will be referred to as the mood of the comments, using a concept borrowed from linguistics.

What is the most appropriate behavior of the commenting tool?

Here, we concentrate on the distinction between passive and active tools. Using a passive tool, the designer has to ask explicitly for comments. In contrast, an active tool autonomously presents a comment as soon as it detects a "mistake" in the designer's work. In the case

of the active tool, we further want to explore at what point the comment should be given in order not to unnecessarily disturb the designer.

In addition to these main questions, we also explore issues related to the commenting timing strategy, i.e., the mehanisms that decide when the system should deliver a comment if an active tool is preferred. Hence, we are interested in investigating the user's (designer's) perceived mental workload (MWL) when using a commenting tool.

The remainder of the paper presents an empirical study undertaken to answer these questions. As the overall goal of the work is to inform the design of appropriate support tools, the paper closes with a number of tentative design recommendations based on the results we present.

THE STUDY

Background

To gain an initial understanding of the phenomena of interest, we performed an exploratory pilot study. In this, professional designers, working in pairs, were videotaped as they collaborated to solve a design task. From the videos, we identified commenting behaviour in the communication between the designers. For example, we found that the commenting designer is primarily active, illustrates comments with examples and often begins by describing "how to do it instead", rather than telling his colleague that "it's wrong". These findings guided our choice of mood and behavior as dimensions for detailed study.

Since we had no possibility to actually design an interface design support, including all the intended variations, we used a Wizard-of-Oz environment for our empirical study (see figure 1). Reasons for using such an environment have been voiced for instance by Dahlbäck, Jönsson, & Ahrenberg (1993). A Wizard-of-Oz environment for studying graphical design support systems was developed and is described in another report (Ericsson, in preparation).

Experiment Design

In the experiment, the participants used a User Interface Builder (UIB) to design the static user interface for a hotel search system. The participants were divided into five groups. One group (control) used only the UIB, and the other four also had a simulated support tool capable of evaluating and commenting upon the design. For the experimental groups, a crossed design was used, where one variation consisted in active/passive mode and the other in declarative/imperative mood.

Task and instructions

The participants were given a description of a design task. In the description, a hotel search system was outlined, and the participant was instructed to work on the visual design part of the systems development.

Advice given

In order to enable the Wizard to work as quickly as possibly, a limited number (33) of guidelines were written. Two versions were used – a declarative and an imperative – with the intent to keep their semantic content similar.

Participants

A total of 16 interface designers participated in the experiment. All of them currently studied at the university. Fifteen were male and one was female; 11 of the participants came from Computer Science and 5 from other study programmes. They each received a movie ticket in return for their participation.

PROCEDURE

The participants were asked to design the user interface, and were (in the active condition) given advice as soon as the Wizard noticed some flaw in their design. In the passive condition, the Wizard provided advice when the participants demanded such. The control group did not receive any comments. During the experiment, all interaction was recordedin a file, and the participants screen was recorded on video.Ê

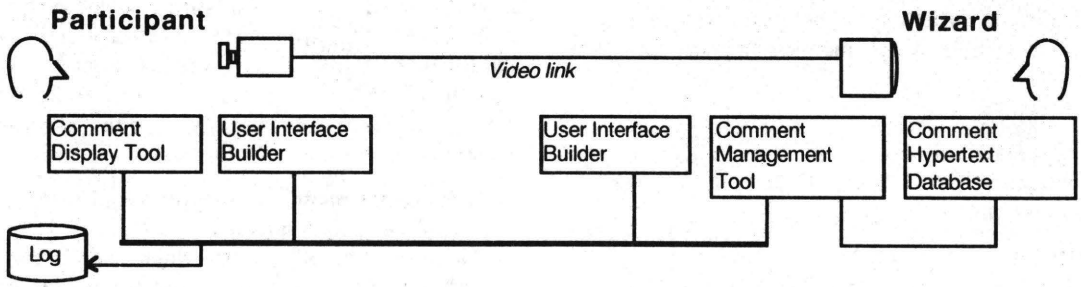


Figure 1 Overview of the CODEK Wizard-of-Oz environment.

After the session, the participants were asked to rate the system (rating scales given below) and to describe their impression of the total system. The participants were then shown three short video clips from their work on the design. For the non-control group participants, one or two of the clips showed situations where the participants received comments, starting one minute before the comment was given. They were asked to rate the particular situation pertaining to this clip in two different ways. First, the participant was given an RTLX-form with which perceived MWL was measured. Secondly, they answered a questionnair with rating scales similar to those used for the total system.

Rating Scales
Likert scales were used for collecting participant ratings of the complete design environment, the commenting support tool and specific comments. The rating questionnaires used included questions about usefulness, understandability, system competence, disturbance, perceived stress and ability to find referred interface objects. The scales ranged from 1 to 8.

Perceived mental workload was measured using RTLX (Byers et al., 1989), which is a simplified version of the subjective task load measurement technique NASA TLX (Task Load indeX). In RTLX, participants rate the workload contribution for a set of factors (performance, effort, temporal and mental demand and frustration); from this an overall task load level (0-100) is calculated.

RESULTS

The control group rated the system as very little disturbing (Mean 1.7) and the other groups' answers indicate some disturbance (Mean 3.7). At the same time, the participants getting comments rated how much they would like to have a system like this in their work environment very high (Mean 7.1) This indicates that they despite the disturbance found the system to be useful.

The linguistic variation in the comments showed that the comments with imperative form had higher understandability than the comments with declarative form (Mean 6.3 vs. 2.8, $p<0.05$), and resulted in lower rated system competence (Mean 4.7 vs. 5.8, $p<0.1$). The mode variation showed that the system was experienced to be more supportive in the active condition than in the passive one (6.6 vs. 5.0, $p<0.05$). Furthermore, it was considered easier to find objects referred to in the comments in the active mode, than in the passive one (6.9 vs. 3.9, $p<0.01$).

When scrutinizing the results from the video-clips it was found that the participants in the active condition sometimes did not notice the comments at all. This mostly happened when mental workload was rated as high (RTLX>50).

DISCUSSION

The aim of this study was to compare some different ways of administering an advice giving support in an interface design situation. The planning and results from the study indicated that general design rules for a design support have to be supplemented with considerations, situated in the particular design moment.

First, the commenting tool was perceived as useful in the support of user-interface design. However, it was also perceived as disturbing. This apparent contradiction does not in fact need to be problematic if we also consider the finding that the comments provided the designers with previously unknown information. A plausible interpretation is that the comments interrupt, and thus disturb, the flow of design but that the value of the commentsÑin terms of their contentsÑjustify the interruption.

It may be proposed that the comment must be "situated" in the particular design moment. The importance of such a "situatedness" is supported by the result concerning the linguistic variation. A comment phrased in a specific form (imperative). may be easier to relate to the particular design moment than a comment phrased in a general form (declarative) .

At the same time, it is interesting to note that providing an easy to understand comment lead to a perception of a less competent system. than the providing of declarative ones. This result is similar to others, showing that systems which are easier to understand are considered to be less competent (cf. Wærn, 1995).

The seemingly straight-forward difference between an active and a passive advice was found to be difficult to administer. In the passive condition, the Wizard was simply too slow. However, despite this drawback, the passive condition was considered to be as helpful as the active advice. This indicates that designers can tolerate a long delay, if they can control the timing of it themselves.

Finally, it was asked if it was possible to find at what time active comments should be provided. We believed that the comments should not be given while the designers were at most busy with their design. The results show that designers did not notice some comments during high perceived MWL. In the cases where a participant noticed a comment despite a higher MWL, the comment and the visual indication concerned an object situated close to the area or object they were currently working on. Those comments were also rated less useful and more disturbing than the others. These results suggest that people do not notice a comment when they are intensely busy with their design activities, unless it falls within the field of visual activity.

The identified relation between perceived MWL and ability to notice comments may be interpreted in at least three ways, each one suggesting a different technical solution.

Timing is unimportant
The exact timing of delivering a comment is not important. When the user's MWL is low enough, she will see it. Comments could be delivered at any time, which means that no special temporal mechanism is needed.

A temporal strategy is needed
It is useless to present a comment when the user's MWL is too high. The comment might even increase the MWL, causing the designer to ignore the comment, or treat it as another disturbing element. A study by Carroll and Aaronson (1988) showed that users were not able to utilize the help (from a simulated active help-system) if it did not appear at the appropriate time. In our system, this means that the commenting tool would need a temporal strategyÑan algorithm delaying comments until the user's MWL has lowered. This requires that the system could somehow estimate the user's MWL.

Better cues are needed
There are comments that really should be read as soon as possible by the user, i.e., the attention must be drawn to the comment immediately to ensure that the violation is detected before it is too late (c.f. Lemke and

Fischer, 1990). More powerful visual cues and presentation techniques are needed in the tool, e.g. sound signals and colouring of violating objects.

Further studies are underway to investigate if it is possible to predict, from the designers' momentaneous work, how disturbing a comment would be. We will also extend our work by investigating the effects of using commenting tools, and other design support approaches, in a professional environment.

ACKNOWLEDGMENTS

We wish to thank the participants of our experiment for their time and efforts. This work was financially supported by the Swedish Research Council for Social Sciences and the Humanities (HSFR).

REFERENCES

Byers, J. C., Bittner, A. C. Jr., and Hill, G. S. (1989) Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In Mital, A., editor, *Advances in Industrial Ergonomics and Safety*. Taylor & Francis..

Carroll, J. M. and Aaronson, A. P. (1988) Learning by doing with simulated intelligent help. *Communications of the ACM*, 31(9):1064-1079.

Carey, T. (1988). The gift of good design tools. In Hartson, H. R. and Hix, D., editors, *Advances in Human-Computer Interaction*, chapter 5. 1988.

Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993) Wizard-of-Oz-studiesÑWhy and how. In Workshop on Intelligent User Interfaces, Orlando, FL.

Ericsson, M. (In preparation). Commenting systems as User Interface Design Support Tools: A Wizard-of-Oz study. Licentiate thesis. Linköping University.

Fischer, G., Lemke, A. C., Mastaglio, T., and Morch, A. I. (1991) The role of critiquing in cooperative problem solving. *ACM Transactions in Information Systems*, 9(3):123-151, April.

Fischer, G., Nakakoji, K., Ostwald, J., Stahl, G., and Sumner, T. (1993) Embedding critics in design environments. In *Proceedings of INTERCHI'93*, pages 157-164. New York: ACM Press.

Henninger, S., Haynes, K., and Reith, M. (1995) A framework for developing experience-based usability guidelines. In *Proceedings of DIS'95*, pages 43-53. New York: ACM Press.

Kim, W. C. and Foley, J. D. (1993) Providing high-level control and expert assistance in the user interface presentation design. In *Proceedings of INTERCHI'93*, pages 430-437. New York: ACM Press.

Lemke, A. and Fischer, G. (1990) A cooperative problem solving system for user interface design. In *Proceedings of AAAI-90*, pages 479-484. AAAI Press/The MIT Press.

Löwgren, J. and Laurén, U. (1993) Supporting the use of guidelines and style guides in professional user-interface design. *Interacting with Computers*, 5(4):385-396.

Löwgren, J. and Nordqvist, T. (1992) Knowledge-based evaluation as design support for graphical user interfaces. In *Proceedings of CHI'92*, pages 181-188. New York: ACM Press.

Morris, N. M. (1987). Designing for user acceptance of design aids. In Rouse, W. B. and Boff, K. R., editors, *System Design: Behavioural Perspectives on Designers, Tools, and Organizations*, pages 245-255.

Nakakoji, K., Malinowski, U., and J., L. (1996). Knowledge-based support for the user-interface design process: A CHI95 workshop. *SIGCHI Bulletin*. 28(1):43-47.

Silverman, B. and Mezher, T. (1992) Expert critics in engineering design: Lessons learned and research needs. *AI Magazine*, (Spring):45-62.

Vanderdonckt, J. (1995) Accessing guidelines information with Sierra. In Nordbyn, K., Helmersen, P., Gilmore, D., and Arnesen, S., editors, *Proceedings of Interact'95*, pages 311-316. London: Chapman & Hall.

Wærn, Y. & and Ramberg, R. (1995) People's perception of human and computer advice. *Computers in Human Behavior*.

Distributed Information Resources: A New Approach to Interaction Modelling

Peter C. Wright, Bob Fields, Michael D. Harrison
Dept. of Computer Science
University of York
Heslington, York, Y01 5DD, UK
+44 1904 432741
pcw@cs.york.ac.uk bob@cs.york.ac.uk mdh@cs.york.ac.uk
http://www.cs.york.ac.uk/~pcw http://www.cs.york.ac.uk/~bob http://www.cs.york.ac.uk/~mdh

ABSTRACT
In this paper, we present a new approach to interaction modelling based on the concept of *information resource*. The approach is inspired by distributed cognition literature but develops a model which applies specifically to human-computer interaction (H-CI) modelling. A number of information resources are described. These resources can be represented internally or externally and can find different expressions when represented externally. The distribution of resources and the properties of their expressions are what shape interaction. The resource model is exemplified by analysing an episode from a Macintosh interaction.

Keywords
H-CI modelling, distributed cognition, design, evaluation

PLAN-BASED THEORIES IN H-CI
Within both the science and engineering of H-CI, most models of interaction are plan-based. Norman's (1988) well known goal-plan-execute-evaluate cycle is perhaps the most widely cited example and has been used to consider design issues like the so-called gulf of execution and the gulf of evaluation. Similarly, Reason and others (Reason 1990) provide a rich taxonomy of human error in the use of complex technology which has a plan-based model of interaction as a central component. They distinguish, for example, between planning errors and errors in the execution of a plan. Within the engineering disciplines associated with the design of interactive systems, plan-based models manifest themselves in approaches to task analysis and task description (see Diaper 1989 for a review). Hierarchical task analysis for example, provides a means by which designers can represent users' tasks as a first stage in the process of task-centred design.

In her study of the use of an intelligent help facility Suchman (1987) argued against plan-based models of human-computer interaction. She concluded that in many environments, users do not create elaborate plans specifying exactly how they will interact with a machine to achieve a particular goal. Rather, their choice of action is determined by the preceding responses of the computer. Interaction is viewed as series of locally managed turns much more akin to spoken conversation. A related argument has been raised by studies of graphical user interfaces. Kitajima and Polson (1995) have produced a model in which action is governed by semantic matching between the user's goals and the meanings of individual menu items. In this model there is no concept of plan or a representation of an intended action sequence. Instead, choice of action is locally managed. The key point that is being made in both these pieces of research is that interaction can occur without the need for the user to have a precise knowledge of what they intend to do. Instead, the user responds to environmental cues such as menu items, on the basis of their relevance to an overall goal which is often only partially defined.

The need for models of interaction to take account of the role of environmental cues in shaping interaction is well made by these and other researchers (e.g., Howes and Payne 1990). But the research has not attempted to postulate any general conceptual framework for understanding the role of environmental cues, that could be of used for example, in the design of a broad range of interactive systems. This paper aims to provide the beginnings of such a conceptual framework. To do this we turn to another field of research.

Distributed Cognition
Distributed cognition (Hutchins 1995) is a synthesis of cognitive, anthropological and social science approaches to the study of collaborative work and has been used to analyse a number of group-work situations including aircraft cockpits, air traffic control, and marine navigation.

The central thesis of the distributed cognition paradigm is that intelligent action in everyday contexts cannot be properly studied by an analysis of an individual's cognitive performance in isolation. It is argued that central to the uniqueness of human intelligence is the ability to recognise our cognitive limitations and to design our environment to overcome these limitations. Thus any analysis of cognition has to include a theory of how individuals use the resources in their environment to extend their cognitive performance. Such resources include other individuals and physical artefacts such as computers, calculators, reference manuals and so on. Within this theory, cognition is not seen as an individual and internal process but as process distributed between individuals and artefacts. The approach dissolves the traditional boundary between the human agent and their environment. Zhang and Norman (1994) capture the importance of this point well,

"A wide variety of cognitive tasks, whether in everyday cognition, scientific practice, or professional life, require

the processing of information distributed across the internal mind and the external environment. It is this interwoven processing of internal and external information that generates much of a person's intelligent behaviour ..... The traditional approach to cognition, however, often assumes that representations are exclusively in the mind ..." p. 87-88.

The work of Hutchins and his co-workers concentrates on resources for action in social situations. Palen (1993), for example, studied collaborative work on an aircraft flight deck. Relatively little attention has been paid to the kinds of resources that might be relevant for analysing human-computer interaction where, for example, a single user interacts with a desk-top application package. But the key idea of distributed cognition – that resources for action can be distributed between individual and artefact – suggests an approach to modelling human-computer interaction. This would encompass both the plan-based and situated approaches at a level of specificity valuable to designers of interactive systems. The model views plans as only one of a number of distributed information resources that can be used to determine next action.

PLANS AS DISTRIBUTED INFORMATION RESOURCES

An example from a Macintosh application

Consider as an example, the use of dialogue wizards in Microsoft's Excel spreadsheet[1] application.

If a user wants to produce a graphical representation of a set of data (referred to as a 'chart' in Excel), he or she can use the *chart wizard*. The user is presented with the first of five dialogue boxes, as in figure 1.

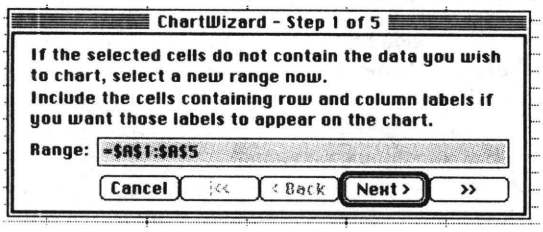


Figure 1.

Each box allows the user to type in some data or make some choices relevant to the goal of charting the data. Once each data entry is complete the user can move to the next dialogue box. As they progress through the dialogue they make choices about what type of chart they require and so on. When they arrive at later dialogue boxes, they are presented with a sketch of what the finished graph will look like. At any point in the dialogue, the user may move forward to the next box or back to the previous one.

From a cursory inspection, it might be argued that such a scenario is a good example of situated interaction. The user need have no knowledge of how to produce a chart using Excel. Buttons and prompts indicate clearly how to negotiate the very next step in the interaction and clear feedback is presented about the success of actions. The whole interaction episode can be locally managed without need for the user to have an explicit plan in mind. However, if we consider the episode more closely from the standpoint of distributed cognition, it can be seen that although there may be no plan represented in the head of the user, a plan does, nevertheless, govern this interaction. But this plan has been externalised and embedded in the interface itself and the dialogue boxes constitute a manifestation or expression of that plan. Each step in the plan is marked by a new dialogue box and the conditional branches in the plan are laid out as a set of choices for the user. The user presses the "Next" button to move forward in the plan. There is information in each dialogue box about how far the user has progressed through the plan (for example the legend "ChartWizard – Step 1 of 5" in figure 1) and as the plan nears completion, there is even a sketch of what the outcome will look like. In addition the user is provided with a way to jump to various points in the plan, by using the "< Back" button and so on.

The chart wizard example illustrates how plans can be usefully viewed as a distributed information structures rather than simply mental constructs. Any interaction may involve different distributions of such information structures. For example, in order to use a system, a user may mentally represent a plan, explicitly recalling each step and executing it. Equally, the user may refer to a manual to find out what steps to take and in what order, to achieve the goal. Alternatively, the plan can be embedded in the interface itself. This embedding make take an explicit form such as a step-by-step checklist, or may be a more implicit representation such as is the case with the chart wizard.

Properties of plan expressions

In order to be used as a resource for action, a plan must be expressed either in the head of the user or in the interface of the computer. The particular way in which plans find expression in interaction impacts usability. Different plan expressions have different usability properties. For example, the same plan for dealing with an engine fire in a modern twin engine commercial aircraft is represented, (i) in the head of the pilot through explicit training, (ii) in a reference manual on the flight deck (iii) in an electronic checklist presented by the electronic crew alerting system. These are three different resource expressions for supporting the same interaction plan. Each expression has different properties which may affect how it can be used. A mental representation of the plan provides rapid access but may be prone to a number of process failures leading to incorrect execution (Reason 1990). The reference manual is portable, but may be slow to access, and is also prone to errors such as forgetting what parts of the plan have so far been executed. An electronic checklist is fast, and not prone to place-losing errors, but it is not portable, can to premature commitment to action, and may place greater attentional demands on the user.

A RESOURCE-BASED MODEL OF INTERACTION

The term resource encompasses a range of meanings. It is often used in the psychological literature to refer to a limitation on the amount of information processing capacity

1. Excel is a trademark of Microsoft Incorporated

available. But in this paper and in the distributed cognition literature, it is used to mean a source of information to achieve a task. For example, an encyclopaedia is a resource for learning.

When we begin to think about interaction from a resources point of view it becomes apparent that plans are only one of a number of information structures that can be used to shape action and which can be distributed between user and artefact. In this section we identify a number of additional resources.

Plans: A plan is a list or sequence of actions that need to be carried out in order to achieve some goal. Usually a plan will involve some marker to indicate which actions have so far been carried out. Most plans involve conditional branches. Which branch is taken is typically dependent on the state of the interaction.

Goals: Goals like plans, are traditionally viewed as mental constructs. How can a goal possibly be externalised in an artefact? An everyday example of an externalised goal is a shopping list. A shopping list represents the state of the world that a person wishes to bring about, namely that all of the listed items are in a shopping trolley. Partial achievement of this goal often occurs when only a subset of the items can be found on the shelves. To consider a more technological example, aircraft flight deck displays include a device called a *flight director*. This indicates to a pilot which way she or he must steer in order to continue on the desired flight path.

Action affordance: The term *affordance* has been used by others in the H-CI literature. Within our resource model we use the term in a more fundamental way, which is to refer to the set of possible next actions that can be taken, given the current state of the system. Menus (both on computers and in restaurants) are an everyday example of one way in which affordances can be expressed because they show the set of possible things that can be chosen.

Interaction history: An interaction history is a list of the actions already taken by the user to get to the current position in the interaction or a list of previous states that the interaction has been in. A UNIX command line interface for example, displays some of the previous commands that the user has issued. Some hypertext systems display miniature pictures of previous nodes visited. Plan-based models such as that of Norman tend not to consider the effect of previous actions on subsequent interaction, yet in situated accounts, history is a key idea and is taken as a primary context for deciding next action. Introducing interaction history and plan as two resources within the same model provides the means of capturing both interaction styles.

Action-effect mapping: An action-effect mapping is a statement of the effect that an action will have if it is carried out. Most user manuals and help systems provide this kind of information.

Interaction state: The interaction state is the collection of relevant attributes of the objects being acted on. In the chart wizard example described earlier, the data being charted is a relevant state attribute, the currently displayed partial graph is another, the displayed dialogue box is a third. The status of the RAM cache is not. Carrying out an action can alter state variables. The user may carry out an action in order to bring about a particular state that is judged to be nearer to the goal.

Processing in the resource model

Within the resources model, an interaction sequence is described as a number of steps from one resource configuration to the next. By *resource configuration* we mean the collection of information structures that find expression at a given step in the interaction and which can be used as resources for action. Each step will be characterised by the execution of two processes.

The first is a decision making process which, given the current resource configuration chooses the appropriate next action. The precise nature of this process depends on the resources available. For example, if the only resource available is an action affordance, then a random choice procedure might be used.

The second process is the re-configuration or updating of resources resulting from action being carried out. For example, in plan-based action, for each step carried out, the plan resource needs to be updated to reflect the new position in the plan.

USING THE MODEL TO ANALYSE INTERACTION

The additional resources and the process model articulated in the above section can be demonstrated by returning to a more detailed analysis of the chart wizard example.

In this section we walk through a scenario with an imaginary user who is familiar with Macintosh interfaces, has some experience with earlier version of Excel but has not previously used chart wizards.

Step 1

The interaction episode commences with an initial resource configuration outlined below, part of which is embedded in the interface, as shown in figure 2.

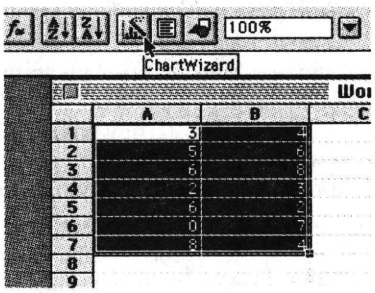


Figure 2.

We attribute to our imaginary user the top-level goal of displaying the selected data as a chart in an area adjacent to the data. This is a resource that is represented in the head of the user. It is not represented in the interface at all. A number of action affordances are presented to the user, most particularly the buttons at the top of the screen. When the cursor is moved over the relevant button, the user is presented with a pop-up legend "ChartWizard". This is an

external representation of an action-effect mapping that maps directly onto the user's goal. The user's (internally represented) goal, together with the action affordances and the action-effect mapping, both represented in the interface, determine the user's next action without the need for a plan.

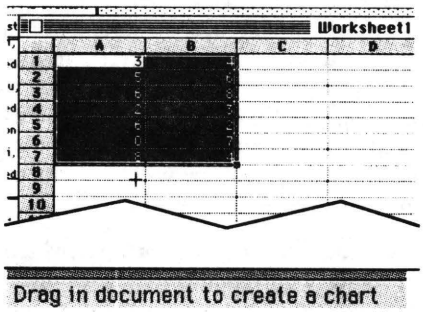


Figure 3.

Step 2
After clicking, the information resources and their presentation in the interface are updated. The screen changes to appear as in figure 3. The user now has an internal representation of the same top-level goal and an interaction history (e.g. some data has been selected for charting, the wizard icon has been clicked on, and the data to be charted is highlighted). As can be seen from figure 3, a cursory glance at the display yields few clues as to what to do next. The cursor has changed shape, the outline of selected data has changed. Neither of these help the user to decide what to do. But at the bottom of the screen, an action-effect mapping has been expressed "Drag in document to create a chart". This action-effect mapping in the context of the user's goal leads directly to action.

Steps 1 and 2 are not supported by an external plan and their correct execution depends on the user having some resources internally represented. In particular the second step relies on the user seeing and understanding the action-effect information at the bottom of the screen. The user must be able to relate that external resource to their mental representation of the top-level goal.

Step 3
Carrying out the dragging action leads to the screen configuration shown earlier in figure 1. The dialogue box has now appeared, marking the starting point of the plan-based of part of the interaction. The length of the plan and the user's current position in it is clearly indicated. This step triggers a new and possibly unexpected sub-goal for the user, which is selecting appropriate data for charting. The user has an action history internally represented which if consulted will indicate that this action has already been carried out. In addition to the internally represented action history, state information expressed on the screen as a box around the data and a range of data co-ordinates, indicates which data is currently selected. Even so, the user must remember that the information selected is the information he or she wishes to chart. Note also that once the next step is taken, the box around the data disappears and the information about which data is going to be charted is no longer visible. The user must rely entirely on memory.

While the goal resource indicates clearly *what* needs to be done, *how* it should be done is not represented externally. The action-effect resource, previously expressed at the bottom left of the screen, is no longer present. The user must simply know that the data can be selected by dragging in the spread sheet or be entering ranges as text into the box. In this step then, a goal is explicitly triggered by events initiated by the wizard, but there are no resources available for determining which action to take in order to achieve this goal. In addition, note that since the correct data range was already selected, the newly triggered sub-goal has already been achieved.

Step 4
Pressing the "Next" button once the correct data is selected leads to a reconfiguration of the resources. The user now has an internal representation which includes a history of choosing an area in which to display a chart and selecting appropriate data for charting. Neither of these have any corresponding external expression. The externally available resources are expressed as in figure 4. Here a new goal is introduced by the presentation of a "gallery" of the kinds of charts that can be constructed. In conjunction with the user's internalised goal of creating a chart, the new resource helps shape a new sub-goal of choosing a chart type that best displays the selected data.

Interestingly, the images serve not only as a goal resource indicating what charts are possible but also as an action affordance since clicking on the images will move the user towards the chosen goal. As well as being expressed implicitly as a gallery of images, The new sub-goal is also expressed explicitly by the prompt "Select a chart type:". Before this point in the interaction, the user may not have had an explicit type of chart in mind, and may not have even known about some of the possibilities. Hence the action taken is an emergent property of the interaction of external and internal resources.

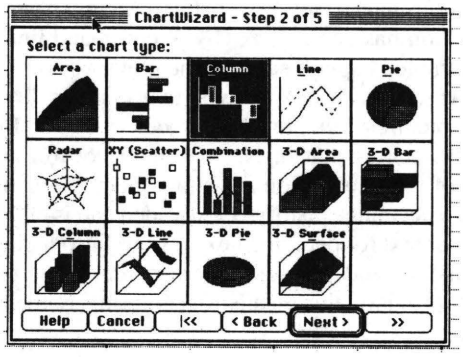


Figure 4.

Step 5 & 6
Selecting an icon and pressing the "Next" button leads to a similar dialogue box where the specific sub-type of chart is chosen. Here again, the new sub-goal and the consequent action is an emergent property, and the sub-goal is

represented implicitly as a gallery of images and explicitly as a prompt to "Select a format for the....chart:".

Clicking "Next" again leads to the penultimate step in the plan. Here the user is presented with a new resource, a representation of what the finished chart will look like as in figure 5. This image of achievement, in conjunction with the user's internal goal of charting his or her data, can be used to evaluate the quality of the finished product even before it is finally produced.

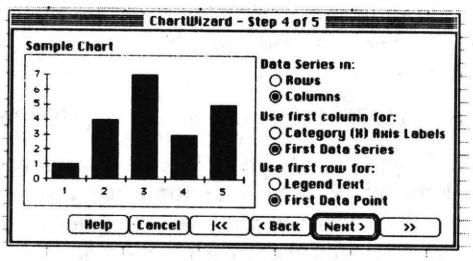


Figure 5.

Results of the analysis

Three key points emerge from the interaction analysis.

i) The first two steps are not supported by a plan resource and rely on the user being able to utilise the action effect resource expressed at the bottom left of the screen in conjunction with some internally held goals. No other external resources are available. The action-effect information which is so central to these first two steps is never used again by the designer.

ii) After the first two steps the action is governed by an externalised plan. The plan that is being followed is not explicitly stated as a "first do this, second do that...." sequence. Rather, it is implicitly expressed as a dialogue sequence. The user is given information abut how long the plan is and where they are in it. But in order to find what the plan is they must explicitly navigate forward through the plan using the appropriate buttons. Similarly in order to determine what they have done so far, they must internalise an action history or explicitly navigate back through the plan using appropriate buttons.

iii) During this plan-based phase, new sub-goals are externalised as galleries of images. These trigger possible choices that the user may have not considered. Hence the action taken by the user is an emergent property of the interaction between internal and external resources. The user's top-level goal of creating a chart is given a rendering in the interface as an image of the final product before it is actually produced. This instantiation of the user's goal serves a resource for early evaluation of the whole interaction.

Our analysis of the dialogue wizard using the resources model allowed us to identify a number of characteristics of the design which enhance usability, and also some potential weakness in the resources provided by the designers of the wizard.

- The early steps are supported less well by resources.
- In the later steps, plans are well supported.
- These plans are not made explicit, with the possible consequence that users may not internalise them over a period of use.
- Implicit resources in general may not support users in developing internal models of the interaction.
- Subgoals and triggers for the formation of goals are well supported in the interface .

The example also showed how both situated and plan-based interaction styles can exist within the same interactional episode. What interactive style emerges as a consequence of the configuration of resources, the form of their expression and their distribution across internal and external boundaries. The absence of any plan resource early in the chart wizard dialogue constrains the user to a situated form of responding. Later, the interaction can be more plan-based because of the presence of a plan resource in the form of a dialogue box sequence.

CONCLUSIONS

Recent work in the study of intelligence has emphasised the importance of what has been termed *metacognition* (Sternberg and Wagner 1994). Traditional cognitive psychology has been concerned with describing the ways in which people carry out intellectual tasks. Metacognitive research, on the other hand, has been concerned with how people choose ways of solving intellectual tasks that are appropriate for the context in which those tasks are attempted, and how they design their environment to make such achievement possible. In this view, problem solving strategies are resources for action and the intellectual skill is in context-appropriate deployment of those resources.

The resources model of human-computer interaction presented here makes a similar move away from viewing planning as a mental process that generates action, towards viewing plans as one of many resources that can be used consciously and intelligently to decide future action. This re-viewing of the status of plans in H-CI theory, leads us to conclude that plan-based interaction and situated interaction rather than being two competing accounts can be more usefully viewed as different interaction strategies or styles. The resources model leads us to postulate that it is the distribution of information resources and the particular nature of their expression which determine the interaction style or strategy that will emerge in an interactive episode. For example, using an action affordance with no reference to a goal resource results in trial and error. Actions are chosen at random and not directed towards any goal. Space has allowed us only to hint at the different interaction strategies or styles that emerge from different configurations and expressions of resources. But this is a promising line of further research particularly with respect to looking at errors in a way that emphasised interaction failure rather than human error. In addition it suggests that issues of agency, initiative and control might be understood in terms of resource distribution.

This re-viewing of human-computer interaction also leads us to raise different kinds of questions about the design and

evaluation of interaction. For example, what are the key resources necessary for an interaction to proceed smoothly? When are information resources so scarce that interaction cannot proceed? When does the interaction require information from the user rather than from the system? What interaction styles are supported by the interface? The analysis of the chart wizard showed how, by asking such questions, strengths and weaknesses in the design interaction can be assayed.

The chart wizard analysis used to illustrate the resource model, is a small example and some examples of larger scale modelling are required. Nevertheless, the example serves to validate the key concepts of internal and external resources and highlights how the configuration and expression of resources shapes interaction. The terms *internal* and *external* pre-suppose a user-centric perspective. A system designer might use those same terms with opposite meanings. Further work might more usefully replace this terminology with *user-embedded* and *device-embedded* resources.

The idea of the resources model was inspired by a reading of the distributed cognition literature and an increasing dissatisfaction with task-based approaches to H-CI. Distributed cognition is particularly concerned with the social context of work and the notion of distributed representational state, and the importance of mutual knowledge in guiding action. Our concern here has been to provide a much more microscopic account of interactional context while at the same time maintaining a conceptual link with the distributed cognition literature.

A major source of our dissatisfaction with task-based models of interaction is the failure of such approaches to say anything about the importance of context in interaction and the distinction between tasks as described and task as observed in practice. This is becoming recognised as a major limitation by many authors. Nardi (1996) has spoken of a "fundamental need for a theory of practice in human-computer interaction" which brings together research in H-CI, CSCW and distributed cognition. A key concern in all of these areas is the problem of characterising the context of action. The resources model is not the bridging framework that Nardi seeks, but it does liberate H-CI research from a pre-occupation with task description an begin to talk about interaction in a way that might make that bridge building easier.

REFERENCES

Diaper, D (1989) *Task Analysis for Human-Computer Interaction.* Chichester: Ellis Horwood Ltd.

Howes, A. and Payne, S.J. (1990) Display-based competence: towards user models for menu-driven interfaces. *International Journal of Man-Machine Studies,* **33**(6):637-655.

Hutchins, E. (1995) *Cognition in the Wild.* MIT Press.

Kitajima, M and Polson, P.G. (1995) A comprehension-based model of correct performance and errors in skilled, displayed-based human-computer interaction. *International Journal of Human-Computer Studies* **43**(1): 65-100.

Nardi, B. (1996) *Context and Consciousness: activity theory and human-computer interaction.* MIT Press.

Norman, D.A. (1988) *The psychology of everyday things.* Basic Books.

Palen, L. (1993) *Distributed Cognition in the Cockpit: The Second Officer's Panel as a Communicative Resource.* Undergraduate Honors Thesis, Distributed Cognition Laboratory, University of California, San Diego.

Reason, J. T. (1990) *Human Error.* Cambridge University Press.

Sternberg, R.J. and Wagner, R.K. (1994) *Mind in Context: Interactionist Perspectives on Human Intelligence.* Cambridge University Press.

Suchman, L.A. (1987) *Plans and Situated Actions: The Problem of Human Computer Interaction.* Cambridge University Press.

Zhang, J. and Norman, D.A. (1994) Representations in distributed cognitive tasks. *Cognitive Science* 18, pp. 87-122

Usability and Database Search at the Swedish Employment Service

Carl Martin Allwood
Department of Psychology
Göteborg University
Haraldsgatan 1
S-413 14 Göteborg, Sweden
tel +46 31-773 1656
Psycmall@psy.gu.se
http://pew.psy.gu.se/cms/allwood.html

Sara Thomée
Department of Psychology
Göteborg University
Haraldsgatan 1
S-413 14 Göteborg, Sweden

ABSTRACT
In this paper we report on the problems the employees at the Swedish Labour Market Administration have when using the current database application (AF-90). This is done from the perspectives of functionality, usability and information needs. We also, from the same perspectives, analyse a current developmental project (AIS-95) which has as its goal to replace AF-90 with an improved program. No systematic attempt has been made in the development project to integrate knowledge about users problems with the AF-90 project into the AIS-95 project.

Keywords
databases, employment service, usability, information needs

INTRODUCTION
Database programs exemplify a type of software that can be used to help people to communicate. Our interest is in how database programs can be improved with respect to their function as mediators of communication between different categories of users. In such work a large part of the total situation in which the program will be embedded needs to be included. By performing an analysis both of users' needs and of the development process wherein the application programs is developed we can increase our understanding of how the adaptation of the software product to the users' needs can be improved.

The program analysed (AF-90) is used by the Swedish Labour Market Administration (AMS, responsible for converting Swedish labour market policy into practice and staffed by 6.600 employment officers), for example to match job-seekers and employers and to look up information in registers. During the 1993/94 fiscal year the public employment offices dealt with more than 1.8 million assignments concerning job-seekers and registered job vacancies. In 1993, 92 % of the published vacancies were registered at the employment services. The present study is mainly focused on those parts of the Administration concerned with the matching of people in search of work (job-seekers) and vacant jobs (job openings); the employment service.

AF-90, a character-based client server application introduced in 1989, is the main application for supporting the employment officers' work. The number of users have since 1989 increased from approximately 2,000 to 8,000. Its main functions are for registration of job-seekers, job openings and employers, making searches in the registers, automatic monitoring for job openings, assigning jobs and administrative support. The AF-90 makes use of six registers: The Employer register, The Order register, The Jobseeker register, The Occupational register, The Educational register and The Employment Training register (Arbetsmarknadsstyrelsen 1994a).

Functionality, usability and information needs
In order to capture more of the full computer interaction situation when analysing users' difficulties or when specifying demands on a new system we believe that, together with the concept of *functionality* (suitable program functions in the program), the concepts of *usability* and *information needs* are fruitful to use. The usability of a program can be distinguished from its functionality (Goodwin 1987). Usability affects the productivity of the final resulting use situation and can be said to include the users' acceptance of the program, their competence to use the program and the user friendliness of the program. By user friendliness is primarily meant the degree to which the software is compatible with and gives support to the users' mental processes (Allwood and Kalén 1993). It thus concerns design of the program functions. Other components of user friendliness is the degree to which the software is accessible to the user, provides accessibility to its various parts, is adapted to or is adaptable to the specific needs of the individual user and, finally, the availability of effective help resources.

A difficulty with the distinction between usability and functionality is that it can sometimes be difficult to distinguish between when a function is absent in the program and when it is present but poorly designed. This suggests that the criteria for establishing the identity of a program function are not clear.

The concept of usability is also associated with usability oriented developmental techniques (Whiteside, Bennet & Holtzblatt 1988). Much research has concluded that by use of such techniques as user testing, the chances for a more usable end-product are improved (e.g. Karat 1992).

Information needs concern how the information which is mediated within the constraints of the computer software relates to the users' broader information needs. Users can be categorized with respect to the social role/s they have in relation to the program (Mantovani & Bolzani 1994). Information needs can then be analyzed for each of the

social roles defined in relation to the system. For example, in the present context we can differentiate job-seekers, employment officers and employers. On an abstract level the concept of information needs can be said to concern the properties of the information needed when solving the task. These properties can be expressed in terms of appropriate content, form (refinement), reliability and accessibility of the information.

Contrasted with information needs, functionality has to do with the program functions and concerns what can be done with the information entered. The concept of information needs complements the concept of usability by bringing attention to the user's (social role's) information needs in connection with the task/s that she is attempting. Finally, it should be clear that the notions of functionality, usability and information needs should be seen as ideal types and that they in practice sometimes overlap.

METHOD

Our data consists of semi-structured interviews with seven placement officers and three guidance counsellors from four employment offices in Göteborg and its surroundings. In addition observations of seven of these persons using the system was carried out. Furthermore, three system development managers at the National Labour Administration Board (AMS) and one local project manager were interviewed. The interview questions were constructed on the basis of eight pilot study interviews. In Allwood and Thomée (1995) a full-scale report of the present study is given.

RESULTS

We first present the results concerning the users' problems when using the AF-90 system. First problems in connection with searching/matching in the system are presented and them problems and deficiencies experienced in connection with other aspects of the system. Finally, we describe some aspects of the current developmental project (AIS-95).

Problems in connection with searching/matching

All of the employment officers who used the system for matching job-seekers with job openings (at least 8 of 10) perceived problems in the matching process. A general impression was that it was difficult to search on a detailed level, for example to find the "exact person wanted". Since the problems reported had to do with an installed function we have not used the functionality category here and classified the problems as having to do with either usability or information needs. We first present problems having to do with *usability*, starting with those that we find most functionality oriented.

The most important usability problems reported in connection with the matching process are shown in Table 1. Most important may be that the users lack the possibility to register certain types of information about the job-seekers, for example that a job-seekers is a candidate for more than two jobs. Furthermore, the occupational classification codes (NYK) were experienced as being out of date. An important usability deficiency with respect to matching was that the codes were difficult to remember. Moreover, search-words used in addition to the NYK codes were felt to be too limited and to be unreliable since they were used in different ways by different users. Also, the system's support in such situations was experienced as inadequate. In general, users had problems when delimiting the search profile in order to arrive at a search result of useful size. The matching function often gave too large search results. Another deficiency was the restrictions on the matchings that could be made with candidates from outside the own municipality. Furthermore, switching between registers in connection with matching and keeping track of information resulting from a matching were felt to demand too much short term memory of the user.

TABLE 1. EXAMPLES OF USER PROBLEMS WHEN MAKING SEARCHES AND/OR IN THE MATCHING PROCESS.

Usability
NYK - The occupational classification system:
- does not cover existing occupations
- is sometimes too specific
- job-seekers can only be registered on two NYKs
- the NYK-coding system is only found in a paper manual
Added search-words:
- arbitrarily used
- not always registered when appropriate
- lack of words
Incompatibility with users' short term memory:
-.not possible to make a search directly from the form in a register
- too much information in codes to remember while leaving a register to form a search question in a search form
Too large search results:
- when using only NYK
- difficult to browse through large search results
- extra work when more than 150 hits in a search
Lists of search hits cannot be saved, while interrupting one search to make another one
Extra work to add job-seekers from other offices/ counties in a search
Information needs
Not enough/inadequate information registered:
- due to employment officers' lack of time
- customers not submitting information
- customers not updating information
Un-registerable information:
- e.g. job-seekers personal qualities
Difficult to find information:
- free text areas are not searchable
The employment officers do not "know" the job- seekers/ employers:
- wish to scan job-seekers photo into the form - important with personal contacts

With respect to *information needs* and matching, the employment officers sometimes wanted access to other

types of information about the job-seeker or the employer than was available in the system. Furthermore, they wanted greater flexibility with respect to the kind of information which can be entered in searchable locations when registering job-seekers and job-orders. Different problems with information needs are shown in Table 1.

Problems in connection with other aspects of the system

Problems associated with other aspects of the system than the matching process were analyzed as related to functionality, usability and information needs. The most important problems are shown in Table 2.

With respect to the system at large and *functionality* an important deficiency reported was that the wording in the automatic letters to be sent out in various connections could not be changed although it was sometimes felt to be inadequate. With respect to *usability* oriented aspects of the system it was complained that there was not enough time for training. Furthermore, switching between registers were felt to demand too much short term memory of the user. Entering and searching the Educational register was experienced as awkward. Long waiting times and system shut downs annoyed the users and made their work more difficult. Users wanted better possibilities to customize the system according to the specific needs of the situation, for example to be able to include or exclude information add-ins according to the wish of the employer. The help function was reported to be nearly only used for looking up code-words and not for any other purpose.

TABLE 2. EXAMPLES OF USER PROBLEMS IN CONNECTION WITH OTHER ASPECTS OF THE SYSTEM.

Functionality
External communication:
- e.g. summons are automatically created and only to a small extent adjustable
- only a simple writing function in the AF-90
Manual handling:
- a wish for computerization of paper forms
- a wish to receive budget feedback within the system
- a wish to be able scan information in to the system
Usability
User competence:
- not enough user- training
- not enough time to practice
- users are not competent on the parts of the system not used so much
- sometimes difficult to identify mistakes
Accessibility:
- shut downs
- slow response time
Information needs
Information spread too slow within the organization
Manual handling of information
Information not updated or incomplete

With respect to *information needs* the most important complaint was that the employment officers did not receive information about changed laws and regulations for the labor market fast enough.

The system development project (AIS-95)

We next describe some of the actions taken in the AIS-95 project in order to see to it that the new system will have better effective usability than the old. Below, we utilize information mostly from the interviews with the four system development managers but also text material and material from the interviews with the ten employment officers. The central project manager pointed out that those who initiated the project four years ago probably had no idea of how large it eventually would become. The size of the project has made it progress rather slowly. Initially, the project had focused mainly on two parts; the handling of cases and the database containing employers/companies (later to become sub projects of the AIS-95 project). Next, it was decided that also the handling of employment training was to be included in the development project, and that a new placement system was to be developed to substitute the AF-90. A thorough analysis of the information processing within the Administration was done (Arbetsmarknadsstyrelsen 1994b), as a complement to a first version of the system description the so called Town Plan. From this analysis data modelling could take place, and then the application would be structured out of this.

Building a system for more than 10,000 users means a large investment, especially since one tries to buy late models when buying new computers. However, it was not considered possible to purchase new computers for all users at the same time and keep up with the fast development in the computer industry. Since 386 PC's are the minimum standard of the Administration today, even though many users already have 486's, the system would have to be built to run on 386's. Forthcoming changes in labor market policy cannot be anticipated, so the aim has to be to construct the system in such a way that adding and subtracting of programs is possible. The aim is also to work with standard technology, products and development methods, so as not to create a Labor Market Administration standard and thus make maintenance and further development difficult.

The manager of the sub project concerned with the new placement system (AIS-F), mentioned problems due to the size of this project. Another problem mentioned by the AIS-F manager was the lack of money. The AIS-F project was started in February 1994, but no resources were supplied, so the following six months had to be focused on getting resources for the project. When money finally was supplied the project work had to start up very quickly. At the time of the interview there was no time-schedule set yet for the new placement system, and the project manager was again fighting for resources. More funding for the project was later received but the planning processes had been somewhat disturbed by the financial uncertainty. Further turbulence was caused in the project when there was a change of manager in the early fall of 1995.

The sub projects involved with the handling of labor market programs (AIS-H) and the purchasing and handling of employment training (AIS-AUTB) was not experienced by the AIS-F manager to have the same financial problems, since these are calculated to save money for the Administration by saving personnel in the future. The development of a new placement system has more to do with making the matching process more effective and to increase the quality in services. No calculations in terms of economic gains had been made. The most important reason for the delays with the new placement system was said to be that the demand specification for the system took longer time to develop than had been anticipated. Part of the reason mentioned for this was that one had high ambitions with respect to the user interface.

User participation in the system development process
Few of the employment officers in the study experienced that they had influence over the system development, but most of them mentioned that they had the possibility to forward their point of views - if they wanted to - through the officers with system management responsibilities to the county board and AMS. The system management officers were considered to have an intensive dialogue upwards. Some informants mentioned having forwarded their points of views with no results. One officer pointed out that it was important to "know the right persons" to have an influence. A couple of employment officers mentioned that the county labor board system management sometimes came to the office for discussions. Several mentioned that the search-words that were continually entered into the AF-90 were influenced by themselves or the office they worked at.

The informant from the local project in Göteborg for introducing the AIS-95, didn't consider the possibility to influence the system development to be large. When testing prototypes, for example, there had been some possibility to influence, but mostly with respect to small changes rather than large ones. According to the informant, the managers of this local introduction project had received input from probably seven hundred out of eight hundred users in the region and taken this into account when communicating with the central project group.

Still, there was much user participation in the developmental project. There were users present in the project groups, there was a consultative group consisting of users, and a network of employees with system management responsibilities in all counties of Sweden.

Users had participated in *project groups* from an early stage in the development project. The Town Plan was formulated in seminars with user participation (e.g. placement officers and guidance counselors). A further example of user participation is the *consultative group*, which consisted of six members, representing employment offices, employability institutes and labor county boards from different areas in Sweden. The group had previously been selected to make a complement report (Arbetsmarknadsstyrelsen 1994c) to the model-report that had been the basis for the Town Plan. Later, this group became a consultative group for the AIS-95 project and had for example evaluated the near prototype of the occupation templates.

There was also feedback from users by means of a *network* on questions having to do with computer support. The network consisted of 70-80 method managers (i.e. non-technical persons responsible for taking advantage of the computers' possibilities to support the work at the employment offices) from each county labor board. These met at least twice every year. Each county had a corresponding regional network. The network was considered a good way to spread information and ideas and to receive feedback. It has been used in order to mediate prototypes, multislide presentation etc for testing.

During 1992, AMS had carried out what was called "the Trouble Hunt", where all employees were to contribute suggestions on what could be improved or rationalized in the organization. The Trouble hunt was not part of the AIS-95 project. However, many of the suggestions came to focus on the computer support and the results from the "hunt" were used as input in the AIS-95 project. Finally, AMS' representatives had participated in seminars in the counties primarily in order to give information, but they had also received feedback.

DISCUSSION

Although the present system appears to be a clear asset for the employment officers, it seems to suffer from deficiencies. The deficiencies reported were analyzed in terms of functionality, usability and information needs. As noted above these aspects may overlap in practice. The work with the analyses gave concrete experience of this. For example, it was sometimes difficult to separate information needs and functionality. For instance, the extent to which a particular program function allows or supports the communication of some particular information is often a matter of degrees. For example, AF-90 allows information about personality properties of the job-seekers but only in the so called free text area where the information is more difficult to search.

In spite of the difficulties that may occur when using the three categories functionality, usability and information needs when analyzing the effective usability of a computer application we still think that they are worth retaining since they can help the system developer to find deficiencies in the application which may otherwise not have been found.

As reported in more detail above, the system, in general, was felt to be cumbersome to use. We next consider the extent to which the organization of and work in the AIS-95 development project has made it possible to avoid these deficiencies in the new system developed.

System development processes are not totally rational. Instead they are constrained by socially and otherwise determined circumstances. This is well illustrated in the analysed system development process. For example, the project was not planned to be of the size it is today. Due to this, no full grip of the situation was taken from the beginning. An effect was that the *project organization* to some extent appears to have developed spontaneously over time

rather than to have been carefully planned from the beginning. This has resulted in one of the sub projects (AIS-F) reaching a cumbersome size, making it difficult to administrate. Other effects of the unplanned development were delays and, in periods, a lack of resources for at least one of the sub projects (AIS-F).

A general lack of money has naturally put constraints on the project. For example, it has limited possibilities to utilize front line technology; the applications developed must be runable on 386-computers. Lack of money also seems to be one, but clearly not the only, reason why the development project is planned to a large extent to build on the existing system. A further reason may be the need not to create too much upheaval in the organization.

To a certain extent it may be possible to foresee the use situation in the new system without any extra attempts at analyses. However, since the new system will build on the old system and will be running in a future stage of the same social context, it seems clear that an important means to arrive at good effective usability is to identify deficiencies in the old system and to see to it that these are avoided in the new system.

To what extent has this been accomplished in the present project? The means to get information about the deficiencies in the present system have been of various kinds. One has established a "report structure" where employment officers report deficiencies to the local system experts who report them to the next level, etc. Furthermore, the consultative group has attempted to reflect the users' experienced problems with the system (Arbetsmarknadsstyrelsen 1994c). That report is oriented towards specification of demands on the new system. Many of the difficulties described Allwood and Thomée (1995) appear to have served as an implicit background to the suggestions given in the report of the consultative group. However, no detailed specification of different problems experienced by the users is given and the extent to which a systematic attempt has been made to analyze such difficulties is unclear.

User participation in a system development project also serves other purposes than gaining information about the system. For example, it can help increase acceptance for the new system. Attempts to accomplish user participation in the project includes the formulation of the Town Plan in seminars with user participation and the establishment of a consultative group representing local organizational units and of a net-work consisting of system managers at different levels. However, our results suggest that these attempts had not been fully successful in creating a feeling of user participation in the project. Many of the interviewed employment officers felt that they had very little influence over the system development. The same feeling was held by a project manager at the local level in Göteborg.

It appears that one reason for this feeling of lack of participation among the users is the great size of the project. As noted above, the present system has 8.000 users and 12.000 users are planned for the new system. A further reason may be the hierarchical and representative approach taken in connection with the user participation issue. Possibilities of e-mailing opinions and wishes with respect to the new system directly to the sub project leaders in Stockholm may improve the situation as may the establishment of general-access computer conferences around different aspects of the new system to be developed.

An good way to reach effective usability in the new system is to work with usability oriented methods, such as user testing. Current wisdom in the system development literature says that, for economic and efficiency reasons, this work should be started as early as possible in the development process. However, in the AIS-95 project user testing appears to have been a late recognized good and other usability oriented work methods, such as the setting of usability goals, not to have been used at all. A further sign that this way of working is not fully integrated into the project is that external consultants have been hired into the project for the usability testing. This risks isolating usability oriented work in the project and of course incurs extra costs on an already weak economy. A better approach could have been to train project members in usability oriented work methods, such as user testing, and to have had the trained personal work within the project.

On a more general level it appears that although the Swedish government has recently attempted to modernize the system development methods for the state sector in Sweden, the AIS-95 project is an example of a specific ongoing development project which is still run according to fairly traditional system development methods.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Swedish Work Environment Fund (AMFO) and by the Swedish Council for Work Life Research. We wish thank all of our informants. Thanks also to Nicolay Beck and members of our projectgroup for comments on an earlier version of the manuscript.

REFERENCES

Allwood, C.M., & Kalén, T. (1993). User-competence and other usability aspects when introducing a patient administrative system: a case study. *Interacting with Computers*, 5, 167-191.

Allwood, C.M., Thomée, S. (1995). Usability and database search at the employment service. *Technical report SSKKII-95.07*, Göteborg University.

Arbetsmarknadsstyrelsen (1994a). Handledning för användare - Förmedlingssystemet AF-90, Yrkessystemet, AMS, Platsförmedlingsenheten. Stockholm

Arbetsmarknadsstyrelsen (1994b). AMV:s stadsplan for informationsbehandling. Stockholm.

Arbetsmarknadsstyrelsen (1994c). Krav på kommande datorstöd inom förmedlingen. Stockholm.

Goodwin, N. (1987). Functionality and usability. *Communications of the ACM*, 30, 229-233.

Karat, C.M. (1992). Cost-justifying Human factors sup-

port on software development projects. *Human Factors Society Bulletin*, 35 (11), 1-4.

Mantovani, G., & Bolzani, M. (1994). Analysing and evaluating multi-actor multi goal systems in use: social contexts and participation in three vocational guidance systems (VGS). *Behaviour & Informational Technology*, 13, 201-215.

Whiteside, J., Bennett, J., & Holtzblatt, K. (1988). Usability engineering: Our experience and evolution. In M. Helander (Ed.), *Handbook of human-computer interaction* (pp. 791-817). Amsterdam: Elsevier Science Publishers (North-Holland).

Utilizing A Concept Model in User Interface Design

Anthony A. Clarke, Basil Soufi, John R. Tyrer & Luise H. Vassie

Department of Computer Studies Department of Mechanical Engineering
Loughborough University, Loughborough, Leics LE11
+44 1509 222 688
a.a.clarke@lboro.ac.uk

ABSTRACT
This paper is about designing the human-computer interface of a Laser Safety Decision Support System (LSDSS). A task structure model, based on the procedures used by laser safety assessors, could not support the design of the interface dialogue. A model based on concepts and schemata used by the assessors was then constructed. The model represents the domains of knowledge, experience, and action used in making a safety assessment. The model was successfully used in designing the prototype LSDSS and its interface dialogues. It is suggested that where the task involves knowledge and its manipulation, concept models may be successfully used.

Keywords
user modelling, interface design, decision support systems, laser safety

INTRODUCTION
Laser safety assessment is a complex task requiring many skills, and even laser manufacturers sometimes experience difficulty with safety assessment. In a survey of safety in 49 laser companies in the UK, for example, just under half had difficulties in assessing laser safety hazards or implementing laser safety Standards (Vassie et al., 1993).

The knowledge needed for safety assessment is not easily available. Much of it is in the heads of a small number of laser experts; the rest is widely spread in various publications. Where information is available it is often very technical and not in a form that can be easily used by safety assessors.

Thus, laser safety assessors have a need for up-to-date practical information in a readily accessible and usable form, and a prototype Laser Safety Decision Support System (LSDSS) was built to meet the need. This paper is about using models in designing the system and its human-computer interface.

THE APPROACH TO THE DESIGN
The Designers' Aims
The designers' aims were to:

1. determine the assessors' model of the laser safety assessment task.
2. use the model for the interface metaphor.
3. make the interface to the knowledge-base transparent.
4. build the prototype system using the model as a tool.

The Assessors
Field surveys showed that laser safety assessors can be:

- expert or not-so-expert in lasers,
- familiar or not-so-familiar with safety assessment, and
- have a lot or little computer experience,
- require general or specific information.

Representative groups of assessors were involved in the design process so that these factors would be taken into account.

The Design Method
It was intended to follow these steps:

1) determine the structure of the assessor tasks and construct a model
2) with the aid of the task structure model, determine the assessor's utterances that achieve the goals in each task
2) segment the assessor's utterances into hierarchically-related subdialogues, then
4) use the subdialogues to design the dialogue structures.

The interface (and the system) designed by using the task model thus derived should reflect the assessor's task, as it is understood. The resulting interface should reduce learning time, and should also give an interface to the system's knowledge base that is transparent (Guindon, 1988).

The Initial Model
Deriving the Structure of the assessor Tasks

A small number of laser safety experts took part in the initial task analysis procedure. There was no opportunity to make recordings of the experts while they made assessments in a working industrial location, so structured interviews were held with them instead. They were invited to verbalise the procedures they followed while 'walking through' a safety assessment in their minds.

The data gathered was then analysed using protocol analysis methods to identify the tasks and utterances involved. After the results had been cross-checked with the experts, a laser safety assessment model was constructed (see Figure 1.)

Laser Safety Assessment as a Procedure

The assessor starts by scrutinising the site and exterior of the building, checking items such as ventilators discharging to atmosphere and storage of gas cylinders. The assessor scrutinises the interior, then proceeds to the work area, checking workers' eye protection, for example.

At the work-space stage, where the laser is doing its work, checks are made of its security, covering safety cut-outs and screening of the laser beam, for example.Then comes the assessment, followed by a report of the assessment. The results of the assessment are fed back into the installation where immediate rectification of faults is required.

Problems with the Procedural Task Model

The procedural model seemed initially to be credible since:-

1) safety assessment has to be procedural because unstructured inspections would yield inconsistent results,
2) the main determinant of the safety assessment task structure was the procedure identified by the expert safety assessors themselves as being the ones they followed in their work, and
3) the items that the experts said they would look for in a safety assessment were the same as items identified separately in the literature.

Attempts to use the model as a programmer's tool, however, revealed that it could not support the implementation of a dialogue structure. Although there was a task structure, the associated utterances of the experts contained no immediate reference, for example, to prerequisite knowledge.

The laser experts were then invited to re-examine the model. They found that it also could not support their problem-solving strategies.

The model was inappropriate and an analysis revealed two main reasons for this:

1) the experts had been unable to identify the bases for their decision making so the model had turned out to be a managerial, ie a non-technical, view of laser safety assessment, and
2) the structured interviews that had been employed to get data were purely 'observational' in nature. Structured interviews are a good general purpose method, but they are poor at eliciting the problem-solving strategies used by experts (Welbank, 1987).

Laser Safety Assessment as a Problem

It was realised that more progress would be made by treating safety assessment as a problem.

The model in Figure 1 may have provided an overview of the procedures, but the mental representation used by people to solve problems consists not only of procedures, but also a set of 'conceptual entities' (Bibby, 1992).

The concepts and schemata used by the safety assessors were required, so they were asked to consider safety assessment as a problem and to identify the most difficult aspect of that problem. They nominated 'making complex decisions using combinations of *laser type* plus *material* plus *process* plus *installation* ..' .

(Definitions: In this paper the term 'concept' is used to mean a mental representation of a category, which allows a person to place stimulii in a category on the basis of similarities between them. The term 'schema' is used to mean an organised body of declarative (knowing that) and procedural (knowing how) knowledge about concepts, a mental representation abstracted from experience. Schema selection is an intrinsic part of problem solving (Howard, 1987).)

In scientific problem solving, experts' schemata are organised in terms of solutions principles, while novices' schemata are based on 'object' categories. However, experts and novices alike proceed in solving problems by first categorising the current problem. This categorisation makes information available which can be used to guide further understanding/solution processes.

The major problem confronting the expert is choosing among candidate schemata, and solution processes are more or less non-problematic once the correct schema has been selected (Chi et al., 1981).

THE CONCEPT MODEL

The concepts and schemata used by the safety assessors were determined. A 'contextual' approach was employed with two groups of assessors – the laser experts and, this time, 15 potential system users who routinely performed laser safety assessment as part of their jobs.

Free-sort tasks were administered to the assessors, where they were given a list of concepts relating to laser safety and asked to group them into categories as they saw fit. In-depth, targeted interviews were also used. In this way the nature of the assessor's concepts, schemata and bases of categorisation were established.

Six major concept domains were elicited. They were

(i) lasers, (ii) materials. (iii) laser energy manipulation, (iv) hazards, (v) workspace, and (vi) environment.

Three major schemata were also elicited. They were

(i) process-energy interaction, (ii) assessment of safety of the laser system, and (iii) assessment of likely failure of components.

This approach proved to be a better way for discerning the assessors' inference strategies, including acquisition and use of procedural and declarative knowledge about lasers, materials, and laser energy manipulation.

Having identified their concepts and schemata, the assessors were then able to explain in depth how they approached and carried out the assessment task. A model was constructed using the the concept domains and schemata as elements (see Figure 2).

Operation of the Model

The assessors accepted the new model as an accurate depiction and were able to use it straightaway as a medium for analysing and discussing the safety assessment task.

Context
The concept model is operated within the context of a meta-plan (or meta-schema) formed by the assessor, where the meta-plan is 'a plan to reason about goals and plans' (Guindon, 1988).

Precursory Level
The safety assessor brings to each assessment a body of knowledge about three major concepts: lasers, materials, and energy manipulation (ie methods and techniques of laser beam delivery).

The extent of any given individual assessor's personal knowledge varies in total, and in each concept. It is bounded by a minimum of qualifying knowledge, at the lower limit, and at the upper limit by the existence of unpublished contemporary research data, or unpublicised industrial applications.

At any given time each assessor could possess different "knowledge fragments". Variation in individual differences among assessors at the precursory level also points to the existence of a knowledge 'core' which experienced and novice assessors share, but the experienced assessors have more elaborate knowledge structures, with extensions to the basic core (McDonald and Schvaneveldt, 1988).

Process-Energy Interaction
Specific 'fragments' of the prerequisite knowledge from the precursory level are applied by the assessor when examining each specific instance of interaction of laser energy in the process-energy interaction.

Each interaction is potentially unique so combinatorial problems could arise. This does not often occur because meta-plans include pre-assessment classifications such as 'welding' or 'etching steel'. These classifications represent the body of pragmatic experience built up by personal experience of safety assessments and preclude most 'impossible' combinations.

Consequential Level
Tasks at this level commence with analysis, evaluation, and interpretation of the hazard (eg reflected beams), the system workspace (eg power and interlock switches), and the system environment (eg fume extraction).

During these tasks certain calculations may also have to be performed to establish values of critical parameters.

The assessor, exercising judgment, then formulates an assessment of a) the safety of the specific laser system, and b) the risk associated with component failure in the laser system under consideration.

Application of the Model
Having been employed to elicit assessment task structures and their associated dialogues, the model was then used to derive the system interface structure, particularly the parts of the interface and structure of the human-computer dialogue responsible for the specification of the problem by the assessor, and the display of recommendations made by the system to the assessor.

Figure 3 shows how the structure of the system interface dialogue and recommendations correspond to the elements of the conceptual model. There is a mapping between the selection sub-dialogues and the precursory-level domains, and between the system's recommendations and the consequential-level domains of the model.

System Dialogue Specification and Design
Precursory Level System Dialogues
Each of the knowledge domains of the precursory level has a corresponding sub-dialogue in the user interface. For example, the laser domain contributed a dialogue for the selection of the laser type, and another for the selection of the laser application. Similarly for the material domain and the energy manipulation domain.

Process-Energy Interaction System Dialogues
At this stage of the assessor-system interaction, the assessor is assumed to have specified a problem situation, and now has to communicate his or her intention so as to get the appropriate recommendations. When this takes place, the system checks the validity of the selections that have been made.

Interface Design for Specifying a Problem
The design of the interface is such that only critical parameters need to be specified by the assessor in order to obtain recommendations. This is in accordance with the principles of minimum critical specifications and economy, that is, users should achieve their task in the minimum number of actions necessary.

When more detailed information is sought, the system requests entry of more detailed relevant parameter values.

Context-Specific Recommendations
The system acts as a decision-support tool by providing context-specific recommendations. This is achieved by displaying advice information that is appropriate for a specified laser material processing situation.

The content of the recommendations displayed by the system arise from consideration of the consequential level of the laser safety assessment model's three knowledge domains.

Hazard Recommendations
Information about hazards is displayed by the system under the categories of optical hazards and non-optical hazards. These are parent categories that lead to a number of sub-headings. For example selecting 'optical hazards' will lead to ten selection options such as eye hazard, skin hazard etc.

System workspace and System Environment Recommendations
Interactions concerning the system workspace and the system environment give information about the laser enclosure and fume extraction, for example, and general guidelines concerning non-laser but laser-related issues (e.g. electrical supply and chemical handling).

THE PROTOTYPE SYSTEM
The prototype was developed with HyperCard and HyperText on an Apple Macintosh computer.

The working system was tested and evaluated by a large number of potential and actual laser safety assessors with varied laser and computer experience. The results of this evaluation indicated that in general the system is acceptable in functionality, is easy to learn and use, and has dialogue structures that show immediate affinity with the structure of the laser safety assessment task as perceived by safety assessors ((Clarke et al, 1996).

DISCUSSION

Knowledge, Expertise, and Objects

The concept model allows the dialogue designer to see that the assessor and the computer form an active dyad, and that some forms of knowledge could be distributed between them. The model also showed how assessors could take different points of view of the problem at hand, so the design had to handle different 'knowledge fragments' (O'Malley and Draper, 1992).

Expert and non-expert assessors had different schemas for the domain in which expertise was examined, although the model at its present level of resolution could not always help to clarify how their schemas differ. The expert and non-expert assessors also shared a 'core' schema, but the experts had more elaborate knowledge structures with extensions to the basic core. Experience, however, may cause even the core schema to be reshaped, making the expert's view quite incompatible with that of the non-expert (McDonald and Schvaneveldt, 1988).

The concept model also allowed the designer to appreciate that an assessor's mental model permits the relationship between objects (category concepts) to be understood, thus reflecting the assessor's understanding of "what the system contains, how it works, and why it works that way" (O'Malley and Draper, 1992).

The existence of objects and the intention of manipulating them also encouraged the designer to consider the assessors' activities and their plan- or goal-directed nature. However, the behaviour of the assessors does not simply result from being driven by internal goals, but is more likely to be the result of interplay between goal-directed activity and the specific situation that they find themselves in (O'Malley and Draper, 1992).

Thus, the focus of activity in the concept model is on the specific, situational, process-energy interaction, whereas there is no real focus in the procedural model.

Placing the Critical Decision Areas

The laser safety experts, as potential users of the computer-based system, used the concept model in articulating their wish for the Critical Decision areas of process-energy interaction and safety/risk assessment to remain in the human decision-making domain.

They were concerned about the risk that less laser-experienced assessors could perceive the system as having a greater degree of vested 'authority' than the current state of knowledge of laser safety assessment would warrant.

They also opted for the computer to be given the routine aspects of the job, leaving them with the creative work and an information handling tool to help them with this work. They were, in effect, supporting the 'enrichment' hypothesis and rejecting the 'deskilling' hypothesis (Eason, 1990).

The Interface Reflects the Task

The requirement that process-energy interaction, safety-, and risk-assessment decisions be externalised from the computer system is reflected in the structure of the system's dialogue, and in the functionality available at the different stages and levels of the interactive dialogue.

On the basis that it is not "user-friendly" to make a system easy to use now if it is later limiting in terms of the user's increasing knowledge over time (Stevens, 1983), the system was also designed to be expandable in the knowledge domains so that growth of knowledge on the assessor's part would not be limited by the system. The concept model suggested how this should be done.

Also, the focussing at the process-energy interaction indicated how the dialogue should be structured at the equivalent point in the sequence of use of the system.

CONCLUSIONS

1) The concept model represents the generality of the approach to assessment, the situational nature of safety and risk assessment, and post-assessment activity. Thus, the complex task of laser safety assessment has been successfully modelled.

2) The design aims were met in that:- the assessors' model of the laser safety assessment task was established, the model was successfully used for the interface metaphor, the interface to the laser safety assessment knowledge-base was made 'transparent', and the model was successfully used as a major tool in the prototype design.

3) Mental models (in general) and concept models (in particular) are more useful for supporting computer system and interface design than procedural models (ie those based solely on user actions) in those circumstances where user tasks involves mental representations of objects and their relations, distributed knowledge and its manipulation, and user experience.

4) There can be some risk in using concept models of the task derived from the user where the user's concepts of the task are limited, 'erroneous' or false, or do not lend themselves to being modelled.

ACKNOWLEDGEMENT

This research was supported by ACME Directorate of the Science and Engineering Research Council, and was done under grant GR/F/94279, in collaboration with the Health & Safety Executive and the National Radiological Protection Board, UK.

References

Bibby, P. A. (1992). Mental models, instructions and internalisation. in Y. Rogers, A. Rutherford & P.A.Bibby. (Eds). *Models in the Mind.* Academic Press, London, 153-172

Chi, M.T.H., Feltovich, P.J. & Glaser, R. (1981). Categorisation and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152

Clarke, A.A., Soufi, B., Vassie, L.H. & Tyrer, J.R. (1996). Field evaluation of a prototype Laser Safety Decision Support System. *Interacting With Computers,* 7(4), 361-382

Eason, K. (1990).*Information Technology and Organisational Change*. Taylor & Francis, London.

Guindon, R. (1988) A multidisciplinary perspective on dialogue structure in user-advisor dialogues, in R. Guindon (Ed) *Cognitive Science And Its Applications For Human-Computer Interaction*. Lawrence Erlbaum Associates, London, 163-200

Howard, R.W. (1987). *Concepts and Schemata*. Cassell Education, London.

McDonald, J.E. and Schvaneveldt, R.W. (1988). The application of user knowledge to interface design, in R. Guindon (Ed) *Cognitive Science And Its Applications For Human-Computer Interaction*.. Lawrence Erlbaum Associates, London, 289-338

O'Malley, C. and Draper, S. (1992). Representation and interaction: are mental models all in the mind?, in Y. Rogers, A. Rutherford and P.A.Bibby (Eds) *Models in The Mind* . Academic Press, London, 73-91

Stevens, G.C. (1983). User Friendly Computer Systems?: A critical examination of the concept. *Behaviour & Information Technology,* 2 , 3-16

Vassie, L.H., Tyrer, J.R., Soufi, B., & Clarke, A.A. (1993). Lasers and laser applications in the 1990s: a survey of laser safety schemes. *Optics and Lasers in Engineering,* 18, 339-347

Welbank, M. (1987). Knowledge acquisition: a survey and british telecom experience, in *Proc. 1st European Workshop on Knowledge Acquisition for Knowledge-Based Systems*. Reading University: Institution of Electrical Engineers, 23-54

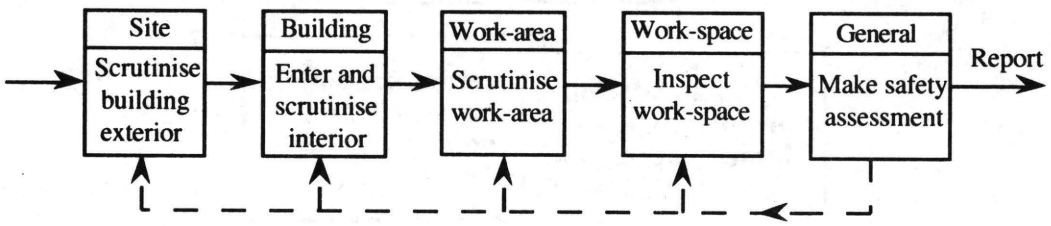


Figure 1. A Procedural Model of Laser Safety Assessment

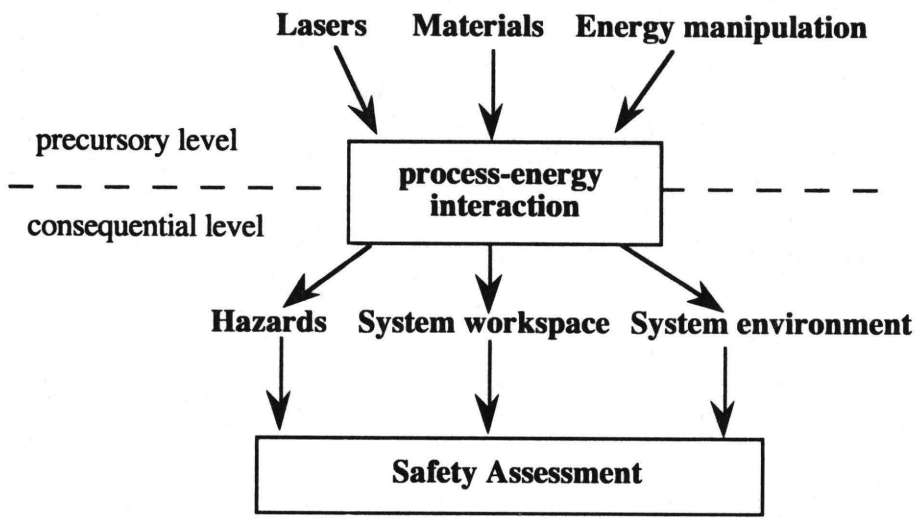


Figure 2. A Concept Model of Laser Safety Assessment

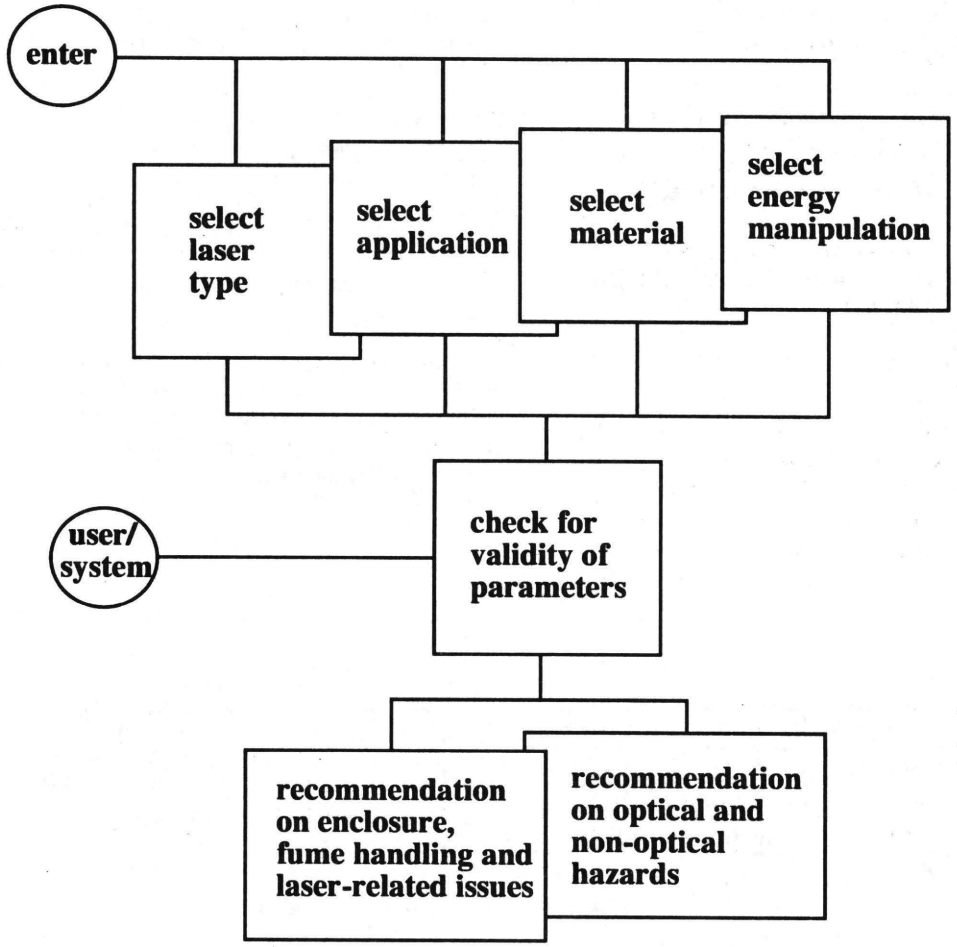


Figure 3. Interface Dialogue Structure mapped from Concept Model

The Diversity of User Participation in System Development: a Video Analysis

Carolyn Axtell, Patrick Waterson, Chris Clegg
Institute of Work Psychology
University of Sheffield, Sheffield S10 2TN, UK
+44 114 2756600
C.M.Axtell@sheffield.ac.uk

John Rae
Dept of Psychology and Counselling
Roehampton Institute
London SW15 3SN, UK
+44 181 392 3612
JRae@max.roehampton.ac.uk

ABSTRACT
Using a video-based analysis we examine users' activities during their participation in system development, the knowledge they require to complete them, and the way they organise their work. We find that the users' activities are diverse and that the availability of the necessary cognitive resources is heavily dependent on the work organisation of the development team. Implications for the conduct of user participation and the role of "the user" are discussed.

Keywords
user participation, activities, knowledge, work organisation.

INTRODUCTION
User participation is widely regarded as an essential part of system development and as a means of ensuring that knowledge relating to the application domain is included within the final system (Curtis et al., 1988). Participation of users can lead to better products as developers gain knowledge about users' practice and anticipated use situations. This is particularly important for the design of systems for "in-house" use (Eason, 1995).

Similarly, within the work organisation and sociotechnical literature there is a strong tradition which advocates the participation of employees in decisions and changes that affect them (Cherns, 1987). Highly participative work teams are advocated with control over the methods they use to conduct their work, and access to all relevant information needed to take action. Participation has been identified as a means of improving motivation, satisfaction, productivity and acceptance of change.

Development teams can be seen as cognitive systems (Curtis et al., 1988). Within these teams, knowledge sharing, the use of artefacts, resources and the organisation of work affect the processes and outcomes of their tasks. Likewise, there has been a recent call for the move to a "process" rather than "product" focus on system development (Grønbæk et al., 1993). The argument is that only by understanding the processes involved and problems that need to be overcome, can we hope to optimise the conditions for including the organisational and human issues required for successful systems.

In order to go some way toward achieving this goal, examples of user involvement need to be explored in more detail than has previously occurred. There have been few attempts to study the structure and diversity of user activity. In particular, there are few detailed examples drawn from real software projects situated within their organisational context. The purpose of our study was therefore, to investigate the work of a group of users involved in system development in an organisation. We aimed to gain a better understanding of the tasks they completed, how these were done and their relationship to the wider organisational and cognitive constraints in operation (Axtell et al., 1995). As such we examine what they did, the knowledge required to do it and the way in which work was organised. We were motivated in part by the need for a process focus, and the re-examination of central concepts in HCI, for example "interface" (Grudin, 1993), and "task" (Draper, 1993). The re-evaluation of such concepts and the increasing emphasis on user participation in systems development prompted us to pay closer attention to the role that users take. With this in mind we offer in this paper an analysis of the concept of the role of "the user", within the development lifecycle.

THE CASE STUDY
The study was conducted in the IT department of a national, UK based organisation, where users were involved in the design and evaluation of a large system for nationwide use. This system (which we will call AS - Administrative System) was designed to support a major administrative function which varied slightly in different local offices. AS replaced and expanded upon an existing computer system. Users from different parts of the country went to Head Office for periods of 1 to 5 weeks to be involved in the development of the system. The method of user involvement in use was intended to mainly address usability issues. Prior to the researchers' involvement, much of the system had already been designed using cooperative prototyping, with users and developers in direct contact. However, due partly to concerns about the time this was taking and a change of deadline, user participation

was severely reduced. As such, users were conducting evaluation exercises as a group, with at least four users present at any one time. There was no direct contact with developers, and written comments were passed to developers almost exclusively via a user support team, which consisted of non-technical personnel. HCI experts were also involved in development in an advisory role.

RESEARCH METHODS

To achieve a closer understanding of user behaviour we employed a range of methods, using video-taped records as a primary resource for the study. Video has been successfully used in similar contexts and has allowed researchers to gain an understanding of user requirements (Brun-Cottan and Wall, 1995) and the complexities of the work setting and work practices associated with it (Brun-Cottan, 1990/91). We supplement our analysis with interviews conducted with users as well as written system evaluation reports.

Video and Observation

In total, we video-taped and observed approx. 84 hours of user activity, during a three week period, excerpts of which we describe within the paper. The researchers were present during the recording and so were able to note user activities at a high level and observe periods or events of interest (examples of different types of activity), which could be reviewed later on the video tapes. The researchers then reviewed the sections of interest (approx. 12 hours) in detail and content analysed them for recurring knowledge and work organisation themes.

Interviews

We interviewed 21 users about their experiences and attitudes towards the system development approach during their stay at head office. The findings from these will not be discussed in detail in this paper, but will act as supplementary to the issues raised by the video analysis.

Written Evaluation Reports

Users completed written evaluation reports which were sent back to developers via the user support team. Users reported problems they came across, or issues they wanted to raise on specific forms. 245 of these reports were collected from users over the development process. Again, this data will only be referred to where it complements issues raised by the video analysis.

FINDINGS

In this section we describe the activities the users were involved with during this period, the knowledge they needed to conduct those activities and the way the work was organised. High level analysis during this period shows that users were involved in a variety of activities. We grouped these into the five main categories below.

Activities

- Entering materials on to the database, so that trials could be carried out on the system;
- Off-line work, including making lists of materials to be checked and tested, and devising work based scenarios which were later used to evaluate the system;
- Designing and writing on-line help screens;
- Running work based scenarios, i.e. carrying out formal trials and role playing typical office activities;
- Trouble-shooting, which included general usability testing (ensuring consistency, intelligibility, positioning etc.) and identifying bugs, errors and suggesting modifications to the system.

Users spent most of their time troubleshooting during this period (see table 1). The percentages of user time reported in table 1 indicate the main activity that the users were engaged in during tasks being completed. However, there was clearly some overlap between these. For example, when running scenarios, many usability issues, bugs and errors were detected. In addition, elements of design were involved in troubleshooting, such as when suggesting modifications.

Table 1: User Activities

Activity	% of user time
Entering database materials	5
Off line work	10
Designing help screens	20
Running scenarios	20
Trouble-shooting	45

Detailed analysis of the videos revealed that there were several dominant knowledge and work organisation themes present within the users' activities. These will be explored below.

Knowledge

It was found that users needed to draw upon a number of technical and non-technical sources of knowledge, which they could not always gain access to, in order to complete their work.

Technical Knowledge

System Functioning

Users needed to know **what** AS could do and also **how** it did it, which was often discussed in relation to how the existing computer system in local offices worked. For instance one aspect of the users' jobs was categorising types of client into one of hundreds of possible standard categories. Two users were certain that the existing system in local offices told them if an incorrect categorisation was made. They experimented with AS to see if it did the same by entering a category that they knew would be incorrect, but the system did not appear to detect the error. The other two users however, were less frequent users of the existing system and were surprised that such a feature was available on it. What followed was a long discussion (for about ten minutes) on how the system could possibly match a category code with a typed description of a client, unless the same specific wording was used each time. Logically, they realised that the system would have to recognise that things like "joiner", "carpenter", and "self employed joiner" should be under the same category. The users could not resolve the issue amongst themselves, but wrote it down on an evaluation form anyway. However,

the user support team did not raise the matter formally with developers.

Current State of Development
Issues here include knowledge about whether certain aspects had been programmed in yet or had temporarily been made unavailable. For instance, there were some standard letters and forms which users had seen as printouts previously, and spent some time looking for unsuccessfully on AS. The forms had not been integrated into the rest of the system at that time. Another issue in this category is knowledge of whether matters they had already raised were being dealt with, or would be in the near future. There were several occasions where users knew they had written down a problem a few days ago, but decided to raise it again, "just to make sure", because it had not been incorporated into the system as yet. Users were also unsure whether their suggestions had been raised by previous user groups and therefore were a waste of time for them to worry about.

Design Rationale
There were some cases where users did not know the reasons why particular features were available. For instance, there was a new entry field added which specified the skill level of clients. Users were not sure if skill level had been introduced so that certain searches could be performed or if it was for another reason. Therefore, they did not know how to respond to the feature and spent several minutes discussing possible explanations. Other instances occurred where AS asked for more information than users would normally have to obtain in their work. Two types of client were dealt with by the users; one of which required the user to acquire less information than was needed for the other. Users were about to suggest that some of the entry fields be removed for the first type of client, when they were informed by an HCI expert, who was there at the time, that the intention was for client screens to be consistent across the system, and as such all the fields would be needed. Occasionally, also, there were terms or labels on the screen which did not make sense to users (e.g. when some abbreviations were used) and so these needed to be explained before changes could be suggested.

Possibility of Changing Features
Users spent a lot of time discussing whether it was possible to change certain features. They did not know if their suggestions could be implemented easily, particularly at this late stage of development, such as altering the number of steps for conducting certain searches, or adding a reminder to do something when closing a particular screen. Sometimes users made requests to alter aspects of screens which turned out to be standard software features, and were therefore impossible to change.

Non-Technical Knowledge
Current Working practices
Users needed to know, and generally did know about current working practices in local offices so that they could identify potential problems for conducting their job on the system. They often spontaneously invented their own work based scenarios when working through different features. However, sometimes there was disagreement on whether someone's reported experience was correct (such as their experience of the existing office computer system) and disagreement on what most people were likely to do most of the time, due to their slightly differing experiences.

Future working practices
Users needed to know about future work practices in their offices. For example, a large part of local office work involved collating statistics from the existing computer, on office performance. There was a concern that certain new features, which were required for other purposes on AS, would cause problems for the usual statistics. Users spent 15 minutes debating whether the statistics they would be required to collect would be different in the future, involving aspects of these new features. Without knowing the answer to this, users could not be sure if their concern was justified or not.

User Involvement
Users needed to know how to be involved in system development. For example, they needed to know how to use the evaluation forms and whether to write down everything they discussed or only selected points, and if so, which ones. Sometimes users argued over whether they needed to write something down, or whether they needed to worry about certain issues. Issues also arose here in terms of their knowledge of how to organise their work, i.e. who had what expertise, who was better at particular aspects of the task etc.

Work Organisation
An important area of user activity which previous research has not addressed is how users organise their own work within system development. In this study, users were often given a task but the method of achieving it was at their own discretion. This might be considered as "user organised activity". "External factors", beyond the control of users are other important areas of work organisation which could be considered as the supports or constraints influencing their work. Aspects of these two areas are discussed below.

User Organised Activity
Division of Labour
At interview a number of users said that they enjoyed organising their own work and would like more input in this area. Users organised who would work together and who would do what. This included deciding what order to do things in, whether to work in pairs, or as a group, who would fill in the evaluation forms, who would use the keyboard, who would play a particular role in roleplays etc. This was usually very effective such as allocating keyboard tasks to the fastest typist, or a role to someone most familiar with it, and certainly maintained their motivation and output. However, there were instances where it was less effective, such as when users were not sure exactly what task they were supposed to be completing. Another example of less effective work organisation occurred when 4 users were working together and one of them was writing the evaluation forms, but was having difficulty keeping up with the group. This was partly because the group moved on quickly after making a decision, and partly because the

user was also trying to take part in the discussion. Often the writing interrupted the flow of conversation while they waited for this user to catch up.

Use of artefacts

The computer system obviously needed to be used to walk through work based scenarios or when detecting bugs, but it could also be used as a vehicle for user discussions, such as when explaining problems to each other or when experimenting to see what would happen when incorrect information was entered. In addition, during the design of help screens the users organised their work on two computer screens; one displaying AS and one displaying the help screens as they typed in relevant help information. They also used other artefacts to help them in their evaluation work, such as the standard office cards that the computer printed out on. In this case users found that the space on the screen where they typed in the information they wanted to print out was too large. As such, not all the information they typed could fit on the card, and so the space on the screen needed to be reduced.

Decision making

Users needed to make decisions about aspects of the system and which issues to raise. Sometimes users were not sure what to do. Usually they erred on the side of writing down all the issues they discussed "just in case" although on occasion, some users felt strongly about an issue which they felt had not been raised in sufficient detail. Decisions were also being made on incomplete information and so many assumptions were made e.g. over whether something would be possible to change, or what the rationale behind certain features was etc. Often users' discussions included reference to several issues, so that problems discussed were nested within others that were being discussed. Conversations about some matters revealed other related problems. This meant that decisions were often made after several issues had been discussed, or in the middle of debates about others.

Written communication to developers

Users had to write down issues to be raised with developers as coherently and fully as possible. This meant that a discussion which may have taken 20 minutes was reduced to something like 5 lines of text. Additionally, users did not always see what the person writing had put, although sometimes one person dictated and the other one wrote on the form. It was difficult to make the written information as detailed as the discussion they had, replicating all the points of view, reasons and concerns that were raised. In one example an evaluation report was written - "Notes box is too limited for normal use, requires some more space", whereas in fact the discussion had been about the fact that they were unable to carry on typing after a certain number of characters, which did not even fill the space in the box. So the problem was not really one of space but of characters within that space. This evaluation sheet was not formally raised with the developers and the user support team wrote on it that "there is more space available than can be seen", meaning that the box should scroll down. Later development work cleared this problem.

External Work Organisation Factors

Boundaries to user activity

Often clear boundaries to user activity were lacking or communicated poorly. This was mainly due to the inability to reinforce these or facilitate user activities. Such boundaries might include what issues to discuss, whether some issues were irrelevant or unalterable, and what aspects of their discussions to write down. Users on the video noted that a lot of what they wrote down were questions rather than errors or bugs to be fixed, but were not sure how appropriate that was. In the interviews, some users commented that on occasions they were unclear about exactly what they were supposed to do. Boundaries were clearer on occasions when user support, HCI personnel or developers were involved in user discussions.

Contact with other groups

There was very little direct contact with developers even though they were only a few desks away. At interview, users said they felt as if it was almost a "them and us" situation between developers and users. Direct contact with user support or HCI experts was much more frequent, but, the HCI experts were usually in a different building and the user support team were often busy preparing materials or other work in a separate room. In terms of written communication with developers, taking one day as an example, 24 error reports were completed and were followed up 16 days later. 42% of these were cleared within a week but still required retesting 9 days after that, and 21% were still waiting to be dealt with. 37% were filtered out by the user support team and not formally raised with developers, because they were either duplicates of other issues raised or the problem could not be replicated, which may have been due to forms being outdated following the latest development work. The development and evaluation cycles were out of synchrony, which resulted in efficiency and wasted effort.

System breakdowns

Sometimes there were system breakdowns which affected what work could be completed. Off-line work was usually done at these times. On other occasions it was just specific screens that did not work, and so certain work based scenarios could not be used or help screens could not be designed etc.

Work provision

During the interviews, users complained that there were occasions when they had to wait a long time to get work to do. This was usually because of system breakdowns or temporary faults which meant that pre-arranged tasks could not be completed, and the user support team had to devise something else for users to do. Users sometimes also felt that there was a lot of repetition in their work such as being told to go back over the errors they wrote down yesterday and decide which ones to raise officially.

It is evident that the issues of knowledge requirements (technical and non-technical), and work organisation (user organised and external) are interconnected. The work organisation aspects such as lack of direct contact with other groups had a great effect on the knowledge, or cognitive resources available. For instance, lack of contact

with developers meant that sometimes items were programmed which did not make sense to users, such as meaningless abbreviations. It also meant that there was a need to have someone writing communications to developers, which was a problematic part of the users' work organisation. Written information is not as effective as spoken communication for relaying ideas. Users also did not know why certain features had been programmed in by developers (i.e. who requested them and why). Greater knowledge of how these aspects might relate to future work practices would have been useful here. The technical knowledge that users required was not well supported, but some of the non-technical aspects also required more support. As a result of the users' knowledge being incomplete in these areas, they were making assumptions and sometimes disagreed about how important issues were. Decisions were being made on limited information and users were writing down things they did not have to, or things that had already been raised. Due to a lack of clear boundaries to users' work, some aspects of their work organisation were less effective, in terms of not knowing what to do or what issues to discuss and make decisions on. User controlled work organisation, however, did enable users to gain some of the knowledge they required. For example, by using the artefacts to experiment on, they were able to gain some technical information about what the system could actually do. The self organising style of the users work was also effective for handling some of the uncertainties intrinsic in their tasks, such as who has what expertise, and how users would typically approach the system and make decisions about it.

The above analysis has highlighted some of the negative consequences of not supporting information needs or work organisation needs. However, it must be noted that there were circumstances where these aspects were managed successfully. The users' evaluation work was most efficient when they had all the information or if the problem was so understandable on AS that they did not require further information. For instance there was one area where the users agreed that an entry field for wages should have a £ sign prefixing it, as they had automatically tried inputting a "£" in the field and the system would not accept it. This was dealt with quite easily and promptly by the developers, once they received the evaluation form. There were also occasions when users had direct access to a developer, HCI expert or one of the user support team, and could clarify whether something had been programmed in, or could find out the boundaries to what they were able to do - e.g. "you can change that label but it can only be 8 characters long" etc. This worked for the benefit of both sides. For instance, on one occasion, a developer could not understand a particular error that one of the users had written down, and so came over to ask him about it. The user was able to clarify the issue for the developer who was then able to correct the problem quickly.

We did not have the opportunity to evaluate the final system, but anecdotal evidence suggests that AS has been successful, as users in local offices have been able to use it and like it. This may owe more to the intensive user involvement that occurred before contact was reduced with developers, than with the type of involvement described in this paper. However, the user involvement described here is perhaps more typical than cooperative prototyping and is doubtless better than not involving users at all, but was inefficient and in need of improvement. Nevertheless, this example has demonstrated the diversity of user activity even within a relatively restricted role. Video recording was an extremely useful medium for the purpose of gaining detailed information on users' activities, particularly when supplemented with documentation which highlights other process issues, and interviews about users experiences of system development.

DISCUSSION

As can be seen, user activities were varied as were the areas of knowledge required to complete them. In addition to the non-technical information that one might expect users to bring to system development, they needed technical knowledge which was harder for them to acquire. Even so, our findings show that it should not be assumed that users will even have all the non-technical information they need. For instance, they may not know everything about future work practices (e.g. the way statistics will be used). In such cases they need to be certain about potential changes, in order to evaluate that part of the system effectively. Not only will the users participating be new to the computer system itself, but to the whole process and organisation of development, and so knowledge of how to be involved in system development is also required.

Another important area of user activity which previous research has not addressed is how users organise their own work within system development. In our study we found that users were often given a task but the method of achieving it was at their own discretion, an aspect that users enjoyed. From the perspective of sociotechnical theory (Cherns, 1987), their tasks can be seen as minimally critically specified and the user group in this case could be considered a self organising unit. The effectiveness of this was largely dependent on the external, organisational aspects which users had no control over. In particular, these external aspects, such as contact with other groups, had implications for the availability of the necessary cognitive resources. As such we can see that the relationship between cognition and work organisation is closely connected.

What do the findings say about the role of users? It suggests that users are concerned with technical information as well as non-technical work related aspects, and indeed need to know about both these in order to complete their evaluation of the system. This is not to say that users involved need to be technical experts as this would make them non-representative of the user community, but they do need answers to their technical questions. Evaluating a system for ease of use is a complex issue for a user. It is difficult if not impossible to separate it from issues of functionality such as "what we usually do in the office", or issues of design such as whether to add a new information box on a screen, as well as technical issues such as what it is possible to change. Conversations tended to cover a range of these issues and discussing one problem tended to

lead users on to another. As such it is difficult, and would be constraining, to try and tightly structure the flow of users' work.

The findings also suggest that users have a role in organising their own work within the user participation process, which is an effective strategy for this type of work. Such a role might include organising the division of labour, defining how decisions should be made, deciding how to use artefacts to help their decisions and how to communicate with those who can take action on their requests. Organising their own work needs to be supported with the necessary information resources and be within defined boundaries which specify what they can take action on and which matters are important.

Instead of trying to narrowly define the role of the user as just acceptance or usability testing, it seems sensible to try and support the activities that users would ordinarily do when trying to evaluate a computer system that will support their job, and ensure that they have the necessary resources to do it. The problem should be reinterpreted from "how can we get users to stick to just usability testing", to " how can we support what users do when they evaluate the system?".

CONCLUSIONS

Our findings have implications for the conduct of user participation, as well as the continuing debate regarding the most appropriate way in which to conceptualise the role of users within systems development. Three main points arise:

- Users' activities when involved in system development are not tightly structured and can be diverse, including input to organising their own activities within development;
- The knowledge or information users require to complete their activities is also diverse, including technical as well as non-technical aspects;
- The knowledge requirements, can be supported through the appropriate work organisation of development teams.

One of the main problems arising from the externally imposed work organisation was the incomplete information available. This could be addressed through greater integration. There needs to be some facilitation of user discussions, preferably with someone who has technical knowledge and someone who has non-technical knowledge. At the very least such information resources need to be close at hand and easily accessible. In uncertain situations, such as development projects, direct, informal communication is critical. Preferably, to keep the development and evaluation processes in synchrony, the technical problems should be dealt with straight away, or as close to the time that users point them out as possible. Otherwise information should be available about what is being dealt with and what is not and why etc.

Users should have control over how they carry out the specifics of their work, to ensure that they approach the system in the way they feel comfortable with and also to keep them involved and motivated in the process of user participation.

Finally, we recommend that organisations and developers think carefully about how to provide support for the different types of knowledge which are manifest in the kind of project we have described. Building a new system is heavily knowledge-intensive and involves pooling various knowledge sources so that they can be exploited. Our study demonstrates the importance of these processes. It is not just a case of accessing information from users, as users require information from others in the process as well. Future research should further elucidate the scope and variety of user knowledge. We suggest that the role of the user, and the dynamics of user involvement in 'live' projects are re-examined in a similar way to which other central concepts in HCI (Draper, 1993; Grudin, 1993) have been recently. There is clearly more to being a 'user' than initially meets the eye. The challenge is how best to take advantage of this, and at the same time, inform the practice of user involvement in systems development.

REFERENCES

Axtell, C.M., Waterson, P.E. and Clegg, C.W. (1995) User participation in system development: the importance of organisational factors. In S. Robertson (ed) *Contemporary Ergonomics 1995*, London: Taylor and Francis, 31-36.

Brun-Cottan, F. (1990/91) Talk in the workplace: occupational relevance, *Research on language and social interaction, 24*, 227-295

Brun-Cottan, F. and Wall, P. (1995) Using video to represent the user, *Communications of the ACM, 38*, 5, 61-71.

Cherns, A.B. (1987) The principles of socio-technical design revisited. *Human Relations, 29*, 783-792

Curtis, B. Krasner, H. and Iscoe, N. (1988) A field study of the software design process for large systems. *Communications of the ACM, 31*, 11, 1268-1287.

Draper, S.W. (1993) The notion of task in HCI. In *Adjunct Proceedings of INTERCHI '93*, Amsterdam: ACM Press.

Eason, K. D. (1995) User centred design: for users by users? *Ergonomics, 38*, 8, 1667-1673.

Grønbæk, K., Grudin, J., Bødker, S. and Bannon, L. (1993) Achieving cooperative system design: shifting from a product to a process focus. In D., Schuler and A. Namioka (eds) *Participatory design: principles and practices*. Hillsdale, N.J.: Lawrence Erlbaum Associates. 79-97.

Grudin, J. (1993) Interface: An evolving concept. *Communications of the ACM, 36*, 4, 109-119 .

Coordination training for distributed worksystems in Emergency Management

John Dowell
Centre for HCI Design
City University
London EC1V 0HB
+44 171 171 477 8418
jdowell@city.ac.uk

Walter Smith
Centre for HCI Design
City University
London EC1V 0HB
+44 171 477 8427
walter@city.ac.uk

ABSTRACT
Coordination is essential to the rescue operation which follows a major catastrophe or disaster. We summarise an agenda of training needs for the coordination of distributed multi-agent worksystems. We apply the agenda to assess the current use of physical 'tabletop' models in training coordination for emergency management. We use the agenda again to comparatively assess two computer-based simulators also intended for emergency management training. The assessments provide an initial contribution towards formulating the cognitive engineering design problem of coordination training.

Keywords
emergency management, coordination, training needs, simulation.

THE ISSUE OF COORDINATION IN EMERGENCY MANAGEMENT WORKSYSTEMS

Catastrophic accidents and natural disasters require many professionals to work together if the emergency rescue operation is to be effective. Coordination between different agencies is especially difficult when, as is often the case, those agencies are faced with severe uncertainties, hazards and time limitations.

Consider the example of the fire at the Kings Cross underground station in London, UK. Without a reliable assessment of the hazards, and with poor communications, the Fire Service were under pressure to decide on an evacuation route. A tactical level decision on the best evacuation route would have required coordination with the railway company, to establish the available exits and to control the movement of trains. In the event, some people were evacuated by the trains which, oblivious to the fire, continued to run as normal. Other people were led towards station exits through the booking hall. However the movement of the trains fanned the fire, and contributed to the fatal 'flashover' into the booking hall. The public inquiry (Fennell, 1988) was later to comment on the lack of coordination between agencies .

Coordination is now a well recognised training need in Europe and the US (der Heide, 1989). In the UK, the emergency services have established a regime of training practices including seminars and 'live' simulation exercises. Increasingly, tabletop simulation exercises are being used with the principal aim of providing training in coordination. Tabletop exercises use a physical model of a catastrophe (displayed on the top of a large table) to promote discussion about the organisation and implementation of the emergency operation. The trainees may include agents from each of the agencies that would be involved in the real emergency operation.

Tabletop exercises are an evolutionary development in training technology. They were developed within the Emergency Management community through the appropriation and modification of techniques used by military trainers. An alternative approach, which might be described as technology-led, is now making its appearance in the form of computer-based simulations of emergency management tasks. This new technology is accompanied by promises of benefits for coordination training. Yet neither the evolutionary nor the technology-led developments seek to specify the training agenda to be implemented in the cognitive behaviours of the worksystem. As a consequence, they are an insufficient basis for the design of effective coordination training. By contrast, a Cognitive Engineering approach would specify solutions to the cognitive design problem of coordination training for emergency management; it would then implement those specifications in an appropriate form of technology and practice (Long and Dowell, 1996).

This paper aims to contribute towards such a Cognitive Engineering approach. It first summarises an agenda of coordination training needs for Emergency Management. An account is then given of how the agenda can be used to evaluate the physical simulation provided by a tabletop exercise for training tactical level personnel from the emergency services. Finally, the same agenda is used in the comparative evaluation of two computer-based training systems. The comparison indicates desirable cognitive properties of both the physical and computer-based systems (see Table 1 for a summary) as a contribution to formulating the cognitive design problem of coordination training.

AN AGENDA OF TRAINING NEEDS FOR COORDINATION

An agenda of training needs has been identified (Dowell, 1995) through a cognitive analysis of coordination failures in emergency management. The agenda identifies specific issues related to the coordination of planning, action, communications and knowledge, each of which must be addressed by training.

Training needs for coordinated plans

The contingency plans of each agency participating in an emergency operation can only be high level, and conditional and must be interpreted for the particularities of each emergency. If this plan interpretation is conducted implic-

itly, coordination will suffer, leading to priorities which are duplicated, in conflict, or which neglect important concerns. Because plan interpretation is conducted under "high temporal constraints" (Samurçay and Rogalski, 1993) and must be 'reactive' to domain events (the "metronome" of coordinated behaviour, Hutchins, 1990), monolithic, decompositional planning must give way to a more opportunistic but fragmented form of planning.

Coordination requires that the contingency plans of each agency are explicitly interpreted, and that conflicts can be recognised and resolved. It requires that the priorities of all agents should conform with a unitary goal structure for the emergency operation. It also requires opportunistic planning to be made explicit.

Training needs for coordinated action

Rather than deriving from a 'master plan', the sequencing of actions across agencies is determined by mutual constraints or "behavioural dependencies" (Hutchins, 1990). Action then becomes reactive to the effects of other agents' actions. Coordination relies on the recognition and negotiation of behavioural dependencies between agents, and it requires that the actions of each agent, and their effects, are visible to those who will need to act on them.

Training needs for coordinated communications

Emergency agencies commonly report that they are "only as good as our information." Coordination requires authoritative, validated and accurate information to be transmitted as rapidly as possible in an emergency operation. It requires an understanding of how others will respond to messages and actions. Coordination also requires the use of a common language for the communications between agencies.

Training needs for coordinated role and task knowledge

Coordination requires agencies and their agents to have redundant knowledge and skills. This redundancy enables differences in the division of labour to be negotiated. Furthermore, the cognitive horizon (Hutchins, 1994) which delimits the uniqueness and redundancy of an agent's knowledge allows recognition of which agents need extra support: if coordination means giving extra support where it is needed, then cognitive horizons must be sufficiently broad.

These issues for coordination might be considered as an agenda of training needs for emergency management. It is an agenda which might be used to assess both the effectiveness of current training practices, and the design of future training practices. The following section uses the agenda to evaluate the effectiveness of coordination training provided by the tabletop exercise.

EVALUATION OF TABLETOP TRAINING

Tabletop exercises are an established form of coordination training, using a physical model and a disaster scenario to promote discussion about the emergency operation. The trainees represent the agencies that would be involved in the real emergency operation. The following is an account of a tabletop exercise concerning a railway accident scenario and observed by the authors.

Participating in the tabletop exercise were tactical level officers from the emergency services and managers of equivalent level from associated organisations including hospitals, the railway company, and the local municipal authorities. In all, some thirty trainees were attending the training exercise which was orchestrated by a 'Moderator' and his assistant. The exercise was directed towards a model of a section of urban railway line. To begin, the Moderator summarised key features of the scenario, such as schools and roads and answered questions from the trainees.

The Moderator then delivered a narration of events to begin the exercise. The narration described how the Police receive a report from a member of the public of a "loud bang" on the railway. The moderator then asked trainees, grouped into syndicates, to provide answers to the *first question*, concerning what would be the initial response of each agency, who would be notified, and how long it would take to arrive at the scene. A debriefing session allowed discussion of responses to the first question.

The emergency scenario was now extended and miniatures were added to the model representing a collision between a passenger train and a tanker train. Trainees from the emergency services were then invited to place on the scenario, models of their respective units representing their current status. The *second question* then asked trainees to consider what would be the priorities of the first persons on the scene from each agency. The trainees again withdrew into syndicates to discuss their answers, and then reconvened for a debriefing of the second question. The tabletop exercise continued in this way; the scenario was extended (the tanker train begins leaking a toxic and volatile liquid), and further questions were posed by the Moderator. The exercise was closed after the debriefing of the final question.

This example of a tabletop exercise can be evaluated in terms of its support for the agenda of coordination training needs identified earlier. The evaluation addresses each of the areas of coordination issue concerning planning, action, communication and knowledge.

Training coordinated planning

The basis of the individual training that preceded the tabletop exercise was the contingency plans of the respective agencies. Those contingency plans were implicitly referenced during the exercise. for example, in the response of a Police Officer to the second question:

> "Our first priority is to scale the incident, and to identify the residual hazards and the casualties. We would contact Scotland Yard and declare it a major incident. We would likely then evacuate those two schools (and) set up a rendezvous point for emergency vehicles..."

The tabletop exercise disclosed the varying interpretation of implicitly referenced plans. For example, whilst the first police officer appears not to interpret the primary residual hazard to be other train traffic, a second police officer does make this interpretation correctly, although the apparent absence of procedure is cause for concern:

> "Hopefully the staff at the station ... should already be aware of what's happened and have stopped the trains to prevent any further disaster."

Conflicting views about priorities were disclosed in the interpretation of contingency plans. In contrast to the

responses of the Police officers, a Fire Services officer correctly identifies the priority of obtaining electrical isolation of the crash, as necessary for the safety of the victims and rescuers.

"Our first priority is to get the line shut down, to protect people coming off the train and the services going in. By this time, you'd probably have shocked and confused people streaming all over the track."

So, through discussion, the tabletop exercise helped all the trainees to recognise common priorities in managing the emergency operation. Tabletop exercises, it would seem, provide training in resolving conflicts between different agents' and agencies' priorities that may often produce coordination difficulties. The tabletop exercise followed a 'storyboard' sequence of events; however it did not have a temporal dimension, merely a direction. Events and their responses were not constrained by any timescale, and it seems implausible that explicit opportunistic planning could be learnt from the tabletop exercise.

Training coordinated action

The tabletop exercise disclosed the dependencies between agents that may dictate the sequences of their planned actions. For example, a Fire Services officer and railway official identify a mutual dependency for confirmation of the electrical isolation of the track before they can proceed with the rescue operation:

Fire Services officer: "But how will we know if the electricity is off? We'll be frazzled if we're not told!"
Railway official: "Well it's our duty to get the lines switched off - to block other trains."
Moderator: "So the priority has to be to stop the trains, then get the power off? Is this the overriding priority?"
Railway official: "Yes - but in a crash like this, the current would almost certainly be off already ... because of short circuits caused by the wreckage."

In this way, tabletop exercises appear to train recognition and negotiation of the dependencies between agents dictating sequences of actions, as necessary for coordination. Yet the process of discussion did not, and could not simulate the visible execution of plans. Consider the later reaction of two other participants to the decision of the first Police officer to establish a traffic exclusion zone around St Michael's church without considering access to the accident:

Fire Services officer: "We're all round the wrong side of the train at the moment. We need to be on the other side - 'cause access will be easier up the grassy bank ..."
Ambulance officer: "There could be 750 injured people in those 8 carriages so I'd have called a major incident by now. We would also get up a rendezvous and a control point on the other

Table 1. Comparative evaluation of three alternative worksystems for coordination training in emergency management

COORDINATION TRAINING NEEDS	Tabletop model	Icarrus	Vistrain
	Multiple trainees with scripted physical simulation of domain	Single trainee with interactive computer-based simulation of domain scenario and coordinating agents	Multiple trainees with scripted computer-based simulation of domain scenario
planning			
explicit interpretation of plans	√	√	√
recognition of conflict	√	x	√
resolution of conflict	√	x	√
unitary goal structure	√	x	√
explicit opportunistic planning	x	√	x
action			
recognition of behavioural dependencies	√	x	√
negotiation of behavioural dependencies	√	x	√
making actions and effects visible	x	x	x
communication			
authoritative and accurate dissemination	x	x	√
timely transmission	x	x	√
understanding others' responses	x	x	√
role and task knowledge			
negotiation of division of labour	x	x	√
extended cognitive horizons	√	x	x

side (of the train)."

Although it is now apparent that access to the accident should be from the other side of the embankment, by this time, the rescue operation would already have attempted to scale the wrong side of the embankment. It would appear that tabletop exercises are unable to train the reactivity of actions that supports coordination.

Training coordinated communication

The tabletop exercise exposed all trainees to the same information at the same time; it did not simulate the progressive dissemination of information that occurs during emergency operations. That information was generally authoritative, since it was provided by the Moderator, and it was generally validated, since it was apparent in the model. The information was also accurate, indeed, at one point a railway official is able to obtain an immediate clarification of facts:

> *Railway official*: "Well our first person on the scene is of course the driver; and assuming he's OK...?"
> *Moderator*: "- He's dead ... and there's no guard"
> *Railway official*: "Ah ... he's dead ... and there's no guard, ... right, so it's not a dangerous goods train, then. So its down to the driver, then, of the tanker train, to inform the signal box and then to put detonators on the track. mmh"

Although the initial report of a "loud bang" better simulates the uncertainty of communications, it would appear that tabletop exercises are unable to recreate the conditions of information dissemination required for coordination training. Because the tabletop exercise did not simulate the 'channels of information' and their contents that occur during a real emergency operation, it could neither recreate the asynchrony of communications; all trainees could address all others directly and immediately. More generally then, tabletop exercises appear unable to train the rapid and timely dissemination needed for coordination, and it is highly questionable as to whether they provide understanding of how others will respond to the information.

Training coordinated knowledge

There was no evidence in the tabletop exercise of differences between the nominal and real divisions of labour, and for two apparent reasons. First, the scenario did not present the exigencies of a catastrophe which might force a difference between the nominal and real divisions of labour. Second, each trainee was playing the role of their entire agency, rather than the role of a single agent. Without evidence of the division of labour being negotiated, it may only be concluded that the tabletop exercise did not provide training in its configuration, as necessary for coordination.

However, if the real division of labour requires a partial redundancy in knowledge and skills, that redundancy could be learnt through the discussion of individual responses. For example, all the trainees are exposed to the conceptual distinction between "stopping the trains" (i.e., via signalling) and "getting the power turned off" (i.e., protecting the victims of the crash and their rescuers from electrocution). This distinction was not apparent in the responses of the Fire Service and Police officers. Those other agencies also appear to learn about the issues confronting the railway personnel in obtaining track isolation - and they learn that the wreckage will almost certainly have already short-circuited the current. Equally, the discussion would extend the cognitive horizon of each trainee. More generally then, tabletop exercises appear to provide the redundant knowledge needed for coordinating the configuration of roles, for example, in recognising which agents need extra support.

Summary of evaluation

In summary, the tabletop exercise provides valuable, but incomplete training in coordination. Trainees learn collectively to interpret contingency plans, to resolve conflicts between different priorities, and to recognise and negotiate mutual constraints. They may also extend their cognitive horizons, as necessary for mutual support and configuring the real division of labour. However, the exercise does not provide training in the reactivity of planning to events, or in the reactivity of actions to the actions of others. It can provide no training in planning and decision-making under temporal constraints. Neither does the tabletop exercise appear to provide training in the dissemination of information, nor in the asynchrony of communications.

COMPUTER-BASED TRAINING SIMULATORS FOR EMERGENCY MANAGEMENT

Advances in computer-based simulation technology, in particular, in agent-based intelligent systems and in dynamic rendering techniques, are supporting the development of training systems for Emergency Management. Those systems simulate physical and conceptual features of the emergency management domain and worksystem which were previously impossible. Two very different examples are the Iccarus and the Vistrain systems. An account of those systems is given below together with their assessment against the training needs agenda.

Iccarus

Iccarus is a simulator for training Fire Service officers in managing large scale fires (Workhouse, 1992). It consists of a set of modules for creating and running an event-based scenario (typically, a fire at a cinema). The simulator trains Fire officers individually, and a training session runs as follows.

To begin, the simulator provides the trainee with a spoken introduction to the catastrophe scenario, explaining that they are currently driving towards a fire. The simulator video monitor presents a sequence of photographic images of a route though an urban area, during which can be heard various recorded radio messages concerning the incident and simulating the radio traffic being relayed to the fire control centre.

The scenario soon describes the trainee as entering the mobile Forward Command Post at the scene of the emergency operation. The trainee must now take over as Officer-in-Charge of the operation. They are briefed by the officer standing down using the standard protocol: "What have I got? What messages have you sent? "What have you done?" The part of the officer standing down is played by the simulator; the dialogue is conducted by the trainee interrogating the officer through the simulator's graphical user interface, textual responses being displayed on the computer monitor. The monitor provides a strategist's representation of the emergency operation: a plan of the

building and the current state of the fire; the current deployment of 'resources' and those available for deployment. This representation re-creates paper-based planning aids typically used in managing emergency operations. It is a view of the computational domain model of the emergency operation as that model is executed by the simulator's transputer.

The domain model is an event-driven, dynamic model simulating the progress of the fire through the building and the location and properties of objects such as hazardous materials; in addition, it simulates the location and behaviours of firemen and agents of other emergency agencies and of members of the public trapped in the building, and it models the location and behaviours of equipment such as pumps. In their role as Officer-in-Charge, the trainee is able to manage the emergency operation by intervening in the state of the domain model, for example, by requesting more pumps and officers, moving pumps and directing specialist teams, etc. These actions are taken by issuing commands, for example, to deploy a breathing apparatus team, the trainee uses the mouse to select the icon 'BA Deploy' and an entrance to the building. The domain model responds to these instructions, for example, the breathing apparatus team are able to act with semi-intelligence to enter the building, locate trapped people or fight the fire. So, the effects of the trainee's instructions, the progress of the fire, and the behaviours of up to 200 'automaton' firemen and other agents are each visible to the trainee.

Photographic stills and video film are stored on videodisc within the simulator. The simulator also provides an acoustic environment of voices and noises typically heard at 'firegrounds'. As events unfold within the executing domain model, the simulator selects appropriate visual images and sounds, for example, distraught people stating that someone is trapped in the building, or other agencies demanding information. The intended effect is to bombard the trainee and induce them to 'suspend disbelief' (Hennessey and O'Shea, 1993). The training session finishes when the fire is controlled, or the building lost. All the scenario events and the trainee's actions are recorded and the record used to debrief the trainee. The debrief addresses at each point whether the trainee has an overview, whether they are deploying resources effectively, whether communications are effective, and whether the trainee is being reactive to events.

As apparent from its debriefing structure, Iccarus is intended to provide comprehensive training in emergency management, including training in coordination issues. With reference to the agenda of coordination training needs, it is apparent that Iccarus might in principle enable the trainee to learn how to plan by experience. It is able to do this because its domain model is dynamic, it represents the time course of events of the emergency operation, and it is responsive to the decisions taken by the trainee. Further, the trainee might learn to plan opportunistically and reactively to domain events. However because all the individuals with whom the trainee interacts are reduced to automata, the trainee cannot learn to recognise or negotiate plan conflicts, nor to coordinate actions, communications and role and task knowledge.

VISTRAIN

Vistrain (Video-based integrated system for training applications) is a multi-user training simulator intended for general application to management training (LTU, 1992) which has also been applied to training Police officers in the management of emergencies. For example, one fictitious scenario relates to a collision between a heavy truck and a train. The system provides team-based training to support tactical level decision-making through the negotiation of plans. Teams are separated, their separations being task and role determined. For example, Vistrain is intended to simulate the activities of a forward control point and the strategic-level command centre at a major incident. Within their different roles and locations, teams have access to different information. For example, the forward control point team might have access to the current status of trapped individuals in an emergency scenario, whilst the command centre would have access to information on spare capacities in the regional hospitals. That information is represented in different media, including simulated news broadcasts and video clips. The different teams are exposed to cues to which they are expected to respond. The cues include video images, and radio and telephone messages. The teams must decide on their responses to these cues and communicate their decisions to other teams. Scenarios are run from a central workstation by a Moderator who is able to call different events and responses to the trainees' actions. However the scenarios are pre-authored and stored on multiplexed videodisc, so the variety of responses and scenario developments is closely limited.

The Moderator monitors the activity of the different teams through shadowing the communications between teams and the information displays to which the teams are responding. The Moderator's workstation provides a storyboard representation of the appropriate activities of each of the teams in response to specific events. This storyboard can be used to record and mark the behaviours of the team for later debriefing. Trainees using Vistrain reported through questionnaires that the system provides a more demanding and realistic training experience than the tabletop exercise (LTU, 1992). However, to assess the training value of Vistrain with respect to coordination, we can refer to the agenda of training needs.

For the training needs of coordinated planning, Vistrain can in principle support learning to plan by experience. It appears to support the explicit sharing of plans, and of the planning process, including the interpretation of contingency plans. However whilst the system possesses a dynamic domain model, that model has only a limited repertoire of responses. Trainees therefore have a limited learning experience of the need for planning to be reactive and opportunistic. Nevertheless, coordination of priorities between the plans of different agencies can be learnt with the system which enables the negotiation of those priorities within a unitary goal structure. With respect to the training needs of coordinated action, the communication between distributed agents and agencies supported by Vistrain allows recognition of behavioural dependencies, and it allows rehearsal of the process of negotiation. With respect to the training needs of coordinated communications, Vistrain appears to support rehearsal of the dissemi-

nation of authoritative, validated and accurate information. It is able to do this because it simulates the properties and contents of different communication channels, making evident the difficulties that occur when information is not authoritative, valid or accurate. Finally, with respect to the coordination issues of roles and task knowledge, Vistrain may allow recognition of the negotiation of divisions of labour between agents. However the form of rehearsal-based training provided by Vistrain is less likely to promote redundant knowledge of the tasks of other agencies, since that knowledge will only be implicit in the activities of those other agencies. As a consequence, Vistrain is unlikely to extend significantly the cognitive horizons of the trainees.

CONTRIBUTIONS TO COGNITIVE ENGINEERING AND THE SPECIFICATION OF COORDINATION TRAINING

The tabletop exercise and the computer-based simulators evaluated in this paper can be seen, respectively, as exemplars of evolutionary and technology-led approaches to the training of coordination in emergency management. The evolutionary approach has the advantages of carrying over the wisdom implicit in current emergency management practice. Computer-based solutions, in comparison, offer formal properties thought to be desirable on logical grounds; for example, interaction with a 'real-time' event-driven domain, flexible scenario generation, behaviour recording and evaluation tools.

Cognitive engineering provides an alternative to both the evolutionary and technology-led approaches. In this view, the starting point must be an explicit agenda of training needs specified as those cognitive behaviours which underlie effective coordination in emergency management. Training technology is then designed with the goal of establishing these behaviours in the trainees. Such an agenda has been presented in this paper and used to evaluate existing training technologies (see Table 1).

When compared against the training needs agenda, the Iccarus and Vistrain systems can be recognised as significantly different solutions to the coordination training problem. Iccarus, the individual user system, appears to provide an enhanced learning of coordinated planning because of the reactive domain model on which the system is constructed. However the learning of coordinated action and communication is negligible, since other agents and agencies are reduced to automatoa. By contrast, Vistrain the multiuser system, appears to provide an enhanced learning of coordinated plan implementation and communication, since it constructs the real time decision-making and communications channels and dynamics of multi-agency coordination. However Vistrain's non-computational model ensures that planning cannot be reactive to domain events, and so learning to coordinate planning is weakened.

A contrast can also be made between the two computer-based systems and the tabletop exercise. Neither Iccarus nor Vistrain appeared to train the vital coordination of roles and knowledge apparent in the tabletop exercise. Indeed, they appear to support a qualitatively different training experience to the one provided by the tabletop exercise: a difference often characterised rhetorically as a move away from 'learning by talking' towards a form of 'learning by doing'.

To conclude, we have used an agenda of needs for coordination training to assess current practices which use the tabletop exercise. We used the same agenda to assess two different forms of computer-based training system. Qualitative and important differences between those training systems and practices are evident. It is similarly evident that evolutionary and technology-led developments will not ensure the most effective training system. We advocate a cognitive engineering approach to specify the training needs and the training system to meet those needs. The agenda of training needs, and the assessment of the tabletop exercise are intended to contribute to that process.

ACKNOWLEDGEMENT

The authors, together with colleagues at University College London and Bangor University, are funded by the Economic and Social Research Council, UK to specify requirements for coordination training in Emergency Management. The project is incorporated in the research council initiative in Cognitive Engineering.

REFERENCES

Auf der Heide, E. (1989) *Disaster Response: principles of preparation and coordination* St Louis: Mosby.

Dowell, J. (1995) Coordination in the management of emergencies and the tabletop training exercise *Le Travail Humain* 58, 1, 85-102.

Fennell, D. (1988) *Investigation into the King's Cross Underground Fire* London: Department of Transport (HMSO).

Hennessey, S., and O'Shea, T. (1993) Learner perceptions of realism and magic in computer simulations *British Journal of Education Technology* 24, 2, 125 - 138.

Hutchins, E. (1990) The technology of team navigation. In Galegher, J., Kraut, R. and Egido C., (ed.s) *Intellectual Teamwork* Hillsdale NJ: Lawrence Erlbaum.

Hutchins, E. (1994) *Cognition in the Wild* Mass: MIT press.

Long, J., and Dowell, J. (1996) Cognitive Engineering or getting users and computers interacting to perform work effectively The Psychologist (accepted for publication)

LTU (Learning Technologies Unit), (1992) *Report of the development of the control of crowds at major spectator events* Sheffield: Dept for Education and Employment.

Samurçay, R. and Rogalski, J. (1993) Cooperative work and decision making in emergency management *Le Travail Humain* 56, 1, 53-77.

Workhouse (1992) *ICCARUS.* Sheffield: Dept for Education and Employment.

Reshaping Task Contents: Operators' Contribution To Systems Reliability

E. Morvan, L. Rognin
LIS - LAAS / CNRS
7 av du Colonel Roche
31077 Toulouse, France
+33 61 33 63 14 / 64 23
{morvan, rognin}@laas.fr

J.-C. Sperandio
Laboratoire d'Ergonomie Informatique
45, rue des Saint-Pères
75270 Paris Cedex 06, France
+33 1 42 86 20 74
sperandio@ergo-info.univ-paris5.fr

ABSTRACT
Socio-technical reliability has long been considered as mainly dependent on technical reliability. Yet, it now appears that human components and interactions between human and technical components have to be focused on. The analysis of real-work performance (activity analysis) allows the identification of human mechanisms supporting the socio-technical reliability. This paper presents some results issued from the analysis of two working situations: (1) maintenance of technical systems in Air Traffic Control and (2) control of a nuclear reactor. The emergence of unexpected activities such as collective (instead of individual) problem-solving and implicit coordination through information sharing is discussed. Human agents reshape their task contents in order to efficiently handle the situation, thus mitigating some technical, human or organizational deficiencies.

Keywords
complex systems, error tolerance, human cooperation, human reliability, regulation mechanisms

INTRODUCTION

Nowadays, technical progress removes most of the technical failures, yet without eliminating them completely. Despite the efforts taken these 20 last years, in order to diminish human errors (with automation, technical assistant systems and training), the situation remains about the same (Amalberti, 1992). Even so, many complex systems seem to be reliable on the whole. This can be explained by error tolerance due to regulative mechanisms. Such mechanisms must be fully analyzed, so as to be appropriately supported.

In the research area, a growing number of studies on human reliability addressed recently the human agent in positive terms. Improving the reliability of complex systems may now require us to cease considering the human agent as a weakest point but rather to identify how this agent is able to contribute actively to the global reliability.

Task analysis proves to be fruitful to point out how human operators contribute to system reliability.

This paper is composed of three main sections. In the first one we discuss the human contribution to systems reliability. In the second section, we describe two working situations, highlighting how the human agents reshape their tasks contents, in order to cope efficiently with the task demands and reduce the consequences of errors. In the last section, we discuss the supports for task contents reshaping and the perspectives of this study.

SYSTEM RELIABILITY: THE ESSENTIAL ROLE OF HUMAN FACTORS

As technical systems are becoming more and more reliable and efficient, human components are generally considered as the weakest point of the system. Due to their limited cognitive and physical capabilities, they appear as the main source of error under certain circumstances (high pressure, time constraints, night work).

Paradoxically, at the same time, it appears that the more advanced a system is, the more crucial is the role of the human operator (Bainbridge, 1987). The automation leads to the design of new systems managing normal operations, but often unable to handle unexpected problems. In such situations, the human agent is in charge of ensuring the safety of the process, being responsible for the diagnosis and the decision-making. Yet, moving from an active and central position (control) to a distant one (monitoring), the human agent is nevertheless expected to understand immediately the situation and choose the most appropriate solution. Such conditions put the human agent in trouble, going against the human role in the systems reliability.

Recent studies address the problem of socio-technical reliability by attempting to apply some basic concepts developed within the framework of reliability (Hollnagel, 1992; Laprie, et al., 1995; Mazet & Guillermain, 1996). In order to ensure system reliability, two approaches must be combined: prevention and tolerance. The first one requires not only identifying the causes of errors, but also finding means to prevent them. The second one rests on the idea that errors are unavoidable events; so the stress must be laid on mitigating their consequences.

Indeed, according to statistical studies, complex systems seem to be reliable on the whole (Amalberti, 1995), in the sense that the delivered service does not deviate from fulfilling the expected function (Laprie, 1992). Our point of view is that in order to balance deficiencies or recover from errors, some regulations are performed within the system by human agents.

In order to obtain a general and representative understanding of positive human contributions to reliability, it seems efficient to analyse usual working activities. Based

on findings from aeronautical field studies, Amalberti (1992) addresses operator's strategies to ensure safety. Bressolle et al. (1996) highlight the continuous presence of incidents during normal periods. The human agents detect and recover co-operators errors, using mechanisms such as intention recognition, objectives assessment, implicit control, environmental resources sharing (Pavard, 1994; Salembier, 1995).

The purpose of the following field studies is also to highlight the operators' contribution to systems reliability. In the next section, we describe usual working activities, underlining how human operators manage to prevent or tolerate technical, human and organizational errors, mainly through the reshaping of their task contents.

Pointing out the Positive Contribution of the Human through Task Contents Analysis

Distinguishing what the operator is supposed to do (task description) from what is actually done (task contents analysis or activity analysis) is a very classic and fruitful method for field studies (Leplat, 1981; Sperandio, 1991), especially when it is undertaken in the perspective of designing or modifying working environments.

This approach has been followed while studying two working situations of systems maintenance and control. During the last two years, an ergonomic analysis was performed through an iterative process, consisting in observations and recording of real activities, data analysis and interviews of the agents.

Maintenance of Technical Systems in Air Traffic Control

General Presentation of the Work Situation

The first study concerns the activities of maintenance operators in an Air Traffic Control Center. The operators are specialized in one particular system (here the CAUTRA[1] system whose main function is to provide the controllers with processed data issued from radars and flight plans). The organization distinguishes clearly two different functions of maintenance. *Operational Maintenance* (OM) is performed by a single operator 24 hours a day, 7 days a week and aims at restoring the system in case of failure. *Specialized Maintenance* (SM) is carried out during office hours and consists mostly in installing and testing new systems, in configuring data, and in managing repairs of faulty components. A team of nine technicians alternate the two functions (three weeks of OM, six weeks of SM).

The analysis of the situation highlights the recent profound changes in both the technical staff (departure of the experienced operators, most of the team are novices) and the technical system (more sophisticated software, higher reliability).

When a part of the system fails, the OM operator is supposed to identify the failing component and find a way to maintain or restore the service. Depending on the context (air traffic flow) and the complexity of the problem, the requirements of the tasks may vary considerably from a simple application of well known rules to a complex problem solving under uncertainty and time pressure. Despite the needs, the situation offers the operators few possibilities for learning. In addition, as the technical system is gradually replaced, more and more unfamiliar situations come out (e.g., novel faults). According to Rasmussen's three level ladder, in this situation fault-finding often requires a knowledge-based level that might be costly for the operator (Rasmussen, 1986).

At first, the main purpose of this study was to understand how the OM operator handles fault-finding in this context. Actually, the analysis of the OM activities showed the unexpected importance of collective activities. Hence we addressed fault-finding considering not only the OM operator, but also his/her interactions with environmental resources, both human and technical (Hutchins, 1995).

In the following examples, we illustrate how the operators adapt themselves collectively to the context (sharing implicitly tasks in case of a technical incident affecting the controllers activities) and to their individual lacks of knowledge (collective problem solving).

Cooperation and Task Sharing

A lot of redundancies have been introduced in the technical system in order to ensure a high level of availability. Most of the time, a failure does not spread up to the controllers' work positions, thanks to a switch (automatic or manual) to another component. However, in some cases, when the function of a crucial device cannot be restored immediately, measures might be taken by the controllers' hierarchy in order to temporarily reduce the traffic flow (freezing plane departures from local airports for example).

We present here such a situation showing instantaneous conversion of individual tasks to a collective one.

As soon as the OM operator got informed by a telephone call, he rushed out of the technical room up to the control room. Three SM operators immediately joined him for a common and quick analysis of the situation. A blocked terminal had made opening additional control work positions impossible. So the control center was unable to adapt itself to the present increase of the air traffic flow and measures had to be taken.

Implicitly, the four maintenance operators shared subtasks in order to restore the failed function. One of them went with the operational supervisor down to the technical room in order to open the required work positions. Another one restarted the device with the assistance of a third one. The last operator remained in the control room to explain the situation to the controllers and their hierarchy.

This example highlights the real-time distribution of tasks and integration of extra supports in order to speed up the recovery of a technical failure. Because of the time pressure, the emphasize was not on fault-finding but on recovery. Yet when time pressure was lower, we observed unexpected collective fault-finding as presented below.

1. "Coordinateur Automatique du Trafic Aérien"

Collective instead of Individual Fault-Finding

Although the OM operator is supposed to carry on fault-finding alone, frequently at least one SM colleague joins him/her for collective problem solving. It may start after a request from the OM operator but most of the time, the assistance is spontaneous. Verbal protocols of the interactions have been recorded in order to analyse the dynamic of the collective diagnosis process.

Indeed, we noticed that the assistance lies in sharing information likely to be helpful for the problem at hand (e.g., a SM operator says:" *...X made a file, there is a file about network problems*"). In complex systems, knowledge can be defined as partial and distributed amongst various supports (human agents, technical devices, paper supports). Firstly, as the system is also quite unsteady (progressive replacement by new technology components, extended possibilities of configurations), nobody knows everything about it. Secondly, no two operators have identical knowledge even if they have been working the same length of time. During operational maintenance activities, they have met different incidents and during specialized maintenance activities they have tested and installed different parts of the system, and written related documentation. Cooperation appears as a means to allow the integration of partial and distributed knowledge (Bannon & Schmidt, 1991; Cicourel, 1990).

Several studies underline the limited capacity of the human agent to take into account multiple possibilities (Elstein, et al., 1979) and a tendency to hypothesis confirmation rather than disconfirmation (Klayman & Ha, 1987). In collective fault-finding, a broader investigation of the situation is possible thanks to the number of alternative hypotheses and strategies that might be considered. Sometimes, the SM operator gives just ideas and goes away (e.g., *"We had something similar, it might be a plane which is not at the required level..."*). But usually the cooperation lasts quite a long time and involves discussing each other's hypotheses and sharing sub-tasks such as component testing. Yet, the decision making control remains ultimately in charge of the OM operator.

Occasionally during the diagnostic process, a co-operator detects an error in the representation of his/her colleague (e.g., an OM operator says pointing to an indicator "a": *"The indicator should flash...!"* The SM operator shows him the indicator "b" flashing below and says: *"Better look at this one".*). Detection by another person seems to be the most efficient means to identify many errors of diagnosis (Reason, 1990).

So the diagnostic process takes advantage from cooperation through the integration of partial and distributed knowledge, hypotheses comparison, error detection and recovery.

As we mentioned above, the situation offers few possibilities for training and this cooperation can be understood as a way for each operator to enrich his own experience. Reshaping operational maintenance from an individual task to a collective one offers opportunities for collective feedback and training concerning fault-finding.

Unforeseen cooperative activities are therefore considered as a means to detect and recover from technical, organizational and human errors.

The Nuclear Reactor Control

The Working Environment

In the observed situation, a team of six agents is working together in order to achieve a common goal, which is the safety of the reactor. Some of the agents are in charge of controlling the reactor (supervisor, reactor controller), the others (electrician and mechanics) are in charge of operating locally on the installation, in order to ensure the availability (readiness for use) and the reliability (continuity of the service) of these technical components.

The activities are distributed inside and outside the control room, according to the agents' objectives.

The supervisor and the reactor controller remain in the control room so that they have at their disposal centralised support of information (control panel, feedback from local agents).

The local agents are the auxiliary controller, the mechanic and the electrician. The auxiliary controller stays during the 12 hours shift in a backroom, potentially used as a back-up in case of destructed control room. The other agents leave only occasionally the control room in order to check the equipement or operate it.

The Communication Flow

Two main flows of verbal communications are observed. Face-to-face communications used within the control room are combined with telephone and intercom-mediated communications between the control room and the local area. With the intercom, the artefact allows verbal exchanges between the agents and broadcasting of information inside the control room, itself contributing to peripheral listening[2]. Inside the control room, non-verbal communications are observed, contributing to updating the situational awareness (Rognin & Pavard, 1996). These communications rest on observing not only the agents location, actions, behavior and reactions, but also the evolution of the common field of work, defined as the part of the world affected by actors' work (Schmidt, 1994).

Accessing verbal and non-verbal communications allow each agent to be able to implicitly control the others ' contributions and integrate his/her task within the collective one.

Error Detection and Recovery

In this situation (fig.1), the mechanic is sent on the plant in order to open a valve. When he leaves the control room, he is just told to *"open the lower valve"*. The supervisor who is supposed to remain behind his desk, moves to the control panel, in order to follow the evolution of parameters. There, he and the controller both notice a problem, as the

2. Peripheral listening refers to the fact that even focusing on specific actions, the agents are able to hear the communications and identify in which way the messages could concern them.

monitored flow is not evolving as expected. The supervisor first asks the mechanic if he did open the *"lower valve"*. The mechanic answers that *"yes (he opened) the lower valve"*.

Discussing in front of the control panel, the two agents formulate an hypothesis regarding the problem, suggesting that the local agent opened the wrong valve. Meanwhile, the supervisor's question leads the local agent (1) to wonder if there is any problem, and (2) to provide more information than expected, specifying exactly which valve he opened.

This input of data confirms the agents' hypothesis and allows the problem to be solved through specifying clearly which valve has to be operated.

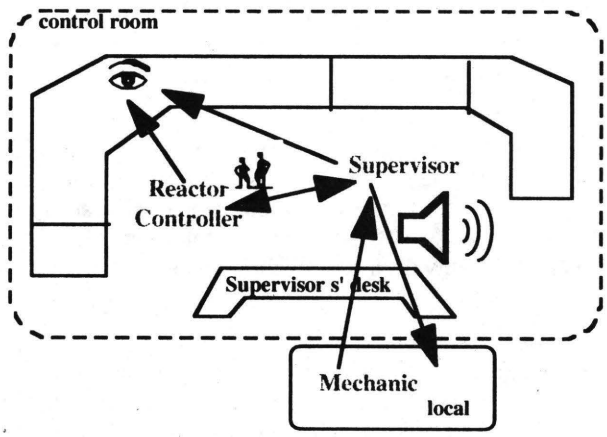


Figure 1: collective error processing

The supervisor moved to the control panel, in order to get the same view as the controller. Sharing the environmental resources allows the two agents to elaborate together an hypothesis and to use implicit communication supported by their co-location. Thus, when the supervisor watches a specific indicator, the controller is acknowledged about the relevance of this specific data.

In this situation, the supervisor adapts his task contents, extending it from the global supervision to the precise monitoring of data, moving from a distant position to a proximal one. This task reshaping aims at improving the problem solving, increasing the sources of information and of suggestions (supervisor and controller) using a shared support of information in order to elaborate a collective diagnosis.

Distributed Decision-Making

The task specification highlights the supervisor central role in the distribution of task and in the decision making process. The supervisor attributes the tasks, monitors the process and updates his own knowledge through centralising every communication (inside and outside the control room).

Yet, we noticed that the distribution of tasks is explicit in the sense that the supervisor expresses what has to be done, but it remains implicit, as the actor is not mentioned. Each agent has to identify who has to perform each task, and who the supervisor is talking to.

When an agent is operating equipment (fig.2), feedback has to be sent to the supervisor, in order to inform the control room about the operations performed and the consequences observed locally.

In this situation, the auxiliary controller is operating on a pump, under the supervisor monitoring. The reactor controller is also monitoring this intervention through reading indicators on the control panel. The local actions have no consequences (the local agent does not notice any evolution), so that the agent suspects a problem. In order to assess his hypothesis (the flow is not strong enough), the local agent suggests the supervisor to increase the flow (1). The supervisor does not answer or accept anything; yet, the reactor controller who heard the local suggestion (2) decides to increase the flow.

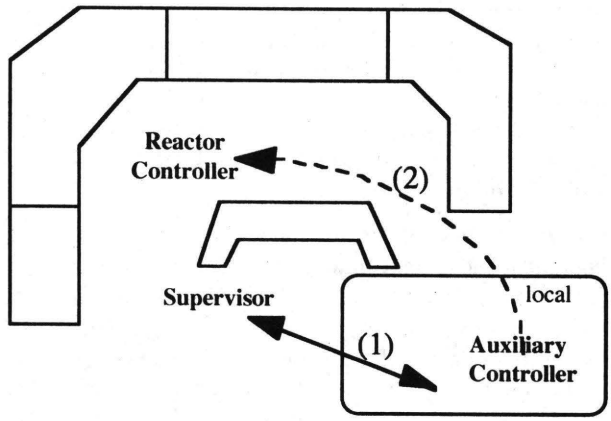


Figure 2: distributed information and implicit order

The communication between the local agent and the control room has then complementary objectives, which are not only to inform the team about local observations, but also to suggest actions or strategies. The supervisor is supposed to accept or refuse these suggestions, and then explicitly give the agents appropriate orders. Yet, we noticed that usually, the absence of externalized reaction from the supervisor is interpreted by the controller as the acceptance of local suggestion and considered as an implicit order.

The local agent who has relevant and updated information becomes at this moment the decision maker. This evolution of the task contents allows quicker and more appropriate decision to be made. Moreover, the use of implicit order reduces the cost of communication and control. This implicit mode of cooperation is supported by two main features, which are the common experience of work and the shared field of work. We discuss these points in the next part.

Reshaping Task Contents

Adaptation by the agents to the task requirements is conditioned on the one hand by the organization and on the other hand by the agents knowledge.

The working conditions imposed by the organization usually take into account neither all the reality of work nor the existence of real-time constraints. In order to deal with these unexpected features, the agents develop informal organization and cooperative mechanisms. However, this remains possible only as long as the working environment does not restrict such an adaptation, allowing flexibility and providing enough margins of autonomy (De Terssac, 1992).

In order to choose the most appropriate actions, the agents have to share knowledge about the others' activities and competencies.

For example, during overloaded periods, an electrician may help a mechanic operating locally a pump. This is possible as long as:

(1) the electrician can access the mechanic' supports of information, enabling him to identify the task to be performed and the needed means,

(2) the management accepts this new task distribution.

Likewise, at night an OM operator specialized in Radar system[3] operating may assist an OM operator specialized in CAUTRA system.

Yet, in order to do more than the prescribed task, an agent has to be to a certain extent polyvalent, knowing the others' tasks. Moreover, the agents have to trust one another so that help mechanisms would be possible.

As we shown in this section, the agents modify the contents of their tasks in order to overcome some limits in their definition. These modifications are allowed when the agents share informations regarding the current tasks. The adaptation of the tasks is made easier when each agent is able to access the others current objectives and needs. This point is supported by implicit acquisition of information, through peripheral listening to exchanged communications and observing the agents actions and locations. Thanks to this reshaping, the human agents appear to be efficient means to tolerate errors, providing external supports for errors detection, but also for errors prevention (e. g., anticipating increase of workload).

CONCLUSION

Due to the increasing complexity of systems and to the unability to prevent or tolerate all technical faults, human operators go on playing a major role in the global reliability of most critical socio-technical systems. This reliability appears to be ensured through regulative mechanisms performed between the systems components. Such regulations observed in the real-world, aim at allowing the system efficiency in spite of local failures (Bressolle, et al., 1996).

Field studies contribute to the identification of unexpected activities, performed by the agents according to the situation real-time constraints. In this paper we underlined how operators spontaneously reshape their tasks' contents, improving its efficiency and reliability (by performing collectively individual tasks, using implicit mechanisms improving the cooperation efficiency, ...). We also highlighted how such activities contribute to the global reliability. Some of these additional activities such as self-training seem to be only potentially useful for future hypothetical situation as Falzon (1994) underlined. These activities contribute to the tolerance of various errors related to the human agent himself, the technical components or the organization.

From the agent's point of view, the collective problem solving improves individual knowledge, in the sense that it increases the elaboration of individual and collective database, through information sharing.

From the designer's point of view, understanding why and how people cooperate (assessing the others' needs, detecting overloaded or stressful periods, anticipating error occurrence) has implications on the design of collective tools and training (Rogalski, 1995).

Underlining the human agents' active role in the adaptation and reshaping of the task contents aims to improve the design of fault-tolerant socio-technical systems.

ACKNOWLEDGMENTS

Part of this work was performed under contract for the Centre d'Etude de la Navigation Aérienne. We thank C. Bailleul, E. Troillard (CENA), J.-P. Blanquart for their fruitful support and T. Green for his careful reading and correction of the paper.

REFERENCES

Amalberti, R. (1992). Safety in process control: an operator-centred point of view. *Reliability Engineering and System Safety, 38* , 99-108.

Amalberti, R. (1995). *La conduite des systèmes à risques.* Paris: P.U.F.

Bainbridge, L. (1987). Ironies of automation. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human errors* (pp. 271-286). Wiley Publ.

Bannon, L.J. & Schmidt, K. (1991). CSCW : Four Characters in Search of a Context. In J.M. Bowers & S.D. Benford (Eds.), *Studies in Computer Supported Cooperative Work. Theory, Practise and Design* (pp. 3-16). Elsevier Science Publishers B.V. (North Holland).

Bressolle, M.C., Decortis, F., Pavard, B., & Salembier, P. (1996). Traitement cognitif et organisationnel des Micro-Incidents dans le domaine du contrôle aérien : analyse des boucles de régulation formelles et informelles. In G. De Terssac & E. Friedberg (Eds.), *Coopération et conception* Toulouse: Octarès.

Cicourel, A.V. (1990). The Integration of Distributed Knowledge in Collaborative Medical Diagnosis. In J. Galegher, R.E. Kraut, & C. Egido (Eds.), *Intellectual Teamwork. Social and Technological Foundations of Cooperative Work* (pp. 221-242). Hillsdale: Lawrence Erlbaum Associates.

3. The radar system feeds the CAUTRA system with data coming from the radar antennas located in the area.

De Terssac, G. (1992). *Autonomie dans le travail.* Paris: P.U.F.

Elstein, A.S., Shulman, L.S., & Sprafka, S.A. (1979). *Medical Problem Solving. An Analysis of Clinical Reasoning.* Cambridge: Harvard University Press.

Falzon, P. (1994). Les activités métafonctionnelles et leur assistance. *Le Travail Humain, 57* (1), 1-23.

Hollnagel, E. (1992). The design of fault tolerant systems: Prevention is better than cure. *Reliability Engineering and Systems Safety, 36* , 231-237.

Hutchins, E. (1995). *Cognition in the Wild.* Cambridge, Massachusetts: The M.I.T. Press.

Klayman, J. & Ha, Y.-W. (1987). Confirmation, Disconfirmation, and Information in Hypothesis Testing. *Psychological Review, 94* (2), 211-228.

Laprie, J.C. (1992). *Dependability: Basic Concepts and Terminology.* Springer-Verlag. New York.

Laprie, J.-C., Arlat, J., Blanquart, J.-P., Costes, A., Crouzet, Y., Deswarte, Y., Fabre, J.-C., Guillermain, H., Kaâniche, M., Kanoun, K., Mazet, C., Powell, D., Rabéjac, C., & Thévenod, P. (1995). *Guide de la Sûreté de Fonctionnement.* Toulouse: Cépaduès-Editions.

Leplat, J. (1981). Task analysis and activity analysis in situations of field diagnosis. In J. Rasmussen & W.B. Rouse (Eds.), *Human Detection and Diagnostic of System Failures* (pp. 289-300). New York: Plenum Press.

Mazet, C. & Guillermain, H. (1996). Dependable Systems: Error Tolerance and Man-Machine Cooperation. In *ESREL'96, International Conference on Probabilistic Safety Assessment and Management.* June 24-28, Crete, Greece.

Pavard, B. (Ed.). (1994). *Systèmes coopératifs : de la modélisation à la conception.* Toulouse: Octarès.

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: an Approach to Cognitive Engineering.* Amsterdam: Elsevier Sciences Publishers.

Reason, J. (1990). *Human Error.* Cambridge University Press.

Rogalski, J. (1995). From Real Situations to Training Situations: Conservation of Functionalities. In J.-M. Hoc, P.C. Cacciabue, & E. Hollnagel (Eds.), *Expertise and Technology. Cognition and human-computer Cooperation* (pp. 125-139). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Rognin, L. & Pavard, B. (1996). Pluri-addressed Messages and Coordination: Impact of the Intercom on the Human Cooperation. In *COOP ' 96, Second International Conference on the Design of Cooperative Systems,* June 12-14, Juan-Les-Pins, France.

Salembier, P. (1995). Intention et action. In *Séminaire de l'atelier Cognition Partagée, Prescot-Cognisciences.* Toulouse, Novembre.

Schmidt, K. (1994). *Modes and Mechanisms of Interaction in Cooperative Work. Outline of a Conceptual Framework* . Risø-R-666 (EN). Risø National Laboratory, Roskilde, Denmark.

Sperandio, J.-C. (1991). Les méthodes d'analyse du travail en psychologie ergonomique. In J.-P. Rossi (Eds.), *La recherche en Psychologie. (Domaines et méthodes)* (pp. 199-237). Paris: Bordas.

Creation and Loss of Cognitive Empathy at an Emergency Control Centre

Henrik Artman and Yvonne Wærn
Department of Communication Studies
Linköping University
S-581 83 Linköping
henar@tema.liu.se, yvowa@tema.liu.se

ABSTRACT
We have analyzed some examples of communication and cooperation in an emergency control centre. The distribution of information and the distribution of tasks is a complex cognitive activity. In some cases, this activity is regulated by institutional rules, in other cases, the persons involved have to decide when to call upon another person, how to transfer data and how to feed back data about activities performed. For time-critical complex tasks like this, strict rule-based behaviour seems to be insufficient, as much of the success of managing this dynamic task seem to depend upon flexibility and what we coin cognitive empathy.

Keywords
coordination, dynamic situation, control, cognitive empathy

PRESENTATION OF AN EMERGENCY CONTROL CENTRE
An emergency control centre in X-town was studied. The Swedish emergency services are organised so that anyone can call a specific telephone number and the telephone system connects the caller to the closest emergency centre. At the emergency control centre all incoming 911 calls are handled and registered.

Usually, three persons (during the night two only) work together in shifts, taking the calls as they come in. As there might be several calls at the same time to the emergency centre, some work distribution is called for. One person is in general responsible for keeping track of the ambulances, another one is responsible for the fire brigades as well as for the use of helicopters. During office hours a third person is responsible for some more commercial tasks, such as different security alarms. However, everybody must be able to take care of every kind of incoming call. This means that there is both a formal organization and some freedom for people to organize their work according to circumstances.

Each person sits in front of a computer. The persons are sitting in a half-circle form so that they can both see each other and have easy access to the computer (see fig 1). They have a headset which covers one ear and which allows the person to have the hands free while talking to the caller at the same time as one ear is free to listen to conversations within the control room. The computer is used to register the call, so that any commission can be followed by anybody who chooses to look at the computer.



Fig. 1. The control room setting. The third operator usually sits to the left of the second operator.

'Talking to the room' and cognitive empathy
In several papers about cooperation where people are colocated it has been reported that people 'talk to the room' (cf. Heath & Luff,1992; Heath, Luff, Nicholls, 1996; Samurçay & Rogalski, 1993). By this we mean that people openly voice bits and pieces of information from what they do at the moment. These bits and pieces of information are not addressed to a specific recipient but rather to anyone who for the moment can make use of the information or would be able to use it in a future task. In tasks where people act independently, 'talking to the room' does not demand a overt response and is sometimes just dropped (Heath, Luff, Nicholls, 1996). In other situations 'talking to the room' might consist in an overt request for sharing a task, (Samurcay & Rogalski, 1993).

In the situations we have focussed, the institutional rules prescribe that the first person taking a call has to ask for assistance when the accident can be regarded as a "priority one" situation (i.e. life-threatening). However, the team members also use it as a way to confirm an informant's reliability when making a decision on how to act on the call. Often such decisions are stressful and demanding and it is advantageous to let the burden be divided between people. The persons working together must learn what we will coin "cognitive empathy". Cognitive empathy implies not only knowing when and how to employ routine task distribution but it also concerns the ability to understand

another person's need for help. In control rooms such as managing the underground (Heath & Luff, 1992), air traffic or industrial process, the persons involved have a clear task assignment and much of the work is quite routinized and well known for all personnel. However, in the emergency centre the presumptive supporting person has not always had the opportunity to understand what the accident is about, what needs the caller has, and what needs the control companion has. Experienced operators in all the above situations have been observed to monitor the control room at every time they either hear another operator is overloaded with work or when they themselves have less workload. In the emergency centre there at times are more acute 'priority one' cases to be handled than there is personnel, which makes it essential, not only to monitor the work load and handle it cooperatively in a routine way, but also to monitor information needs and emotional stress of fellow operators as well as assessing the accident in order to be of any help to manage the case. All this has to be managed within a minimal time so that any rescue services can be deployed as fast as possible.

Creation of cognitive empathy

If the case is judged to be a "priority one" accident, the interviewer is urged by institutionalised rules to call for "co-listening" from whoever is free at the moment.

Table 1. Overhearing and co-listening in an asthma case

	Person 1	Person 3	Comment
1.	90000, emergency centre		# 3 is free and is checking
2.			if there is anything he can help
3.	What do you say?		
4.			
5.	I don't hear		
6.	anything.. Yes		
7.	Hello, have you got asthma		#3 turns his attention to 1, as soon as he
8.			hears the word "Asthma", but then seems a
9.			bit relaxed until he sees #1 turn towards the
10.			"co-listening button"
11.			
12.			
13.			
14.			
15.			
16.	May I get your telephone		# 1 asks for co-listening, #3 takes it
17.	number?		
18.			
19.	And the area code is?		
20.			
21.	And where do you live?		
22.			
23.		I am with you	
24.	Is it		
25.	x-street?		
26.	Yes	Oh, this one is really bad!	
27.		I'll get an ambulance	
28.	And your name?		
29.			
30.	Can you open the door		
31.	when we arrive?		
32.	It is open, OK		
33.	We are on our way		
34.			
35.	Thanks Bye		
36.			
37.	Oh...	This one was really bad	
38.			
39. 6 seconds later		She must have waited very	
40.		long before she called!	
41.			
42.	Uhmm, yes...		
43. 3 seconds		She had no air left	
44.	No...		#3 calls ambulance

Hereby the activities get divided between the interviewer and the co-listener.

Now the interviewer may be able to collect even more information from the accident reporter, which s/he later can transfer to the rescue service personnel in order for them to prepare certain instruments or the like. At the same time the co-listener listens to the interviewing and can call the rescue services and give them the appropriate address etc.

In less straightforward cases the interviewer might ask for a co-listener because s/he is uncertain about the case or about the informant. In such cases the operators confer within the room in order to make a decision of whether or not to send rescue services.

Often we can see how a presumptive co-listener monitors other operators and their voices, to prepare themselves to take over parts of the task. In serious cases or just in strange cases both operators seek eye contact and hereby tell whether or not there will be any transferring of activities. If an operator notices any possibility that any operator would transfer a case to them, they immediately turn to their computer and search for the case record. By that they

Table 2. A fake call?

48			43 back at the station
49	I do not hear		Ready, end.
50.#1 calls for co-51.listen-	Sure	He's lying outside [street]	[takes the co-listen-
ing	[telephone number]	But what number?	ing]
52.		[nr] OK	
53.	yes	I send an ambulance	[I am] with [you]
54.	(?)		
56	Do you know their name or	They'll be there in about te	
57.	so	the fifteen minutes	
58		[ends this call]	
59.	[towards 3]		
60.	has got a bottle in the throat		
61.	(?)		
62.			
63			
64.#1 talks to # 3			
65.	[name]		
66.#3 asks # 2	well, how old is he?		
67.about			[towards 2]
68. ambulances			#2 which car is it, is it
69.			142 ?
70.	No, they are not free		
71.			in Östhammar
72.			
73.			42 are free, or?
74.			
75.	42 are not free. I've got		
76.	them on the shelf here		Aha
77.			
78.	17 years.Yes, they'll arrive		
79.	immediately		
80			The water tower,
81.	Yes, bye.		Långegrund yes
82.			But then we'll have to
83.			take forty-
84.			
85.			
86.		Have you got something	
87.		there too,	
88.		Noo,	Take 43 then
89.		you are not sane and I have	
90.		got one in Öre-Schimo	
100.			
101.			
102.			Gimo?
103.			What are we going to
104.			do then?
105.			
106. Ambulance 107.calling	Hallo		Is it urgent or?
108.(Observe how		This...	
109.#2 argues for		This I thought was fake, you	
110.his own case)		see,	
111.	You are at home,		
112.	yes, OK	Actually	But you...

can see how far the interviewing person has come. Consider the following quite straight forward case, where an asthma patient calls the emergency centre and hardly can talk.

Operator #1 receives the call. Operator #3 has just closed conversation with operator #2, and has no emergent task, and overhears #1's answer. At first he is just routinely attending this emergency call as there is nothing else to do, but by the sound of the operator #1's voice as well as the hand gesture towards the "transfer button", he is able to predict the call for co-listening. Then #1 pushes the button (16) and thereby requires support. Person #2 is working independently and is not transcribed here.

All operators have several possibilities (on the computer screen, by the signal, and by a display) to see that it is a call to the 911 number so it is no suprise that operator #2 and #3 stop their conversation immediately when #1 has taken the call (5). When operator #1 asks the caller what she/he says, operator #3 starts to turn his chair attentively (7). This might be a way to show that you for the moment is available, as Heath & Luff (1991) have suggested . When operator #1 says "asthma", #3 very clearly direct his full attention towards #1. When #1 moves her hand towards the button where they call for co-listening,.#3 turns to the computer and starts co-listening to the call almost at the same time as #1 asks for help. #3 tells #1 that he is with her (23) and thereby confirms that he will take care of the task. After only being able to hear a few "words" he is confirming through the headset phones to #1 that she has assessed the situation correctly (26) and that he is about to call an ambulance (28). We might also point out that #1 is

Time	#1	#2	#3
113.	[to the caller]	but- let's take ambulance,	
114.	We have a (?)	then	
115.			In Gimo yes
116.	on its way		
117.			
118.	yes		
119.	Mm		
120.			Take Gimo
121.			Shall I check with 42 and
122.			hear where they are doing?
123.			Was it a
124.		Hm, I do not know at all	
125.		I don't believe this, you	
126.		know	
127.			
128.		42 should be at home, but	43 on the station, you know
129.		they should call for infor-	
130.		mation	
131.			
132.			
133.	They are at home. I HAVE GOT		
134.	THEM HERE	Yes	
135. New call			
136.		Yes, but take them then,	
137.		there, on your	
138.			
139.	Yes, but shouldn't we call the		
140.	doctor out there		
141.			Yes, we ll do this,
142.			but take each one a car first
143.			and they should have yes.
144.			And I'll take 90 000 here
145.			then
146.			
147.		Yes, I'll take Gimo on mine	
148.		first	
149.			90000
150.			
151.			ambulance
152.			
153.		Very unclear what this is all	Hello, yes
154.		about	
155.			
156.			Yes, ambulance, fireguard
157.			Hello
158.			

flexible in her cross-checking of the authenticity of the call because of the obvious problems the caller have in speaking, #1 rephrases some questions in yes or no form (24, 30).

Later in the case, while waiting for the ambulance to connect, #3 initiates a conversation with #1 of his assessment of the mission (37). Hereby they each give one another feedback on the situation and confirm that everything that should be done actually has been done. This episode show the importance of having some more personal debriefing of a stressful situation. Clearly, #3 assesses that #1 had problem with the interviewing the patient and possibly wants to initiate an discussion about that. This is a particular form of cognitive empathy which is more than just the instrumental work of the team and deals with the emotional situation as well.

Loss of cognitive empathy

In the following example, a more complex case is illustrated, as well as a higher workload. Two cases are going on simultaneously. Operator #2, sitting in the middle, is working with a call concerned with a person lying on the street. He considers this call to be fake and tells this 'to the room' several times. Operators #1 and #3 at his sides are involved in collaborating on a case concerning a bottle in the throat. There are several instances of turbulence and jumping between these two cases.

When we enter the situation, #1 uses the computer to tell the other operators that she needs assistance (50). #3 takes it and tells her that he is with her on the phone and #1 turns herself towards #3 and tells him briefly and in an upset voice about the case (61). #3 understands at this moment that the assistance she needed was not to assess the situation, but rather to call an ambulance as fast as possible. #2 finishes his call and has promised that an ambulance will come (64), and #3 asks #2 (the ambulance responsible) about the ambulances whereabouts (66). #1 must overhear the question as she immediately answers it in a quite stressed and loud voice and that she's got them in her computer (77) ("on the shelf" means that you have got touch with them). #2 has not noticed the turbulence about the case #1 is taking care of and asks in a suprised voice about what is happening (86). #3 has not noticed that #2 has got a job and starts dividing the resources between them (88). #2 now tells them that he thinks this case is a fake (108) but does not at all get any empathic response, rather #3 tries to redirect #2 attention from assessing the situation right to actually send out an ambulance by saying "but you" (110). #2 clearly has a problem deciding whether or not this case is serious or just a fake, but he does not get any response from the other while pointing it out. The situation continues below with #2 accepting to send an ambulance to the potentially fake alarm. Notice especially how insecure #2 is (in bold).

This episode shows very clearly all the talking to the room (in italics). Much of it is related to overt local coordination, but #2 several times raises his doubts about 'his' mission (124-127, 153-154). We can see in this stressful situation that the operators clearly ignore all the emotional and more problematic questions about assessing #2's situation. The other two operators instead urge #2 to send an ambulance (120) and the local coordination is mainly concerned with getting all this right (124-138). #1 takes care of the ambulance for her/his case but asks someone to call the local doctor (139-140), which #3 takes care of after taking the three incoming calls (148-157). Later, #2 even calls for a "big ambulance", which is an ambulance with a medical doctor (not shown).

Despite the fact that #2 is uncertain about the seriousness of the case, he sends out an extra ambulance still later in the process. At first, #2 has taken this case rather calmly. Just at the end of this episode, all operators are occupied, so even if #2 would like support in assessing the situation he couldn't. When he talks to the ambulance drivers he again tells them he thinks the case is a fake. And afterwards he again tells the other operators within the room, but again with no response. The stress and the high pressure on local coordination seem to let them forget empathy. It seems as if #2 got enticed by #3 to order two ambulances without having any good ground for considering the case *that* serious. This may be due to #3:s heavy commitment (which #2 transfers to his own case) and to the long deprivation of any attention from #1 or #3.

Discussion and conclusion

The examples show that both the planning laid down in the organization and the situated collaboration affect the temporary coordination. An important feature in collaboration seems to be people's preparedness to watch for each others' needs, shown in the first example. Another important feature is people's need for others' help, here shown in the second example. Cases of mental or collaborative overload may then lead to nonoptimal decisions, as in the second case when sending two ambulances to an accident one thinks is fake.

Acknowledgement. This study was supported by a grant from the Swedish Work Environment Fund.

References

Heath, C., Luff, P. (1992) Collaboration and Control. *Computer Supported Cooperative Work* (CSCW) no.1, p 69-94.

Heath, C., Luff, P., & Nichols, G.M. The collaborative production of the document: context, genre and the borderline in design. In: Y. Wærn (Ed.) *Coordination of process management.* Report to COST 14, June 1996.

Samurçay, R., Rogalski, J. (1993) Cooperative work and decision making in emergency management. *Le travail humain*, 56 (1) 53-77.

Construction of Shared Representation Through Cooperative Dialogues of Shift Changeover

Corinne Grusenmeyer
French National Research and Safety Institute
Ergonomics and Industrial Psychology Department
PO Box 27
54501 Vandoeuvre Cedex, France
+33 83 50 20 00
grusenme@inrs.fr

ABSTRACT
Functional communications are considered as a means of cooperation and adjusting operators' representations. This function of dialogues seems crucial during the phase of shift changeover. Dialogues should allow each operator to assess his own contribution to the collective aim, and the incoming operator to construct a representation of the situation. The paper presents an analysis, based on interlocutionary logic, of such a dialogue in a nuclear power plant. It shows the construction by the operators of a shared representation of a malfunction and their cooperative problem solving. These results emphasize the importance of such exchanges for the reliability of systems.

Keywords
cooperative dialogues, shared representation, shift changeover, interlocutionary logic, nuclear power plant

INTRODUCTION

Functional communications in work situations, and especially in collective process control, are essential. These communications are indeed assumed to be a means of coordination, cooperation and adjusting operators' knowledge and representations (De Terssac et al., 1990; Lacoste, 1993; Leplat, 1994).

This last function of communications seems crucial during the phase of shift changeover. Indeed, this particular situation of collective work requires cooperation between operators.

The succession of teams on the same process gives rise to cooperative work, as team members collaborate in the same activity, and have a common global aim, namely to carry out the production process. However, it constitutes a specific situation of collective work, insofar as the operators involved have the same goals, the same tasks to achieve and the same abilities, but also because cooperation is, in such a situation, partly delayed. Direct cooperation can only take place at the time of shift changeover.

This situation can then be described by means of Searle's collective intentionality formula (Searle, 1991). According to him, collective intentionality is neither a conjunction of individual intentionalities, nor this latter conjunction with mutual beliefs added. It consists rather of a means-to-end relationship, in which the end is collective and the means are individual but not necessarily identical. In the case of a shift changeover, the collective intentionality, namely to carry out the same process, is performed by individual means: each operator contributes successively to the collective aim.

Hence, the shift changeover phase seems to require an adjustment of operators' knowledge and representations. This phase must indeed allow production to continue. Thus, operators must firstly cooperate to receive feed-back regarding the effects of their own actions, and secondly, both continue actions instigated by other operators and construct a functional representation of the situation. Finally, they must cooperate in order to estimate, in real time, their own contribution to the collective intentionality. Thus, due to the fact that they have a common goal, the same tasks and abilities, operators are assumed to construct a shared functional representation through dialogues.

We will use the terms "shared representation" and "collective reasoning", rather than "common representation" and "distributed reasoning", in order to distinguish the present situation from those where people involved in cooperative tasks have different abilities (Rogalski, 1994; Schmidt, 1994). Indeed, according to Searle's definition of collective intentionality (Searle, 1991), the individual means for operators to contribute to the collective goal are, in our case, the same (contrary to a situation such as that of a football team, for example, where each player has a specific role in the team).

HOW IS ACCESS GAINED TO THE CONSTRUCTION OF A SHARED REPRESENTATION BY OPERATORS ?

Verbal exchanges have, for a long time, been recognized by psychologists, and notably by work psychologists, as a means of structuring and/or modifying cognitions (Perret-Clermont et al., 1991; Piaget, 1976; but also De Terssac et al., 1990; Falzon, 1991; Lacoste, 1993; Navarro, 1993; Savoyant, 1977, 1985).

However, studies in work psychology were more involved in interactions between operators and the process environment than between operators themselves. Hence, the verbal exchange analyses often concentrated on the contents of the exchange, insofar as they provided evidence of the activities of the operators. These analyses consisted of isolating and classifying parts of exchanges according to their contents. This enabled exchanges related to such and such a content to be quantified, and highlighted changes in

communication according to the operators' level of experience and/or the requirements of the work situation.

However, these content analyses are not, in our view, suitable for assessing the role of adjustments in verbal exchanges. In such analyses, the utterances of each operator are often viewed as a whole and considered as a single contents category. It is consequently difficult to access the dynamic characterictics of the verbal exchanges and their evolution. Hence, this method does not allow access to what role verbal exchanges may play in modifying the operators' representation. At best, the shortening of verbal exchanges could be understood as evidence of operators' shared knowledge.

Interlocutionary logic (Brassac, 1994; Ghiglione et al., 1993; Trognon, 1993) proposes a clinical method that appears to be suited to the aim of gaining access to the potential construction of shared representation by operators. This interlocutionary logic method seems to be relevant for many reasons.

First, this clinical method is based on a constructivist assumption. According to this assumption, experienced events are only part of individual knowledge. Thus, the underlying perspective is that people also build their knowledge on the basis of interactions with others (Resnick et al., 1991; Schegloff, 1991).

Second, and this time from a strictly methodological point of view, this clinical method has the advantage of following the development of dialogues step by step, and enables the operators' representations to be described as the dialogue unfolds. Consequently, this method gives access to the processes of joint elaboration of information insofar as:

- it is based on sequencing in conversation. Hence, speech acts are analyzed recursively. It is then feasible to take the evolution of the exchange into account;
- it takes the logical, semantic and pragmatic relationships between successive interventions into account. The dynamic and interactive characteristics of verbal exchanges can then be considered by studying what the intervention of an operator modifies or confirms in relation to the previous utterances;
- it proposes evidence of shared representation: the double agreement. A double agreement corresponds to a sequence of three turns of speaking, where a participant confirms the previous utterance of his colleague, and the latter, in turn, gives his agreement to this previous agreement.

This method should consequently allow the analysis of the role verbal exchanges play in adjusting operators' cognitions.

The aim of the study was twofold:

- firstly, analyzing the contents of verbal exchanges during shift changeovers and their potential changes according to the work situation and the process state, with a view to gaining a better understanding of this work-phase and of the contents of exchanges occurring at this time;
- secondly, carrying out interlocutionary logic based analyses of a number of exchanges in order to assess their role in adjusting operators' representations at the time of shift changeover.

This paper will essentially focus on this second point (see Grusenmeyer, 1995a, for more information on the first analyses).

SITUATION ANALYZED AND METHODO-LOGY

First, the process and the workstation studied will be presented. Then, the methods used and the data collected will be described.

Process

Verbal exchanges between operators during the shift changeover phase in a nuclear power plant were gathered. The process is highly automated and computerized. It is a continuous process with a very short response time.

The supervisors of the nuclear power plant were observed. They run the primary and secondary sections of the station. These sections of the installation provide the energy for the heart of the reactor, the circulation of primary fluid allowing extraction of the heat produced by nuclear fission, and the heat exchange between the primary fluid and the secondary water source. This secondary water is transformed into steam, thereby producing the power for the turbo-alternator.

Both the supervisors and the auxiliary-equipment operators (the latter responsible for the running and monitoring of the reactor auxiliary circuits) have identical training and alternate successively at the work stations.

Method

Two types of method were used:

- systematic on-site observations (activities analysis) using an observation grid, a video camera and microphones;
- after completion of the shift changeover phase, systematic verbalizations of the incoming operator and, if possible, the outgoing one, on the basis of the videotape.

The following procedure was used:

- on-site observations of the outgoing operator's activities during his shift and before the meeting between the two operators;
- recording of the communications between the supervisors by microphone and of their activities by means of the video camera;
- on-site observations of the incoming operator's activities during the beginning of his shift;
- consecutive verbal explanations (verbalizations) based on the videotape with the two operators, after completion of shift changeover.

Data Collected

13 shift changeovers between supervisors were observed.

These observations varied according to:

- the time of shift changeover (7 a.m.; 7 p.m.);

- the state of the process;
- the last shift of the incoming operator (ending 12 hours earlier, 48 to 96 hours earlier, or longer);
- the "meeting rate" of the operators (because of 6 teams working twelve hours shifts, two supervisors could succeed one another frequently, relatively often, or rarely).

RESULTS

Analyses of the Exchange Content

The first analyses, based on content categorization, show that verbal exchanges can be very different depending on the work situation, namely the process state, the last shift of the incoming operator, and the time of shift changeover.

Only a few results will be presented (for further details, see Grusenmeyer, 1995a):

- the "meeting rate" does not seem to change the content of verbal exchanges;
- a large part of the exchanges (about 20% of the exchanges) is related to the process control system;
- the contents of exchanges vary according to the time of shift changeover. This is related to the fact that the production team is alone during the night shift.

Some Results about the Format of the Exchanges

A number of the analyses carried out were aimed at testing certain hypotheses regarding the format of the verbal exchanges between operators. They concerned the length of the exchanges, the degree of participation of the incoming operator, and the operator initiating the exchange, these variables being related to the exchange content.

For example, it was assumed that, since the supervisors are skilled and probably have common knowledge, exchanges related to malfunctions or incidents would be the longest and should result in greater participation on the part of the incoming operator. Indeed, in these cases, skilled operators should construct in situ a shared functional representation that is not needed when the contents of exchanges are usual.

Indeed, certain studies (Savoyant, 1977, 1985) have shown that when operators have common knowledge this leads to shortened exchanges, whereas the need to elaborate a common referent gives rise to additional exchanges. Moreover, the studies conducted by Falzon (1991) highlight that when skilled operators exchanged problematic or unusual contents, namely contents for which operators do not have a common referent, they no longer use an operative language but a natural one. More importantly, these conversations and the turns of speaking are unusually long (about common referent, see Clark et al., 1991; De Terssac et al., 1990). In our view, these results have to be related to the need for elaborating a common representation jointly.

The results obtained from our data confirm the hypothesis. In particular (for more details, see Grusenmeyer, 1995a), they show that, whereas verbal exchanges take the form of information transfer by means of a few words from the outgoing operator to the incoming one when usual events are evoked, they constitute true dialogues when related to malfunctions or incidents. In this latter case, the operators are then probably discussing or cooperating in order to pool their own information and to construct jointly a shared functional representation of the problem.

Example of Diachronic Analysis: the Shared Diagnosis of an Anomaly

More detailed analyses based on interlocutionary logic were carried out on a number of verbal exchanges to assess the role of these exchanges in adjusting in situ operators' representation.

As pointed out earlier, skilled operators probably do not need to adjust their own representation concerning all the contents evoked. Some verbal exchanges, for which it was assumed that the operators should elaborate jointly a shared representation, were then chosen according to their length, the degree of participation of the incoming operator, their contents, and the operator initiating the exchange (because of unequal positions of the incoming and the outgoing operators relating to the available information).

Three verbal exchanges were then chosen and analyzed by means of interlocutionary logic. An example of such an analysis will be briefly presented (for more details, see Grusenmeyer, 1995b, 1996; Grusenmeyer et al., to appear).

This exchange between two supervisors is taken from a 7 a.m. shift changeover in a shutdown situation with nominal heating conditions. The operators have to maintain this state until the beginning of the afternoon. The last shift of the incoming operator ended 12 hours earlier. He had had to supervise the auxiliary equipment. Whilst carrying out his end of shift round, he had observed an anomaly concerning the nitrogen circuit of the installation. This circuit is annexed to the installation and allows water degassing and the operation of an automatically triggered neutron neutralisation safety system.

The nitrogen circuit is located as shown in Figure 1.

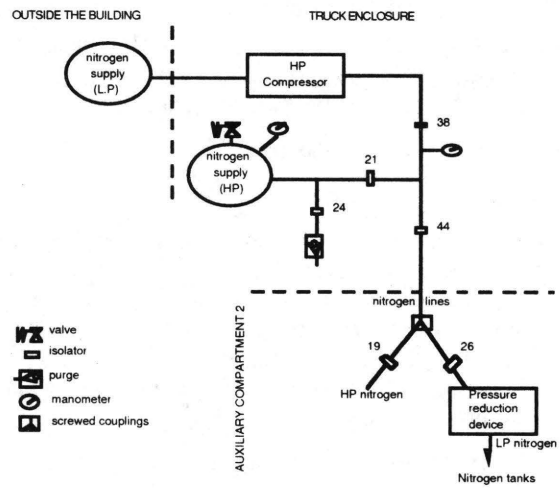


Figure 1. The nitrogen circuit

This circuit comprises:

- a low pressure (L.P.) liquid nitrogen supply outside the building;
- a compressor located in the truck enclosure allowing the nitrogen pressure to be increased;
- a high pressure nitrogen supply located in the truck enclosure and pressurised by means of the compressor (H.P. nitrogen). A valve allows evacuation when the pressure is too high;
- nitrogen lines supplying various tanks with either H.P. or L.P. nitrogen by means of a reduction valve. These nitrogen lines and the various tanks are located in the AC 2 (Auxiliary Compartment 2);
- two manometers located in the truck enclosure that indicate the circuit pressure; a purge upstream from the H.P. nitrogen supply; and different isolators that can be open or closed as required.

The incoming operator had noticed during his last shift that the H.P. supply pressure was low and suspected a leak. To test his hypothesis, he had increased the circuit pressure slightly and it appeared that there was. He transmitted this information to the incoming auxiliary-equipment operator and supervisor (the latter being the outgoing supervisor of the present shift changeover). This information had to be confirmed since the test carried out by the incoming operator only allowed him to suspect a leak. On his return the following morning at 7 a.m., the incoming operator was informed by his outgoing colleague of this anomaly.

Exchange and Operators' Reasoning

During this meeting the two operators discuss the anomaly and its diagnosis. The first part of the exchange allows agreement on the diagnosis of the anomaly (turns of speaking OG 1 - OG 4). The two operators deduce from the tests carried out by the outgoing operator that there really is a nitrogen leak. Being in agreement, they then reason together in the second part of the exchange (OG 4 - OG 8) to locate the malfunction. Their reasoning is then inductive (cf. Da Silva Neves, 1994). They infer new information, namely the cause of the anomaly. This leads to a new agreement between the operators.

The exchange is as follows (OG and IC designate the outgoing and the incoming operators respectively, and the numbers, the turns of speaking of each operator):

OG 1: That nitrogen leak there...

IC 1: Yeah.

OG 2: You know, it's a bit odd, because me, I put it back up to x bar at the tank and I isolated the entire truck enclosure on the two isolators. In one, in the first two hours, it dropped by nearly, practically y or z bar. F., he did the purge at the bottle, he pressed again a little, he reopened it and pressed again and, umm...

IC 2: He closed the purge at the bottle, then.

OG 3: Yeah he did it but it was closed, though there was no noise, nothing. And apparently we lost less, we lost less but we still lost a bit all the same.

IC 3: So, there really is a leak then.

OG 4: Now, maybe we should look at the bottom to see, umm, section by section.

IC 4: But, umm, what, at the bottom in AC 2 you mean ?

OG 5: Yeah.

IC 5: Umm, at the bottom, umm, it was isolated yesterday.

OG 6: Mm.

IC 6: As I didn't touch it well, at first sight, it can't come afterwards, can it ? Or else, that's what it is it's really one of the points...

OG 7: Between the two. That's it. Between the two, yeah.

IC 7: Mm. That's right.

OG 8: Must check it.

IC 8: Apart from that (...)

Identification of the Malfunction at the Origin of the Anomaly Observed

At the end of his previous shift, the incoming operator had made the assumption of a nitrogen leak at the origin of the observed anomaly (low pressure of H.P. supply). This diagnosis had to be verified.

In fact, the outgoing operator had tested the hypothesis of his colleague. That is what he explains in OG 2 - OG 3. To do this (cf. figure 1), he had filled the nitrogen tank, increased its pressure (*"Because me, I put it back up to x bar at the tank"*), and closed isolators 38 and 44, thereby allowing isolation of the truck enclosure from the nitrogen lines (*"I isolated the entire truck enclosure on the two isolators"*). Isolator 21 was open. Isolator 24 was closed, and this is confirmed later in the exchange (*"F. he did the purge at the bottle, he pressed again a little (...) he did it **but it was closed, though** "*). After a two-hour wait, he read the pressure of the HP nitrogen tank on the manometer. He noticed that it had dropped by *"y or z bar"*. This results tends to confirm the hypothesis of a nitrogen leak.

He then puts forward a first hypothesis concerning the location of the leak: the nitrogen is escaping via the purge (that is isolator 24 is not sealed). This test, carried out by a colleague (F) necessitates:

- ensuring that isolator 24 was closed (*"but it was closed, though"*);
- ensuring that no leakage noise could be heard (*"There was no noise, nothing"*);
- checking, by operating the purge, that it was indeed closed when theoretically closed (*"F., he did the purge at the bottle, he pressed again a little, he reopened it and pressed again");*
- waiting a while and comparing the value on the pressure gauge with the previous one (*"And apparently we lost less, we lost less but we still lost a bit all the same").*

This result means that the purge was indeed badly closed, and consequently that isolator 44 was not totally sealed.

However, this is not enough to explain the entire initial loss of pressure.

Logically, the outgoing operator is in a position to deduce from the test that there really is a nitrogen leak. In fact, the incoming operator makes the final deduction of the reasoning of his colleague: "***So there really is a leak then***". In uttering this deduction, the incoming operator implicitly agrees with the reasoning of his colleague. The two supervisors therefore reason together. The outgoing operator then confirms this deduction. By utterring "*Now, maybe we should look at the bottom to see, umm, section by section*", the outgoing operator proposes a means of looking for and locating the leak. And making such an utterance necessarily pressuposes that he has taken the fact that there is a leak for granted.

Hence, a double agreement is observed. In terms of interlocutionary logic, the incoming operator confirms OG 1 ("*That nitrogen leak there...*", which is a simple proposition) and the reasoning of his colleague by uttering the same propositional content in IC 3 ("*So there really is a leak then*"), an utterance which is a deduction of this reasoning. In turn, the outgoing operator gives his agreement by making the propositional content of IC 3 (that is there is a leak) a preparatory condition of OG 4 (the preparatory conditions determine the propositions that a speaker must presuppose to accomplish a speech act having a certain illocutionary force). Thus, each of these three utterances OG 1, IC 3, OG 4 has the same propositional content (there is a leak). Only the illocutionary force or the status given to this content is modified, since from a simple proposition it becomes a preparatory condition of the following utterances.

Location of the Malfunction

At this point, both operators share and take for granted that there is a leak, but they still have to locate it. That is what they try to do from OG 4 to OG 8. The outgoing operator has already tested isolator 44 and the purge. These elements appear to be insufficient to explain the result obtained. Consequently, the leak is probably located outside of this part of the circuit.

The outgoing operator therefore proposes another hypothesis: the leak is located downstream from this part of the circuit, that is downstream from isolator 44 (this isolator is not sealed). Indeed "*Now, maybe we should look at the bottom to see, umm, section by section*" is a proposal to isolate AC 2 ("*But, umm, what, at the bottom in AC 2 you mean ? / Yeah*") and therefore to test isolator 44.

Both operators then cooperate in order to assess this hypothesis. The incoming operator indeed provides information to his colleague which contributes to its assessment.

Indeed, because he had been auxiliary equipment operator during his last shift, he knows that, at that time, isolators 19 and 26 were closed: "*Umm, at the bottom, umm, it was isolated yesterday*". Hence, the incoming operator takes into account the situation put forth by his colleague, and reasons as follows. As isolator 44 was closed and if isolators 19 and 26 are still closed, the probability that the leak is after isolators 19 and 26 is very small. This is what he utters in IC 6: "***As I didn't touch it well, at first sight, it can't come afterwards, can it ?***". This hypothesis about the location of the leak indeed supposes that two isolators (44 and 19 or 44 and 26) are not sealed.

More than just assessing the probability of the hypothesis of his colleague, the incoming operator starts to propose another one, which is a precision of the previous hypothesis of his colleague. "*Or else, that's what it is it's really one of the points ...*" means: or else, for the leak to be in AC 2, it really must be coming from a coupling. The hypothesis treated here corresponds to the following situation. The leak is located between isolator 44 on one side and isolators 19 and 26 on the other (the couplings located in this part of the circuit are indeed potential nitrogen-escape points).

In fact, this new hypothesis of the incoming operator is only partially stated as he does not end his utterance. The outgoing operator clarifies this hypothesis in OG 7. "*Between the two. That's it. Between the two, yeah*" means between the two previously mentioned sections, in other words the initial section created by the outgoing operator by closing isolators 24, 38, 44 and the purge; and the second one, evoked by the incoming operator, concerning the part of the circuit located after isolators 19 and 26. It could also mean between isolator 44 and isolators 19 and 26. Regardless, it corresponds to the coupling, as this is the only part of the circuit that could give rise to a nitrogen leak.

The new hypothesis of the incoming operator is therefore clarified by his colleague. And the incoming operator agrees in turn in IC 7 ("*Mm. That's right*"). The two operators then jointly construct a shared functional representation concerning the most probable hypothesis of the location of the malfunction. A double agreement is also observed in IC 6 - OG 7 - IC 7.

The operators now have to test this hypothesis in order to continue their reasoning. That is what they propose to do: "*Mm. That's right / Must check it*".

CONCLUSION

This diachronic analysis, based on interlocutionary logic, allows the collective reasoning of the two operators to be followed as the dialogue unfolds. Hence, in the first part of the exchange, the two operators deduce the diagnosis of the anomaly from the tests of the outgoing operator. In the second part of the exchange, their reasoning is inductive, since they infer new information: the most probable location of the anomaly, namely a defective coupling.

This analysis shows that the verbal interaction between the two supervisors allows them to construct jointly a shared functional representation of the malfunction evoked. Whereas, at the beginning of the exchange, the only information shared by the operators relates to the existence of a malfunction and an unverified hypothesis, at the end of the same exchange they share the diagnosis of the malfunction and the most probable hypothesis of its location.

This type of analysis highlights the joint construction of a shared functional representation in all the chosen verbal exchanges. It should be noted that in one exchange observed in another situation (paper manufacturing), the problem evoked by the operators has already been solved. However, the operators have a long exchange, probably due to the fact that it allows them to maintain and improve their expertise.

Furthermore, a situation (also observed in the paper manufacturing) where the operators had to shorten their exchanges because of production constraints, was analysed. It highlights that the representation of the incoming operator was less rich than that of the outgoing one because of a process similar to confirmation bias (Grusenmeyer, 1995b).

Finally, these analyses point out that verbal exchanges during shift changeover not only allow actions related to the production process to be continued, but also cooperative problem solving and improvement of operator expertise. In consequence, they emphasize the importance of these exchanges in the reliability of the production process.

REFERENCES

Brassac, C. (1994) Speech acts and conversational sequencing. *Pragmatics and Cognition*, 2, 191-205.

Clark, H.H., & Brennan, S.E. (1991) Grounding in communication. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington: American Psychological Association.

Da Silva Neves, R.M. (1994) Théories de la psychologie de l'induction. *Psychologie Française, 39*, 123-140.

De Terssac, G., & Chabaud, C. (1990) Référentiel opératif commun et fiabilité. In J. Leplat, & G. De Terssac (Eds.), *Les facteurs humains de la fiabilité dans les systèmes complexes* (pp. 111-139). Marseille: Octarès.

Falzon, P. (1991). Cooperative dialogues. In J. Rasmussen, B. Brehmer, & J. Leplat (Eds.), *Distri-buted decision making: Cognitive models for cooperative work* (pp. 145-190). New-York: Wiley.

Ghiglione, R., & Trognon, A. (Eds.) (1993) *Où va la pragmatique?* Grenoble: PUG.

Grusenmeyer, C. (1995a) *De l'analyse des communications à celle des représentations fonctionnelles partagées. Une application à la relève de poste.* Thèse, Université René Descartes, Paris.

Grusenmeyer, C. (1995b). Shared functional representation in cooperative tasks. The example of shift changeover. *International Journal of Human Factors in Manufacturing, 5*, 163-176.

Grusenmeyer, C. (1996). L'analyse dynamique des échanges verbaux entre opérateurs en phase de relève de poste. *Performances Humaines et Techniques*, numéro hors série de mai 1996, 31-38.

Grusenmeyer, C., & Trognon, A. Structures of natural reasoning within functional dialogues. To appear in *Pragmatics and Cognition* .

Grusenmeyer, C., & Trognon, A. Les mécanismes coopératifs en jeu dans les communications de travail: Un cadre méthodologique. To appear in *Le Travail Humain.*

Lacoste, M. (1993) Interaction située et dimension collective du travail. In F. Six, & X. Vaxevanoglou (Eds.), *Les aspects collectifs du travail* (pp. 29-49). Toulouse: Octarès.

Leplat, J. (1994) Collective activity in work: some lines of research. *Le Travail Humain, 57*, 209-226.

Navarro, C. (1993) L'étude des activités collectives de travail: Aspects fondamentaux et méthodologiques. In F. Six, & X. Vaxevanoglou (Eds.), *Les aspects collectifs du travail* (pp. 91-106). Toulouse: Octarès.

Perret-Clermont, A.N., Perret, J.F., & Bell, N. (1991) The social construction and cognitive activity in elementary school children. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 41-62). Washington: American Psychological Association.

Piaget, J. (1976) Logique génétique et sociologie. In G. Busino (Ed.), *Les sciences sociales avant et après Jean Piaget* (pp. 44-80). Genève: Droz.

Resnick, L.B., Levine, J.M., & Teasley, S.D. (Eds.) (1991) *Perspectives on socially shared cognition.* Washington: American Psychological Association.

Rogalski, J. (1994) Formation aux activités collectives. *Le Travail Humain, 57*, 367-386.

Savoyant, A. (1977) Coordination et communication dans une équipe de travail. *Le Travail Humain, 40*, 41-54.

Savoyant, A. (1985) Conditions et moyens de la coordination inter-individuelle d'opérations d'exécu-tion sensori-motrices. *Le Travail Humain, 48*, 59-79.

Schegloff, E.A. (1991). Conversation analysis and socially shared cognition. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 150-171). Washington: American Psychological Association.

Schmidt, K. (1994). Cooperative work and its articulation: Requirements for computer support. *Le Travail Humain, 57*, 345-366.

Searle, J.R. (1991) L'intentionnalité collective. In H. Parret (Ed.), *La communauté en paroles: Commu-nication, consensus, ruptures* (pp. 227-243). Liège: Mardaga.

Trognon, A. (1993) How does the process of interaction work when two interlocutors try to resolve a logical problem? *Cognition and Instruction, 11*, 325-345.

Computerised Ward Monitoring Systems: Understanding the Stakeholders

M.A.R. Kirby, R. Kane
HCI Research Centre
University of Huddersfield
Queensgate, Huddersfield HD1 3DH, UK
+44 1484 472 524
m.a.r.kirby@hud.ac.uk

K.D.C. Stoodley, J. Bates
Cardiac Research Centre
Killingbeck Hospital
York Road, Leeds LS14 6UQ, UK

ABSTRACT
Stakeholder analysis encourages system developers to consider both the social and technical requirements of a system, normally during the early stages of systems design. In this case, the analysis is applied to a system after it had been developed and proven unacceptable to some of the stakeholders. The intention was to understand why it was unacceptable and enable development of an acceptable version. The system is a computerised ward monitoring system in the Cardiac Paediatric Intensive Care Unit at Killingbeck Hospital, Leeds. This paper reports on the stakeholder analysis of the system and reviews the benefits of conducting the analysis.

Keywords
socio-technical systems, ward monitoring systems, stakeholder analysis, systems design

INTRODUCTION
The success of computer systems depends upon social acceptability as well as technical capability. This is argued by proponents of socio-technical theory who have developed participative methods for developing computer systems in office and industrial environments, for example Eason (1987), Harker (1985) and Mumford & Sutton (1991). The aim of these methods is to involve the users in computer system design in order to achieve a successful balance between social and technical factors. To achieve this balance, job design issues have to be considered as well as technical design issues. Computer systems invariably have an effect on the way people work and their interactions with other people. Social problems usually arise from changes in the work system caused by a computer system rather than the computer system per se. This is not unique to industry and office environments. Medical information systems also require solutions that are socially oriented as well as technically oriented and which require a good understanding of the users' jobs and ways of working, as discussed in Kay (1990) and Nygren at al. (1992).

Socio-technical methods normally require extensive user involvement in the development process. If socio-technical methods are used in conjunction with structured or formal design techniques, it is also necessary for users to develop some understanding of the modelling techniques and terminology used by the developer. The main benefits are that the users' requirements are well articulated, social problems are predicted and the users feel greater ownership of the new system.

However, circumstances may arise in which it is not possible for users to be heavily involved in the development process. This is the case for suppliers of information technology to the mass market, in which there is a wide variety of users or the final users may not be known. The USTM methodology [Macaulay et al. (1990)] is a socio-technical method developed for a mass market supplier. This approach enables the developers to ask the right questions about generic users in order to generate a list of social and technical requirements.

It may also be difficult to involve users heavily in the design of systems for individual organisations even though the users are known. This paper addresses such a situation in the Cardiac Paediatric Intensive Care Unit (PICU) at Killingbeck Hospital, Leeds. Here, substantial organisational and cultural change would need to take place to allow the time and responsibility for full participation by users in systems design. However, the use of a customised version of USTM, called CUSTOM [Kirby (1991)], has allowed a socio-technical approach to be taken in this case. In this approach, all stakeholders are analysed, not just end users. This paper reports on the stakeholder analysis conducted in the PICU at Killingbeck.

OBJECTIVES
The objective of the study is to address stakeholders' problems with a computerised ward monitoring system that had been developed for the PICU at Killingbeck Hospital in Leeds.

An earlier study, Stoodley et al. (1992), identified problems that occurred during preliminary trials of the system. These included technical problems, ergonomic problems and difficulties in arranging training for nurses. Provisions have since been made for dealing with those problems.

There were also some problems associated with the change from paper charts to computerised displays. These problems seemed more difficult to solve. They may be regarded as stakeholder problems because they are concerned with changes to the nature of stakeholders' jobs.

For this second study, it seemed important to address the stakeholder problems and try and ensure that any job changes would be acceptable to the stakeholders.

METHOD: STAKEHOLDER ANALYSIS

The definition of 'stakeholder' is 'anyone who holds a stake in the success or failure of a system, whether they are affected directly or indirectly by that system'. The aim of stakeholder analysis is to gain a thorough understanding of stakeholders and their requirements. To study stakeholders, the developer simply asks questions about them. The CUSTOM manual outlines the questions which need to be asked and provides a structured format for recording and cross-referencing the answers. For each stakeholder, there are questions covering such aspects as his/her role in the organisation, knowledge and skills, sources of job satisfaction and dissatisfaction, the nature of the job and the nature of the main objects handled in the job. Questions are also asked about working relationships with other people and characteristics of the workgroups to which stakeholders belong. The answers provide a description of the current situation. When the answers have been gathered about the current situation, there is a need to consider whether any change to the current situation would be desirable or undesirable. Ideally, the stakeholders should be involved in generating the answers and in reviewing whether changes would be desirable or undesirable. This is best done in a group session. Where this is not possible to arrange, the answers to the questions should be at least be verified by checking them with the individual stakeholders. At the end of this process, a list can be generated of the social requirements which must be met in order for a system to be acceptable to each stakeholder.

The above description of stakeholder analysis reflects the part of the CUSTOM methodology that was used in the Killingbeck Case. The methodology also allows for further analysis to be carried out on appropriate allocation of function between human and computer, and appropriate provision for training, documentation and user support.

To address the stakeholder problems associated with the computerised ward monitoring system in the PICU at Killingbeck, a stakeholder analysis was conducted. The aim was to try and understand any mismatch between the system and the jobs of the stakeholders. If stakeholder analysis is conducted during the early stages of development, it may be possible for developers to conduct the analysis with an open mind. After investing much time and effort in developing a system, a developer is more likely to be defensive about the design decisions that have been made. So, it may be advisable to use an independent reviewer if stakeholder analysis is conducted during the later stages of development. An independent reviewer was used in this case.

The stakeholders were identified and then interviewed. The purpose of the interviews was to generate descriptions of their jobs in terms of stakeholder, job and workgroup characteristics. The characteristics were recorded in a structured format, using forms provided by the CUSTOM methodology.

The stakeholder types included nursing staff, patients, surgeons, anaesthetists, cardiologists, physiotherapists, dieticians, managers, researchers and systems developers. Some individuals belonged to more than one of these stakeholder types. Procedural tasks such as 'completing the paper chart' and 'reading monitors' were considered within the wider job context of caring for the patient. The task of 'updating computer displays' was also considered for comparison with 'completing the paper chart'.

RESULTS

The main results of the study were presented as a list of needs that should be addressed to ensure that the system is acceptable to the stakeholders. The analysis highlighted needs which arise from a broad consideration of patient care in a PICU. The needs also reflect the different views and priorities of different stakeholders.

The following is a summary of the main stakeholder needs in this case.

Need 1: The system needs to be seen to be beneficial by each of the stakeholders. The benefits were clear for some stakeholders, for example managers and researchers. The original system had been designed to achieve two main objectives: to replace paper charts with computerised records and to act as a data collection and analysis vehicle for research programs. Provisions made to meet these objectives were in line with the needs of managers and researchers. However, some stakeholders did not perceive a benefit, for example nursing staff. The nursing staff were in favour of using information technology but did not see how this particular system improved their ability to provide patient care.

Need 2: The provision of training needs to take into account four interdependent factors: most users have a low level of computing expertise; staff resources permit little time for off-the-job training; on-the-job training can be stressful and fragmented in a PICU; and there is a fast turnover of staff working in the PICU.

Need 3: There is a need to address some workgroup interrelationship problems that indirectly affect the acceptability of the computer system. In particular, nursing staff feel that some medical staff underestimate their role in the provision of patient care. This leads them to see the computer system as a heavy additional burden on them to give a marginal benefit to the medical staff.

Need 4: The computerised system needs to support the whole process of patient care, from admission of a patient to the transfer of a patient to another ward or hospital. One consequence of this is that it may be necessary to provide appropriate hard copy facilities for collaborative work with external stakeholders, such as other hospitals or to ensure appropriate legal status of drug records.

To address the four main needs, it was possible to derive some implementation needs from the stakeholder analysis. Some of these are not directly concerned with the design of the computer system itself, but are concerned with improving the job situation. For example, to address stakeholder needs 2 and 3, there is an implementation need for discussion between developers, nursing staff and other workgroups. The aim would be to help each workgroup to see the mutual benefits and mutual costs of the new system. There is also an implementation need for greater resource provision to allow users time to learn and establish use of the system without stress. Both of these needs would be partly supported by greater user involvement in design of a new improved human-computer interface

Improving the human-computer interface is one of the implementation needs, which is more directly concerned with the design of the computer itself. For another example, to address stakeholder need 1, there is a need to give the nursing staff direct benefits by identifying repetitive tasks that could be reduced, and by using more automated input from monitoring devices. These would enable nurses to devote more time to more important and interesting aspects of patient care. There is also a need to provide at-a-glance data retrieval for nurses as well as supporting the information retrieval needs of medical staff and researchers.

ACTIONS TAKEN TO ADDRESS STAKEHOLDERS' NEEDS

This section describes actions taken to enable the stakeholders to appreciate the benefits of the system. The approach was to first gain a fuller understanding of the perceived advantages and disadvantages of the paper based and computer based systems, then to try and address the concerns of the stakeholders.

From the nurses' perspective, there were a number of concerns. Firstly, they felt that it was difficult to input data to and retrieve information from the computer system. To improve the system, the entry of data on the computer chart has been simplified by the use of menus for non-numerical data and by the introduction of more comprehensive editing facilities. Secondly, the nurses had a problem with carrying out quality checks when using the paper based system, namely that it is difficult to double check their entries on the paper chart when there are other critical tasks to perform. They have been shown how the computer system can help them by performing quality checks on numerical data, reducing the pressure on them to do so. They have also seen how the computer can help them by quickly and accurately performing calculations for them, e.g. fluid and blood balance totals and pump drug dosage levels. Thirdly, they were concerned about the reliability of the system, having experienced the system crashing and becoming unavailable. Extensive testing has been carried out to considerably improve the reliability of the system, to demonstrate the improved reliability, and explain precautions taken to ensure that data is not lost. Fourthly, the nurses have been provided with a facility to obtain a hard copy of patient information, which they required. Fifthly, an attempt has been made to improve the siting of the computers so that they do not obstruct the nurses' view of patients or monitoring equipment, although it has not been possible to reach an entirely satisfactory solution to this problem. Sixthly, unfinished work has been carried out to improve the groupings of information displayed on the screen. This arises from the way that nurses (or medical staff) can quickly gather information from different parts of the paper chart. The biggest relative advantage of the paper chart is that a single chart shows the whole of the patient's history over 24 hours. This cannot be achieved on a computer display. However, as the developers gain greater understanding of which information is viewed on the chart during different medical activities, it should be possible to improve groupings of information on the computer system to support those activities. Seventhly, the nurses felt that they could easily adjust the paper chart to enter additional information when required, e.g. by dividing a column into two, and were concerned at losing this facility on the computer system. This concern has been addressed by introducing facilities for altering chart headings and using blank columns on the computerised chart.

In addressing these concerns of the nurses, greater effort was made to provide appropriate training for more staff. This training included explanation of how the computer system addresses their concerns as well as taking more time to explain how to use the system. Continual training was provided successfully for ten months, covering nearly all established and new staff, despite the high staff turnover in the PICU. Unfortunately, the new training regime was hit by a general funding problem that stopped all further work on this project.

The medical staff and clinical manager shared some of the concerns of the nurses, particularly concerns over reliability and the need to improve the way information is grouped on the computer system. The medical staff have an additional need to find information on several different variables from several different charts. To address this, a way has been found of providing such information in graphical or tabular form with a delay of less than 1.5 seconds, which is a considerable improvement on scanning several large paper charts. These displays help the medical staff to quickly see correlations and trends in the data. A review has been attempted with the medical staff to see if these displays meet their needs. Another problem for the medical staff is how to retain their newly learnt computer skills when they use the system in at irregular intervals, whereas the newly learnt skills of the nurses are reinforced by frequent regular use of the system. This concern has been addressed by allowing most information to be accessed via function keys rather than requiring more complex use of the keyboard.

The research staff had relatively few concerns, as they understood the advantages of using the computer for retrospective analysis of patient data and testing of research algorithms.

While considerable progress has been made in gaining the active participation of the nursing staff in the project, it has been more difficult to interest the medical staff, with the exception of the physiotherapists. In particular, the lack of

active participation of a senior member of the consultant staff may have been a crucial factor in failing to gain more support from the medical staff. To address this problem, it will be necessary to further investigate the way that medical staff use information on the paper chart and to demonstrate how the computer system can support them in carrying out diagnosis and other medical activities. Some preliminary work has been done in this area, which has influenced unfinished development of facilities for entering nurses' notes and producing patient reports. It has also been identified that it may be useful to try and integrate the ward patient care plan into the computer system.

DISCUSSION AND CONCLUSIONS

It can be difficult to fully assess the benefits of stakeholder analysis. This is because the main benefits often emerge before the analysis itself is complete, and may be attributed to other interventions in the situation. In this case, it is arguable that two main benefits arose as a result of involving stakeholders in the conduct of the analysis. These arose from being involved in discussions about the work as well being involved as interviewees. The first benefit was that the developers and some other stakeholders became more aware of the social issues associated with the computer system, where they had previously focussed mostly on the technical issues. The second benefit was that some of the stakeholders began to understand the views of other stakeholders and gain an appreciation of their needs. There were also other benefits that stakeholders gained enhanced interest in the project, which may be partly due to the Hawthorne Effect, that is their motivation increased because of the greater interest shown in them. As a result of the two main benefits, progress was made in addressing the training and support needs of users, and there was increased cooperation of stakeholders with the developers.

Although considerable progress was made, as discussed in the last section, resource problems arose which provided immovable obstacles to further progress. The first was lack of resources for translating relevant stakeholder needs into an improved human-computer interface design. The CUSTOM methodology has processes that facilitate interface design, based on stakeholder analysis. These processes may have been useful in this case, but were not applied because the independent analyst had to move on to other work and resources were not available to employ another analyst to continue with the interface design stage. One danger of using an independent external analyst is that some of the value of stakeholder analysis may be lost if the analyst does the work and moves on; this can be avoided if the developer and stakeholders are encouraged to carry out the analysis themselves. The problem of resourcing interface analysis and design was part of the more significant problem that no more funding was available for any further development of the system.

It is particularly difficult to make an assessment of the value of stakeholder analysis in this case, because the project was not followed through to a natural conclusion. It appeared that the analysis, and the way it was conducted, enabled progress to be made towards developing a more acceptable system. This was not just because the developers gained a better understanding of the stakeholders, but because each stakeholder group was beginning to understand the needs of each other group. It is not known whether this would have definitely led to an acceptable system or whether it would have raised false expectations that could not have been met. Despite this reservation, it may be worth pursuing an approach that appears to have a positive affect on the design of acceptable computerised ward monitoring systems.

ACKNOWLEDGEMENTS

K. Stoodley and J. Bates were supported by Killingbeck Children's Heart Surgery Fund during the course of this work.

REFERENCES

Eason, K. (1987) *Information Technology and Organisational Change.* London: Taylor & Francis.

Harker, S. & Eason, K. (1985) Task analysis and the definition of user needs. In Mancini, G. et al. (eds) (1985) *Analysis, Design and Evaluation of Man-Machine Systems* . Oxford: Pergamon Press.

Kay, S. (1990) Towards relevant medical information systems. *Medical Informatics, 15, 4.*

Kirby, M. (1991) *CUSTOM Manual Dpo/std/1.0.* Huddersfield: HCI Research Centre, University of Huddersfield.

Macaulay, L., Fowler, C., Kirby, M. & Hutt, A. (1990) USTM: a new approach to requirements specification. *Interacting with Computers, 2, 1.*

Mumford, E. & Sutton, D. (1991) Designing organisational harmony and overcoming the failings of current IT methodologies. *The Computer Bulletin, August 1991.*

Nygren, E., Lind, M., Johnson, M. & Sandblad, B. (1992) The art of the obvious. In Bauersfield, P., Bennett, J. & Lynch, G. (eds) (1992) *Proceedings of CHI '92 Conference.* Reading: Addison-Wesley.

Stoodley, K., Walker, D., Crew, A. & Marshall, J. (1992) Problems in the development of a computerised ward monitoring system for a paediatric intensive care unit. *International Journal of Clinical Monitoring and Computing, vol. 8.*

A Preliminary Model of the Planning and Control of the Combined Response to Disaster

Becky Hill & John Long
Ergonomics & HCI Unit, University College London,
26 Bedford Way, London WC1H 0AP, England.
Tel: +44 (0)171 387 7050.
Fax: +44 (0)171 580 1100.
E-mail: b.hill@ucl.ac.uk.

ABSTRACT
The emergency management combined response system is a planning and control system, set up to manage disasters. This paper presents a preliminary model of the planning and control of the combined response system and its domain. The model is based on a framework for modelling planning and control for multiple task work and on data from a training scenario. The application of the framework has extended its scope to the new domain of emergency management combined response. In addition, the preliminary model is used to identify tasks and behaviours of the combined response system which highlight issues of co-ordination .

Keywords
planning and control, multiple tasks, emergency management

INTRODUCTION
This paper presents an informal analytic assessment of a model of the emergency management combined response (EMCR) worksystem with respect to the data from a training scenario. The model was derived by applying a framework for modelling planning and control in multiple task work (PCMT) to the data from an emergency management training scenario. This framework has been applied to develop models of three office administration planning and control domains - Medical Reception (Hill, Smith, Long and Whitefield 1995) -Secretarial Office Administration (Smith, Hill, Long and Whitefield 1992) and Solicitors Legal Work. The intention here was to extend the scope of the PCMT framework by using it to model a domain other than office administration. The first section of the paper characterises the domain to be modelled, identifying issues with the current design of the EMCR worksystem. The following section outlines the PCMT framework. Next, the data collection, training scenario, and example data are described. The model of PCMT-EMCR is then presented. Using the model to identify overlap between worksystem behaviours is exemplified. Last, issues related to the use of this framework for modelling the domain of EMCR are identified, along with plans for future work.

EMCR WORKSYSTEM
The planning and control system, which is set up for emergency response to disaster, is that of the 'combined' response (EMCR). The EMCR worksystem has a command and control organisation with a three tier structure. Different agencies use this structure to organise their own planning behaviours, so that they co-ordinate effectively with one another. The three levels of command and control are operational, tactical and strategic. At each level the agencies have their own commander for co-ordinating the response. At the strategic level, these commanders make up a 'senior co-ordinating group'. The operational response is carried out by each agency, concentrating on its specific roles (tasks) within its areas of responsibility. For example, the Fire Service fight fires. The tactical response determines the priority in allocating resources, for example additional fire engines. It also plans and co-ordinates the overall response, obtaining other resources as required. The strategic co-ordinating group formulates the overall policy within which the response to a major incident is made. This EMCR worksystem has been specified to ensure better co-ordination of the different agencies involved in the response. Co-ordination in this context is defined as the 'harmonious integration of the expertise of all the agencies involved with the object of effectively and efficiently bringing the incident to a successful conclusion.'(Emergency planning college document 1995). The principal agencies involved are the Police, Fire Service and Ambulance Service. The primary objectives for all the emergency services are: 'to save life, prevent escalation of the disaster, to relieve suffering, to facilitate investigation of the incident, safeguard the environment, protect property and restore normality' (Home Office Publication 1994).

However, a succession of public enquiry documents analysing various disasters have identified co-ordination problems within EMCR. The Home Office Emergency Planning Research Group has also identified problems relating to co-ordination within EMCR (see acknowledgements).

Each emergency service has its own major incident plan. Each plan specifies a set of roles/tasks that need to be supported, when responding to a disaster, for example, evacuation, setting up a casualty clearing station etc. Some of the tasks, at the most generic level of description, are the same for the three emergency services, for example the Police and the Fire Service both support evacuation tasks. Most of the behaviours which support these tasks for the two services are different, but some are the same. Sometimes the behaviours overlap between the different services,

when they are supporting different tasks. The Fire Service may be attempting to set up an inner cordon to contain a hazardous scene; the Ambulance Service may be attempting to access the same scene to locate casualties. It is proposed here, that when behaviours overlap, co-ordination between the services may become a problem. (Behaviours do not necessarily have to be the same to overlap.)

In developing the present preliminary model of EMCR with respect to a training scenario, it is intended to identify such overlapping behaviours to support better the identification of the co-ordination problems raised by the system.

FRAMEWORK

A framework was developed for modelling the planning and control of multiple task work (PCMT) in office administration. The PCMT framework makes a fundamental distinction between an *interactive worksystem*, comprising one or more users and computers/devices, and its *domain of application*, constituting the work carried out by the worksystem. This distinction allows for an expression of the effectiveness (performance) with which the work is carried out, conceptualised as a function of two factors: the quality of the task (i.e. whether the desired goals have been achieved), and resource costs (i.e. the resources required by the worksystem to accomplish the work) (Dowell and Long 1989).

In the framework, a domain of application (or work domain e.g. the disaster) is described in terms of *objects*, which may be abstract or physical. Objects are defined by their *attributes*, which have *values*. The attribute values of an object may be related to the attribute values of one or more other objects. An object at any time is determined by the values of its attributes. The worksystem performs work by changing the value of domain objects (i.e. by transforming their attribute values) to their desired values, as specified by the work goal.

The framework defines a number of worksystem structures for the planning and control of multiple task work. These structures are expressed at both abstract and physical levels of description. First, the framework describes the worksystem's abstract cognitive structures. These structures comprise four processes (*planning*, *controlling*, *perceiving* and *executing*), and two representational structures (*plans* and *knowledge-of-tasks*). The four processes support the behaviours of planning, control, perception and execution respectively. A complete description for this set of structures, and the associated rationale, is to be found elsewhere (Smith, Hill, Long and Whitefield, 1992).

At the first (abstract) level of description, *Plans* are specifications of required transformations of domain objects and/or of required behaviours (to achieve goals). They may be partial (in the sense that they may specify only some of the behaviours or transformations). They may also be general (in the sense that some behaviours or transformations may be specified only generally, and not at a level which can be directly executed). Planning behaviours thus specify the required transformations and/or behaviours to support domain object transformations.

Perception and *execution* behaviours are, respectively, those whereby the worksystem acquires information about the domain objects and those whereby it carries out work which changes the value of the attributes for those objects as desired. Information about domain objects from *perception* behaviours is expressed in the *knowledge-of-tasks* representation. *Control* behaviours entail deciding which behaviour to carry out next, both within and between tasks.

The second level of description of planning and control structures is physical, wherein the framework describes the distribution of the abstract cognitive structures across the physically separate user and devices of particular worksystems.

This framework was chosen to model the combined response worksystem because:

(i) the EMCR worksystem is a planning and control system;

(ii) the work involves multiple tasks; and

(iii) there is a need to identify the tasks and behaviours of the worksystem in relationship to the work to better identify co-ordination problems within the EMCR worksystem

DATA COLLECTION

Data were collected by means of a training scenario which took place at the Home Office Emergency Planning Training College. The trainees were members of the emergency services and local authority emergency planning officers. The exercise required the trainees to describe their response to the disaster scenario from initial response to restoration of normality. Data were collected by transcribing the descriptions given by the trainees, and through presentations given by the trainees about their response.

There are various phases in response to the disaster scenario, each of which can be clearly defined, and which involve different worksystem configurations and different worksystem behaviours. In this paper, only one worksystem configuration and its behaviours is modelled- that of the initial response phase to the scenario.

Scenario

'At 9.30am on a weekday during school term time, a tanker train en-route from a refinery to an airport fuel depot is derailed whilst passing over a railway bridge. The railway bridge carries the railway over the main access road to a town from east to west. The main access road bisects the town in a north/ south direction. There are market stalls set up on the roadway on either side of the bridge. At the time of the accident, a tourist' bus carrying 45 foreign tourists is passing beneath the railway bridge. One of the tank cars is ruptured during the derailment and aviation fuel flows down the sides of the embankment onto the roadway. Flammable vapours from the fuel have been ignited by an open gas burner from a catering caravan. The explosion has created severe structural damage in a 30 metre radius and moderate structural damage in 100 metre radius. At least 50 people have been killed including some of the foreign tourists, many people have received burns, and many people are trapped. A number have also been contaminated with aviation fuel. The leaking fuel has run down the road

and is entering the canal and watercourses at the bottom of the incline. There is evidence to believe that vandalism may be responsible for the derailment.'

DESCRIPTION OF THE INITIAL COMBINED RESPONSE DATA

A major incident was declared within 15 minutes of assessing the situation. Thus, all the emergency services initiated their major incident plans. Each service has a set of defined roles or tasks specified in their individual plans. For this scenario, examples of initial Police roles/tasks were:

- setting up and manning an outer cordon around the site (a larger area than the scene).
- preserving and managing the scene
- logging all personnel entering the outer cordon
- establishing routes and rendezvous points and marshalling areas for all the emergency services

Examples of the initial Fire Service roles/tasks were:

- set up an inner cordon
- control the fire
- carry out immediate rescue
- Stem the flow of the hazardous substance
- Monitor the health and safety of all inside the inner cordon

Examples of the initial Ambulance Service roles/tasks were:

- set up a casualty clearing point
- set up a triage point
- get medics to the scene
- help with evacuation of bedridden people

For each of the tasks the personnel required were identified by, and are shown in, the model.

PCMT-EMCR MODEL

The model describes the EMCR worksystem and its domain for the initial response scenario data. The initial response in this scenario had no strategic level of command. Thus, although the strategic level physical worksystem users and devices are represented in the model, they are included for completeness. No further description is offered. Also, the operational physical users and devices change rapidly over time. The model represents only those entities present for initial response. The local authority response has not been modelled.

EMCR Domain

Based on the PCMT framework, the EMCR domain is expressed as those objects, whose transformation constitutes the work of the EMCR worksystem.

In the model, the domain is conceptualised as having a single disaster object comprising other abstract objects, such as lives, property, environment etc. The disaster object and its component sub-objects are abstract. The sub-objects of the domain are based on the identified primary objectives of EMCR (see earlier). The domain in addition contains physical objects.

The work carried out by the EMCR worksystem is thus the transformation of a 'disaster object's' attribute values. Each task carried out by the EMCR worksystem transforms the attribute values of the disaster object. For example, one attribute is stability, which has values along a continuum (i.e. very unstable, slightly unstable, unstable, nearly stable, moderately stable, stable). The work of the EMCR worksystem is to transform this attribute to a desired level of stability. In order to transform this attribute, the attributes of the component sub-objects must be transformed. This attribute value changes by manipulating the values of the attributes of the other objects of the domain, for example, transforming the 'lives object' attribute *survivor* from not rescued to rescued. The transformation of the abstract object attributes result from the manipulation of the physical domain objects of the worksystem. Attributes may be affordant or dispositional. Affordant attributes are transformed by the worksystem; their transformation constitutes the work performed. Dispositional attributes are relevant to the work (they need to be perceived by the worksystem, but are not changed by the worksystem). For example for the 'lives object' the attribute *survivor* has a value of mobility injured or mobility uninjured. This value needs to be perceived by the EMCR worksystem, so that the Ambulance Service knows there is a casualty to be transported. However, the worksystem does not change the injured state of the survivor. In Figure 1, all the starred attributes are dispositional.

Figure 1 shows the physical domain objects (but not their attributes) in the model. Included are all the objects identified from the scenario data. The objects include those associated with the example which follows.

EMCR Worksystem

The model of the EMCR worksystem (see Figure 1) shows the cognitive structures of the PCMT framework. The cognitive structures in Figure 1 embody the planning and control behaviours described earlier. The framework comprises four process structures: planning; controlling; perceiving; and executing and two representation structures: plans and knowledge of tasks. For EMCR, these cognitive structures were identified as follows:

A **perceiving** process which acquires information about property, lives and other domain objects, such as their risk status. This information appears in the **knowledge-of-tasks** representation. A task in the EMCR will be different for the different agencies. A task for the Police could be setting up a casualty clearance point. A task for the Fire Service could be making sure all personnel are safe within the inner cordon.

An **executing** process which transforms domain object attribute values. For example, moving injured survivors away from the scene transforms the 'lives object' attribute *survivor* from not rescued to rescued.

A **controlling** process which decides which of the other processes should be carried out next based on the **plans** *and* **knowledge-of-tasks**. For example the **controlling** process might direct the **executing** process to perform next the execution behaviours of evacuating people (based on the **perceiving** process that identified where people were,

and the **knowledge-of-tasks** information that people in this situation are at risk), rather than directing the **planning** process to specify how to preserve the scene. (Thus, there is prioritisation of the tasks carried out by EMCR.)

A **planning** process, which constructs plans based on knowledge-of-tasks. For example using information from **knowledge-of-tasks** that the scene needs to be preserved for investigation, the **planning** process constructs a plan of the sequence of behaviours to carry out this task.

The **Plans** representation structure embodies the different types of plan used in the combined response.

A **knowledge-of-tasks** structure, which represents knowledge of relevant aspects of the work domain. For example, the type of hazardous substance.

The physical worksystem comprises the Police and their devices, the Fire Service and their devices, the Ambulance Service and their devices (plus other agencies and their devices which may be involved). The framework allows the construction of alternative models of the distribution of cognitive structures across the user and devices. Thus, it supports reasoning about allocation of function. In the case of the EMCR worksystem, planning and control are distributed across the different levels of the worksystem hierarchy. In the model, the distribution of the structures across the physical worksystem is not shown. However, these physical worksystem structures support its behaviours. In the example which follows, the worksystem behaviours are identified for a selection of data from the training scenario to illustrate this distribution across the worksystem.

Relationship between the Domain and the Worksystem

This section describes an example of two tasks being carried out by the combined response system, in terms of the planning, control, perception and execution behaviours and the transformations these behaviours perform in the domain. This description helps to identify potential conflicts within the combined response system behaviours which may cause co-ordinatio oper safety requirements which would allow them to enter the inner cordon. Therefore, the execution behaviours of transporting casualties cannot be carried out. The primary objective of EMCR is to save life. Here, there is an overlap of behaviours which is hindering this objective. Also, if the Ambulance Service is not allowed to carry out its execution behaviour, in trying to increase the saving of life, the fire service must carry out rescue execution behaviours to move the casualties to the edge of the inner cordon. These rescue execution behaviours will decrease the resources available for carrying out the execution behaviours of controlling the hazard, thus decreasing the effectiveness of the response to the secondary objective of preventing escalation of disaster

This section describes an example of two tasks being carried out by the combined response system, in terms of the planning, control, perception and execution behaviours and the transformations these behaviours perform in the domain. This description helps to identify potential conflicts within the combined response system behaviours which may cause co-ordinatio oper safety requirements which would allow them to enter the inner cordon. Therefore, the execution behaviours of transporting casualties cannot be carried out. The primary objective of EMCR is to save life. Here, there is an overlap of behaviours which is hindering this objective. Also, if the Ambulance Service is not allowed to carry out its execution behaviour, in trying to increase the saving of life, the fire service must carry out rescue execution behaviours to move the casualties to the edge of the inner cordon. These rescue execution behaviours will decrease the resources available for carrying out the execution behaviours of controlling the hazard, thus decreasing the effectiveness of the response to the secondary objective of preventing escalation of disaster

To attempt to rectify this situation, the Ambulance senior officer informs the Ambulance incident officer of the inner cordon regulations. Thus, the Ambulance incident officer carries out perception behaviours that update his knowledge-of-tasks about the inner cordon regulations. He then carries out control behaviours that direct him to carry out planning behaviours to specify in the plan the required equipment for all resources directed to the scene. Until these regularised resources arrive, the Ambulance incident officer needs to communicate with the Fire Service incident officer to try and negotiate access for the operational personnel. The Fire Service incident officer should have informed the Ambulance incident officer of the regulations, to obviate the problem of co-ordination.

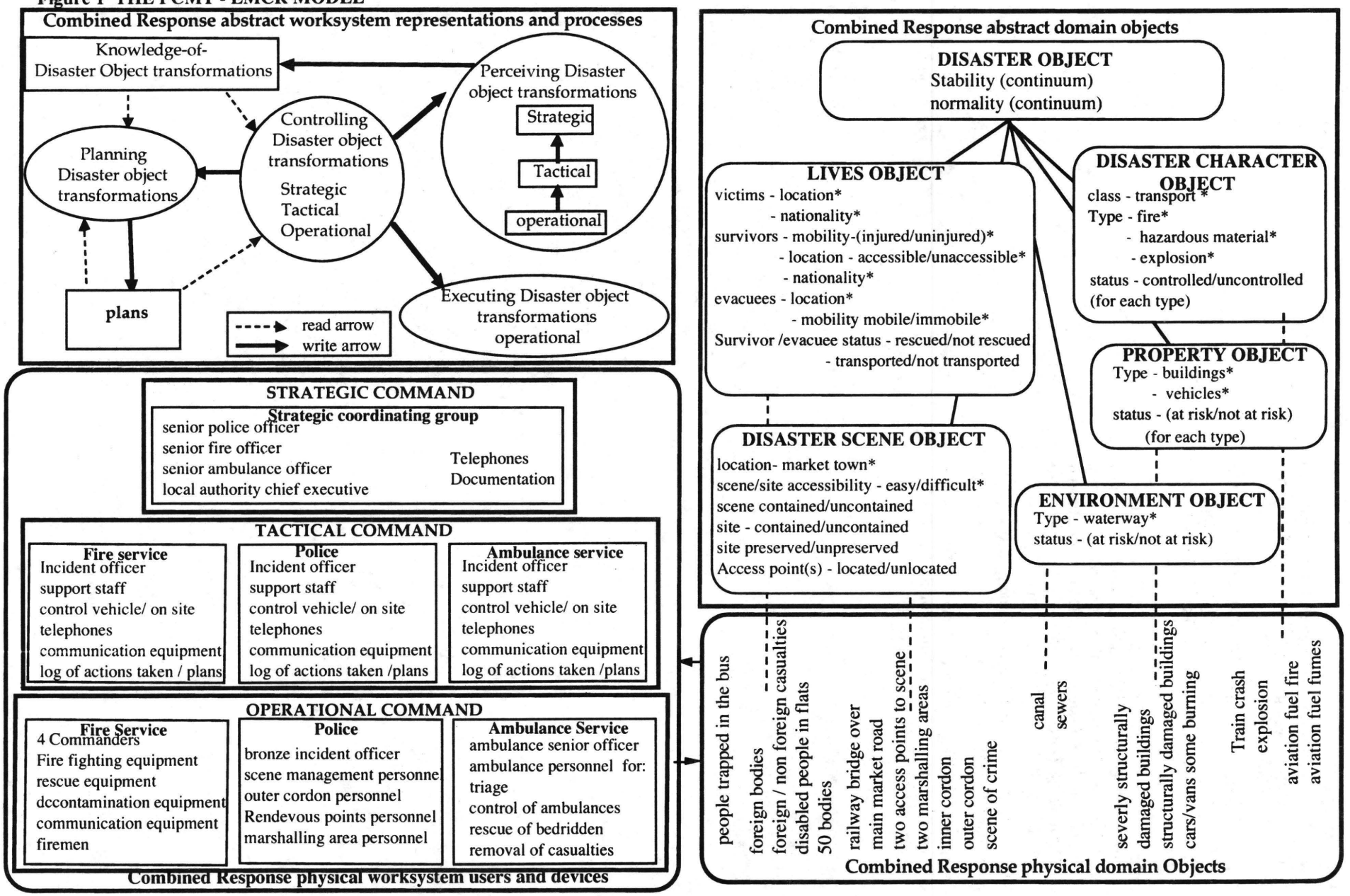
Conclusions

This paper has described the application of the PCMT framework to modelling the domain of Emergency management combined response (EMCR). A preliminary model has been developed based on data from a training scenario. This final section will discuss issues related to using the PCMT framework for modelling this domain, and the use of the model for identifying overlapping behaviours that highlight co-ordination issues within EMCR.

The PCMT framework is for use in planning and control for multiple task domains. EMCR is such a domain, but there are differences between this domain and the other domains already modelled. First, EMCR has a changing worksystem. The PCMT representation does not currently represent a changing worksystem. Thus, the preliminary model only represents one phase of response. Second, EMCR is multiple task, but unlike the previous domains studied, there are not multiple objects at the highest level i.e. multiple disaster objects, rather there are multiple sub-objects, such as multiple lives objects. Third, EMCR is made up of multiple agents within a complex three tier command structure. The PCMT framework has so far only modelled domains with a single level of operation. Thus, interactions between the different horizontal layers and different vertical layers of the system are difficult to describe in terms of the present framework.

The difficulty can be appreciated in the example, where the different commanders liase, at the same horizontal

Figure 1 THE PCMT - EMCR MODEL



REFERENCES

Dowell, J. and Long, J. (1989) Towards a conception for an engineering discipline of human factors. *Ergonomics, 32,* 1513-1536.

Hill, B., Long, J.B., Smith, W. and Whitefield, A.D. (1995) A model of medical reception - the planning and control of multiple task work, *Applied Cognitive Psychology,* 9, 81-114.

Smith, M.W., Hill, B., Long, J.B. and Whitefield, A.D. (1992) Modelling the Relationship Between Planning, Control, Perception and Execution Behaviours in Interactive Worksystems. In D.Diaper, M.Harrison and A.Monk (Eds) *People and Computers VII; Proceedings of HCI '92.* Cambridge University Press.

Dealing with Disaster 2nd Edition (1994). Home Office Publication HMSO London.

Emergency planning college document (1995) Management levels in response to a major incident - The Principles of Command and Control.

Artefact Usage & Conversation Patterns At Meetings

B.P. Bright & N. A. M. Maiden
Centre for Human-Computer Interface Design
City University, London
Tel:+44-171-477-8994
E-Mail: [B.P.Bright,N.A.M.Maiden]@city.ac.uk

ABSTRACT
Current Requirements Engineering (RE) practice is not well-understood. One reason for this is the lack of useful studies of "good", as well as "not so good" practice. As such it is difficult to design new methods and tools to improve on poor practice and encourage good practice. This paper reports (1) a predictive model of canonical patterns of communication and artefact usage in RE meetings and (2) findings from an empirical study of two meetings. The model is anticipated to provide a basis for new methods and tools to aid planning and facilitation of RE meetings.

Keywords
meeting, artefacts, conversation patterns, meeting facilitation, requirements engineering

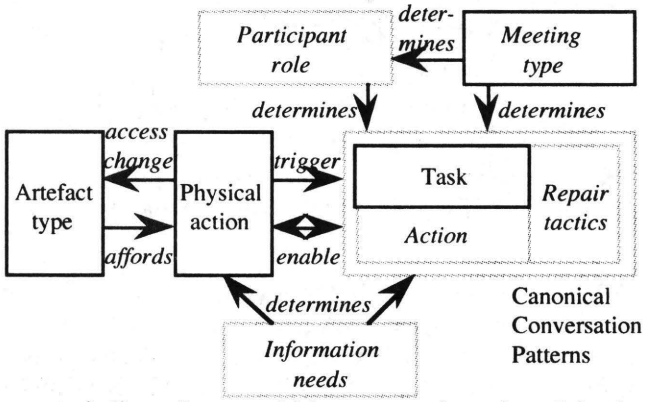
INTRODUCTION
Meetings are important as they are at the core of the way most organisations do business and are used for virtually any intellectual task that requires co-ordination or agreement of several people. Studies have shown that office workers spend as much as 30-70 percent of their times in meetings (Stefik et al, 1987). Requirements engineering (RE) is the complex social, cognitive and technical process during which different groups of requirements engineers, users and business managers (e.g. Curtis et al. 1988) produce requirements for a software-intensive system (Jarke et al. 1993). These groups learn, communicate, negotiate, acquire and agree requirements for the new software system. RE was chosen as the domain area as many RE decision-making tasks such as Requirements Acquisition and Validation often occur at meetings. This framework focuses on meetings in the same time and place, since most RE meetings will be of this type for the foreseeable future. Furthermore, document-type artefacts are important in such meetings, as most communication between people during requirements engineering is artefact-based (Kensing & Munk-Madsen, 1993). People often communicate through documents rather than face-to-face, even in simple meetings. Studying meetings is not new and meeting activity has been categorised previously (e.g. Olson et al, 1992, Saeki et al 1996). However, recurring communication patterns with associated artefact usage in meetings, has received little research attention and therefore we wish to explore this in order to inform design of better planning and managing of meetings.

This paper first outlines models of behaviour for various RE meeting tasks. Second, an empirical study of two meetings for validating the models, is described. The results from the empirical studies are then discussed. Finally, future work is outlined.

MODELS OF RE BEHAVIOUR
Requirements Acquisition was one of the tasks researched. The literature suggests that this task may be comprised of several sub-tasks including brainstorming and ideas structuring (Takeda et al 1993), creating a list of objects by brainstorming (Macaulay 1993), and evaluation and agreement on a list of objects (Macaulay, 1993).

Our model asserts the existence of canonical patterns of communication between participants during these common tasks and predicts recurring communication patterns during a task in a meeting, and how different types of artefact influence this communication see Figure 1. These artefact types include flipcharts, whiteboards, projector slides, overhead projector transparencies, video tape/films, audio tapes, A4-size paper documents and simple electronic artefacts, all of which are often available in an office environment. The model claims that the canonical communication patterns in a meeting are influenced by (at least) meeting type, the formal role of participants in the meeting, information needs of these participants and the types of artefact in use.

Figure 1: overview of predictive model. Greyed-out components indicate factors which were not investigated in the studies reported in this paper



We have developed a first version of the models using the Conversation for Action model (Winograd & Flores,

1987) and drawing on theories of speech acts (Searle, 1969) since many RE activities involve face-to face verbal communication and can be modelled as conversations. The models were populated using guidelines in the literature. Some of the models are:

- Walkthrough in validation: based on Yourdon (1978);
- Requirements Acquisition: based on general interview techniques e.g. Yourdon (1988);
- Presentation: based on presentation techniques, Norman & Draper (1986), socio-emotional and task specialists behaviour Wilke & Knippenberg (1988), direct action Westley & Waters (1988);
- Negotiation where people come to a meeting to discuss their conflicting viewpoints: based on Sycara & Lewis (1991);
- Argument that can occur at any meeting, needing facilitator intervention: based on Westley & Waters (1988).

Figure 2 shows an initial version model for acquisition. Questions are categorised into those which seek information (e.g. elaborate, solution, enable, example), explanation (e.g. background, cause, result, example) and confirmation (e.g. yes, no). Replies are divided using a similar scheme. Such general categories of speech act were identified to minimise over reliance on conversation content and to reveal meaningful interactions between patterns of artefact usage and speech acts.

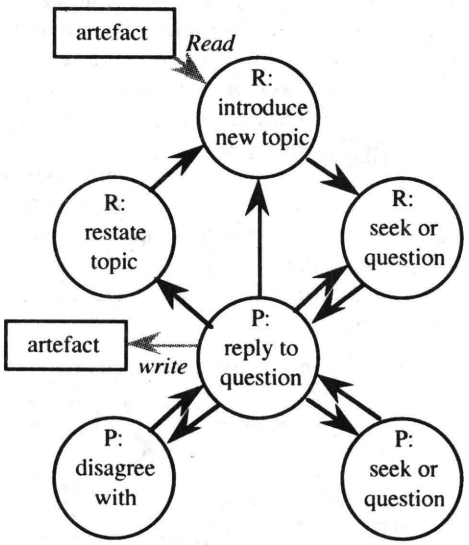


Figure 2: canonical pattern of conversation anticipated during acquisition (R: requirements engineer, P: other participants)

Empirical Studies of RE Meetings

This and other models provide hypotheses for empirical studies of RE behaviour. We need to understand RE processes and to make RE expertise activities and behaviour available. Existing models (e.g. Rolland, 1994) describe method guidance rather than actual behaviour. Given that practitioners tend not to follow such methods these process models cannot be relied on. In contrast, our work presents predictive models of requirements engineering behaviour founded on empirical observation of current practice. From the models we aim to inform the selection of speech, artefacts to use, and how facilitators should intervene at various meeting stages, to inform design of how best to conduct an RE meeting.

A meeting room was set up to observe and evaluate RE meetings. In the first meeting, the two stakeholders were from Personnel, City University, whereas in the second meeting, two members of the Library, City University. For both meetings, an RE consultant, who was to act as facilitator for that meeting, was brought in from 'outside' and had not met the participants previously, although, the facilitators had communicated with the participants by E-mail.

At both meetings, participants sat around a table and discussed their system's current problems and requirements for a new system. Each meeting was two hours in total. To aid the task and communication between people, whiteboard, flipchart, video, audio tape, paper, simple electronic (computer), projector and slides, and overhead projector and slides were provided. The meeting was recorded by video, which was to be focused on all participants, and audio tape. After the meeting, a retrospective questionnaire was used to question the facilitator about RE meetings they attend in general, and also about that particular meeting e.g. their requirements acquisition technique. The same Room layout, used for both meetings, is shown in Figure 3.

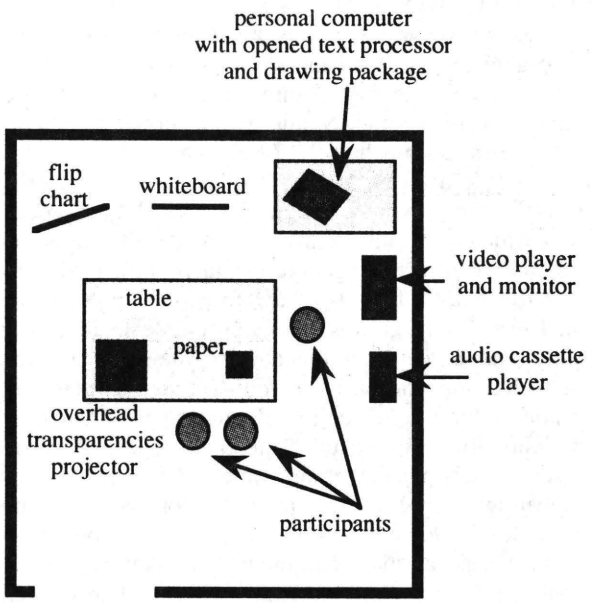


Figure 3: meeting room layout: all people were able to move around the room during the meeting

Results

Both personnel and library staff expressed general levels of satisfaction with the meeting in a retrospective questionnaire (all rated the usefulness of the meeting as either 6 or 7 on a scale of 1-7, where 7=very useful). The basic sequences of both meetings are described.

Meeting-1: the meeting adhered to the agenda set by the requirements engineer. The requirements engineer asked questions prompted by the agenda. During the next 30 minutes the requirements engineer prompted the two participants for current system problems which were written on overhead projector transparencies. A data model was also drawn on the whiteboard. The next hour was spent defining system requirements which were also written on overhead transparencies. The requirements engineer also wrote on the flipchart and sometimes changed the data model. During the last 20 minutes all requirements were prioritised by ranking them on new overhead transparencies. There was little observed disagreement between participants throughout.

Meeting-2: this meeting's focus and tasks were similar to those of meeting-1, however artefact usage differed. First, no agenda had been set. Rather, the meeting began with a short discussion about a system description sent by e-mail to the requirements engineer beforehand. In the next 40 minutes the requirements engineer prompted the participants for current problems which he then wrote on different pieces of flipchart paper stuck to the meeting room walls. A data flow diagram (DFD) was also drawn on the whiteboard, and the requirements engineer referred back to the original paper description of the system. The requirements engineer also used overhead projector slides to give a short presentation. More sheets of flipchart paper were stuck to the walls and written on during requirements definition. Another short presentation using overhead projector slides was given. During the prioritisation task the flipchart sheets were changed to indicate the importance of each requirement. All existing sheets and the DFD on the whiteboard were referred to throughout. Again there was little observed disagreement between participants.

Artefact Usage

Retrospective questioning revealed that artefact usage throughout both meetings was planned rather than accidental. Table 1 summarises reasons for artefact usage elicited from the two requirements engineers during retrospective questioning.

Both requirements engineers used artefacts for specific roles (e.g. problem list, requirements list, data model). For example, prepared overhead projector slides were available for typical topics which arise in meetings (meeting-2). On the other hand whiteboards were more effective for modelling, "*...for a first-cut data model whiteboard is very good, because it does change, its very dynamic*". Furthermore, whiteboard was not used for permanent recording as "*.. but I really do like things to be able to be recorded, and with whiteboards it means somebody's got to copy it down and you run the risk of it..*". Flipcharts were seen as an informal medium of recording, and also permanent as flipcharts sheets could be carried away after the meeting as a permanent record (meeting-2). Both requirements engineers were against the use of computers in such meetings, unless in the hands of a non-participating meeting scribe. Reasons cited were facilitators felt it would create a barrier between themselves and the stakeholders. They also felt that participants would not be able to see the changes made on the computer due to the computer's small screen size. Furthermore, if the changed computer information also appeared on a large screen, the participants would be forced to look in two places. Both facilitators claimed that electronic technologies do have a role however, in their experiences, most technologies were inadequate. Making information visible to participants was considered important by both requirement engineers and was achieved by the use of overhead transparencies (meeting 1) and sticking flipchart sheets to the wall (meeting 2). Furthermore the facilitator in meeting 2 stated that sticking flipchart sheets to the wall "*made it look like work is being done*", "*something was happening*".

Art. type	*Requirements engineer in meeting-1*	*Requirements engineer in meeting-2*
F/C	Useful for recording environment Informal medium No need for lighting or technologies	Large and portable Can be stuck to walls Makes results visible
OHP	Useful for recording requirements but not environment Permanent medium	Prepared slides indicates special effort been made Limited to showing one at a time
W/B	Data model changes Not permanent medium	Anticipation mistakes during modeling Need for more whiteboards in general Larger whiteboards can enourage more groupwork
P/S	Room has to be darkened, inhibiting communication Useful for presentations	
A/T	Useful in large meetings to play relaxing music	No relevant audio tapes available
V/T	Room has to be darkened, inhibiting communication Useful for presentations	No relevant video tapes available. Useful learning device in teaching environment
SEA	Unnatural for drawing Useful for retrieving and many models Screen too small unless projection is possible Room has to be darkened, inhibiting communication	Useless unless scribe available Inhibits communication Screen size too small

Table 1: summaries of reasons for artefact usage elicited from RE during retrospective questioning: F/C = flipchart, OHP = overhead projector slides, W/B = whiteboard, P/S = projector slides, A/T = audio tape, V/T = video tape, SEA = simple electronic artefact

Speech Act Patterns & Associated Artefact Usage

Speech acts were identified from conversational utterances using categories derived from Mann & Thompson's (1988) scheme for text segments. These speech acts include:

- seek-detail: speaker requests information from another, including seeking elaboration;
- give-detail: speaker imparts information to another. Includes answering questions, elaborating on a topic;
- restate: repeats or summarises a current conversation, without adding new content to the conversation;
- action: meta-level statement about the planning and conduct of the meeting, for example "Let's write these requirements on the whiteboard".

We produced models of speech act patterns encompassing speech acts, physical acts and artefacts. Each speech act was either linked or not to a physical act (such as write or point) and the artefact involved (e.g. whiteboard, flipchart). Table 2 shows total numbers of speech acts linked to physical acts for both meetings, broken down according to physical act and artefact type. The most often-used artefacts (overhead projector slides, flipcharts and whiteboards) were both read and written on throughout both meetings. The distinction in the usage of flipcharts and flipchart sheets arose in meeting-2 when flipchart sheets were stuck onto the walls of the meeting room. Sheets on the flipchart stand were read less because sheets which had been written on were then turned over and were no longer visible until stuck onto the wall. Overall, more writing than reading was observed, which is not surprising given the acquisition nature of most of both meetings. There were also fewer changes to artefact contents than both requirements engineers had expected. This might have been due to the nature of the meeting, the number of participants, or both. Pointing to and marking information were also both important physical acts, especially for artefacts which were not to hand in the meeting room. Another important physical act was lining-up information. On 12 occasions that the requirements engineer in meeting-1 lined-up existing overhead projector slides on the table during prioritisation. This was to show to the participants the relative priorities of different groups of requirements which had already been prioritised by them.

	o/h	f/c sh.	f/c	w/b	ppr	o/h dis.	o/h tab.
Read	74	55	11	38	24	2	2
Write	152	80	85	63	0	0	0
Point	8	14	50	54	0	0	0
Change	0	0	0	2	0	0	0
Make visible	17	0	0	0	0	0	0
Put on wall	0	0	2	0	0	0	0
Pick Up	3	0	0	0	0	0	0
Line-up	12	0	0	0	0	0	1

Table 2: totals of artefact usage by artefact type for both meetings: o/h=overhead projector slides, f/c sh=single flipchart sheets on wall, f/c=flipchart sheets on flipchart stand, w/b=whiteboard, ppr=paper documents, o/h dis. =overhead projector transparencies displayed, o/h tab = overhead projector transparencies on projector screen

Combined patterns of artefact usage and speech acts were further categorised using one of the following behaviour categories:

(i) no artefact usage during speech act;

(ii) artefact usage unconnected to the speech act (e.g. conversation continues while something is written down);

(iii) the speech act triggers artefact usage (e.g. saying something causes it to be written down);

(iv) artefact usage triggers speech act (e.g. reading causes someone to say something).

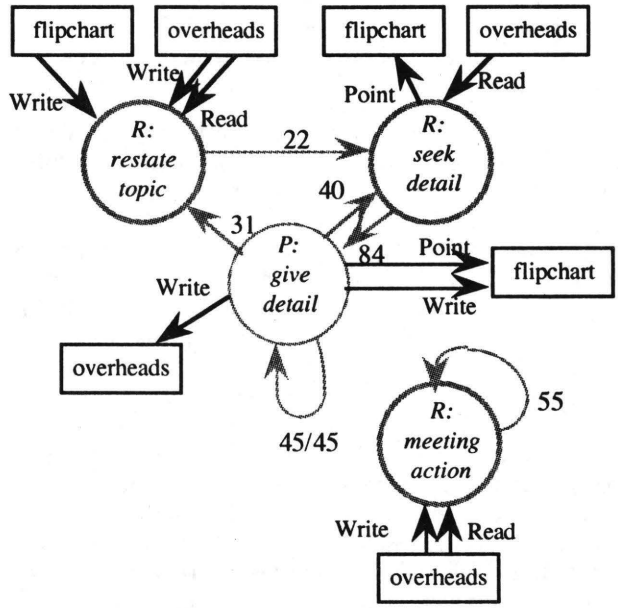


Figure 4: patterns of physical and speech acts during meeting-1, acquisition. Artefact types are shown in boxes. Frequencies greater than 2% of the total of all physical acts for this task are shown. Arrows show link between speech act and artefact type. Arrows also indicate the sequence of physical/speech act (i.e. speech act follows physical act or vice versa). Split totals of transaction acts indicate speaker changes. First total shows number of same-participant transitions, the second shows different-participant transitions

Furthermore, for each instance of the latter two categories, the artefact type (e.g. whiteboard, flipchart) and physical act (e.g. read, write, change, point) were identified. This analysis enables patterns of artefact usage to be superimposed on patterns of speech acts.

A combined pattern of speech and physical acts for acquisition during meeting-1 is shown in Figure 4. The core communication pattern shows the requirements engineer seeking details, participants giving details then the requirements engineer either restating the topic prior to asking another question, or asking another follow-on question straight away. One seek-detail act was often followed by multiple give-detail acts. Furthermore, both participants tended to give details when answering the requirements engineer's questions. Such give-detail acts from both participants were often interleaved. Indeed, the number of speaker changes between give-detail acts in both meetings was unexpectedly high. Most artefact usage occurred during the seek-give-restate cycle. Seek-detail acts were often triggered from reading overhead projector slides, and

in particular from pointing at information on flipchart sheets. Participant give-details were often followed by the requirements engineer writing the given information on both overhead projector slides (during problem identification and requirements definition) and the whiteboard (during data modelling). It was when writing down the problems and requirements that the requirements engineer tended to restate topics.

Patterns of speech and physical acts for acquisition during meeting-2 are shown in Figure 5. Similar patterns of seek-details, give-details were found. Also, as predicted, the participants sought-details from each other. However differences with the predictive model were also found. The requirements engineer tended not to restate topics. Again most artefact usage occurred during seek-detail and give-detail. Seek-detail acts were often preceded by the requirements engineer reading the paper document and reading and/or pointing at the whiteboard. Participant give-detail acts were often followed by writing this information on the whiteboard and the flipchart sheets.

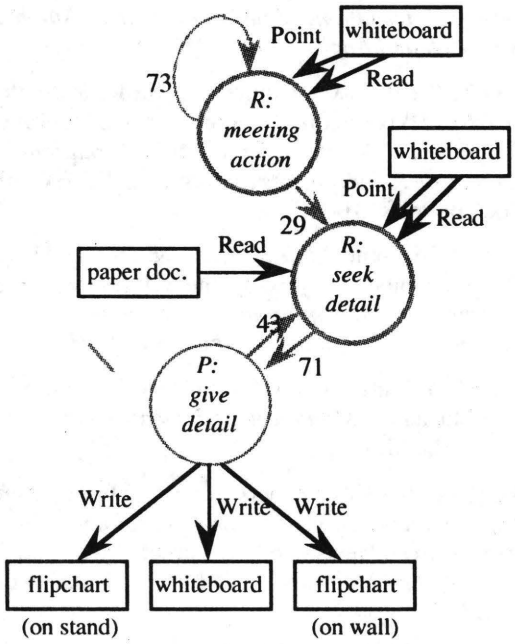


Figure 5: patterns of physical and speech acts during meeting-2, acquisition task. Frequencies greater than 2% of the total of all physical acts for this task are shown.

CONCLUSIONS

Reoccurring patterns of task, speech act, physical act and artefact usage were observed in both meetings. In particular:

- both meetings had a similar task structure;
- overall use of artefacts in both meetings was similar;
- artefact usage was planned rather than accidental and controlled during both meetings. Principle affordances for acquisition were the need for visible information, and the need to make information permanent in most cases;
- writing was the most common physical act, although reading, pointing and lining up information in artefacts were also important.

DISCUSSION & FUTURE WORK

Observation of both requirements acquisition meetings has implications for future research directions. First, both meetings show the acquisition task as a session whereby a facilitator predominately asked questions, generally to the group, and the participants replied. The studies reported here do not show cases of acquisition whereby a facilitator asks individuals one at a time, to relate their problems, but we have looked at these types of meetings. Second, the literature suggests that argument could occur during any meeting (e.g. Westley & Waters, (1988). However, this did not arise at either meeting, due probably to the small number of participants attending both meetings. Therefore further investigation is need, first to discover if argument occurs at a larger meeting, and second to see if the "argument" model will be interleaved with a requirement task. Furthermore, these factors imply that there may be more than one pattern for a specific task (e.g. acquisition with few participants; acquisition with many partcipants).

The physical and speech acts models derived from the meetings are specific to acquisition. Although the "Seek" and "Give" details pairing may be common to most tasks, some discussion is needed to explore if the models showing speech act pairings with associated artefact usage, can be generalised over other tasks.

Information content, such as conceptual models, list of Requirments or problems list, may play an important part in the choice of artefacts used at meetings, and needs to be explored.

Our aim is to better understand communication and artefact usage during complex meetings, to inform method and tool design. Observation of both requirements acquisition meetings has implications for future research directions. First, observed "good and expert" practice will be made available in the form of handbook that will:

- support planning of the meetings: aid facilitator to set up the meeting providing "the best" available artefacts for the task;
- support facilitation during the meeting: advise facilitator in which type of speech acts to use to achieve be assured of certain responses (certain speech acts). Additionally, at what stages to use certain artefacts, & what type of information is best suited to a type of artefact.

From the handbook, as discussed above, a Decision Support System (DSS) tool can be constructed to advise the facilitator in planning & during the meeting.

Additionally, a meeting tool will be constructed to simulate the work of a facilitator in a same time, same place meeting. It would be comprised of two parts, the simulation model and the tool architecture. The Simulation Model will consist of speech and physical act models

information, simulating, for example, the structuring of information from one task to another (e.g. from acquisition to prioritisation). It will advise the choice of artefact, based on task, anticipated associated speech acts, information need (e.g. easy to change, permanent) and the information content. This model will serve as both a stand-alone method, and an executable model to be utilised by the tool architecture to operate. The partcipants at the meeting will have no direct access to it directly.

The Tool Architecture will consist of facilities usual to computer-supported collaborative work and collaborative writing, such as version control, traceability support, coordination support, and private and public workspaces (e.g. Takahashi et al, 1996). It will be controlled by a scribe, acting as a facilitator, if a facilitator using "good and expert" practice was not available. Furthermore, the tool will provide computerised artefacts simulating natural artefacts to enable facilities such as informal working, and permanence of information, during a task.

More empirical studies are on-going. We are carrying out studies of two types, controlled studies and observation. Furthermore, we are going to carry out studies of other types of meetings, with different stakeholders. Additionally, there is a possible follow-up to the two studies reported here.

ACKNOWLEDGEMENTS

The authors wish to thank all participants, in particular the Atlantic Systems Guild, for their invaluable time and experience. Bernice Bright is funded by a studentship from City University's School of Informatics.

REFERENCES

Curtis B., Krasner H. & Iscoe N. (1988) 'A Field Study of the Software Design Process for Large Systems', *Communications of the ACM 31*, 11, 1268-1287

Jarke M., Bubenko Y., Rolland C., Sutcliffe A.G. & Vassiliou Y. (1993) 'Theories Underlying Requirements Engineering: An Overview of NATURE at Genesis', *Proceedings of IEEE Symposium on Requirements Engineering*, IEEE Computer Society Press, 19-31

Kensing F. & Munk-Madsen A. (1993) PD: Structure in the Toolbox, *Communications of the ACM 36*, 4, 78-85

Macaulay L. (1993) Requirements Capture as a Cooperative Activity, *IEEE International Symposium on Requirements Engineering*, San Diego, 174-181

Mann W.C. & Thompson S.A. (1988) Rhetorical Structure Theory: Towards a Functional Theory of Text Organisation, *Text 8*, 3, 243-281.

Norman D.A. & Darper S.W. (eds.) (1986) User Centred System Design: New Perspectives, *Human-Computer Interaction*, Hillsdale N.J., Lawrence Erlbaum Assoc.

Olson J.S., Olson G.M., Storrosten M., Carter M. (1992) Small Group Design Meetings: An Analysis of Collaboration, *Human-Computer Interaction 7*, 347-374

Rolland C. (1994) Modeling the Evolution of Artifacts, *Proc. 1st International Conference on Requirements Engineering*, IEEE Computer Society Press, 216-219.

Saeki M., Matsumura K., Shimoda J. & Kaiya H. (1996) Structuring Utterance Records of Requirements Elicitation Meetings Based on Speech Act Theory, *Proc. 2nd International Conference on Requirements Engineering*, IEEE Computer Society Press, 21-30.

Searle J.R. (1969) Speech Acts. An Essay in the Philosophy of Language, Cambridge University Press

Stefik, M., Foster G., Bobrow D.G., Kahn K., Lanning S., Suchman L. (1987), Beyond the Chalkboard: Computer Support For Collaboration And problem Solving In Meetings, *Communications Of the ACM 30*, 1, 32-47

Sycara K.P. & Lewis C.M. (1991) Modelling Group Decision Making and Negotiation in Concurrent Product Design, *International Journal of Systems Automation: Research and Applications (SARA) 1*, 217-238

Takahashi K., Potts C., Kumar V., Ota K. & Smith J.D. (1996), Hypermedia Support for Collaboration in Requirments Analysis, *Proc. 2nd International Conference on Requirements Engineering*, IEEE Computer Society Press, 31-40

Takeda N., Shiomi A., Kawai K. & Ohima H. (1993) Requirements Analysis by the KJ Editor, *Proc. 1st International Symposium on Requirements Engineering*, IEEE Computer Society Press, 98-101

Westley F & Waters J.A. (1988) Group facilitation Skills For Managers, *Management Education and Development 19*, 2, 134-43

Wilke H. & Van Knippenberg A. (1988) Group *Performance in Introduction to Social Psychology, a European Perspective*, Hewstone M., Stroebe w., Codol J-P, Stephenson G.M. (eds.), Ch. 14, Basil Blackwell ltd., 315-349

Winograd T. & Flores F. (1987) *Understanding Computers & Cognition*, Addison-Wesley

Yourdon E. (1978) *Structured Walkthroughs*, 2nd edition, Yourdon inc.

Yourdon E. (1988) *Modern Structured Analysis*, London, Prentice-Hall International.

Interruptions In Computerised Mental Work: User Interface Design Issues

Ivan Burmistrov and Anna Leonova
Moscow State University, Department of Psychology, HCI Laboratory
8-5 Mokhovaya Ul.,
K-9 Moscow, 103009 Russia
E-mail: ivan@ht.cogsci.msu.su, aleon@chair.cogsci.msu.su

BACKGROUND

In this paper we present results of an exploratory experimental study of the effects of interruptions on the execution of computerised clerical tasks, provide classification of interruption handling strategies, and give some recommendations for the user interfac design for frequently interrupted work conditions.

The experimental study was conducted jointly by HCI Laboratory of the Moscow State University (Russia) and Work and Organization Research Centre, Tilburg University (The Netherlands). Here we report results obtained by Russian team.

Interruptions were chosen as a topic for study for a number of reasons. First, due to the fact that computerised mental work places high demands on the cognitive system, it is likely that this type of work is very sensitive to interruptions. Secondly, interruptions appear to be typical for the working conditions of many occupations in which computerised mental work predominates (e. g. office workers, secretaries, all kind of experts), which makes it interesting to find out how people deal with them while carrying out their duties. Thirdly, interruptions may negatively affect the person's state and performance, thereby exerting an influence on workers' well-being and productivity. The ultimate goal of the study was to propose cognitive engineering solutions to the user interface of computer systems aimed at preventing negative consequences of interruptions.

RESEARCH QUESTIONS

(1) How do interruptions affect the performance (quantity and quality)? (2) What interruption handling strategies do interruptions invoke? (3) What are the implications of the use of interruption handling strategies for subsequent performance?

Design Of The Research

31 subjects performed a computer-assisted task (text editing) bearing high similarity to real-life office tasks during two pairs of experimental sessions on two consecutive days. The experimental task was to make corrections in a computer file, based on a hard-copy version of a text containing hand-written corrections. During the experimental sessions subject's work activity was disturbed by a number of interruptions (phone calls). During the telephone call the subject was told to perform another task, referred to as "interrupting task". The subsequent analyses have been mostly based on the videotaped experimental transcripts.

The independent variables were the presence/absence of interruptions and the complexity of interruptions (two levels of interruption complexity were investigated). Performance characteristics such as editing latencies (times to complete the editing task or specific editing operations such as typing in new text or moving a paragraph to a new location), number of errors, subject's state, mental effort and subjective feelings were the dependent variables. Personality characteristics have been taken into the analysis as moderating variables.

RESULTS

Statistical analyses revealed the significant effects of both presence of interruptions and interruption complexity on the editing latencies for complex editing operations (e. g. moving a paragraph to a new location), while the performance indices for simple editing actions (e. g. typing in a new paragraph) were not affected by interruptions. Further analysis showed that increase in editing latencies is mostly caused by (a) additional compensatory activities (referred to as "strategic activities") invoked by interruptions and directed at either immunisation (taking away the influence of the disturbance) or recovery (resuming the work activity at an appropriate point), and (b) orientation in the primary task after completing interrupting secondary task.

Statistical answer to research question (3) will be ready for presentation at the time of submitting camera-ready copies of the paper.

INTERRUPTION HANDLING STRATEGIES

An empirical classification of interruption handling strategies (for text editing tasks) has been proposed:

[0] No attempts to continue main task after phone ring (full switch to the secondary task)

[1] Attempts to complete the current operation:

- [1A] Before picking up the phone
- [1B] During the secondary task (switch to main task after some work on secondary task)
- [1C] In parallel with secondary task (e. g. subject speaks to the experimenter and edits the text at the same time)

[2] Activity of memorising the current state of the main task:

[2A] Solely visual memorisation

[2B] Use of software environment (e. g. subject positions mouse cursor on a 'Paste' icon after cutting a paragraph, and then switches to the secondary task)

[2C] Use of hardware environment (e. g. subject positions and keeps his finger on a keyboard in order to memorise the action he should do next, e. g. <PgUp> or <Delete>)

[3] Activity on preventing possible errors:

[3A] 'Undo Cut' or 'Paste' into inappropriate place (i. e. subject returns the text block from clipboard to the screen in order to avoid possible loss of that block)

[3B] Deselecting of selected text (situation with selected block considered "dangerous", because occasional press of any key may replace selected text with that key)

RECOMMENDATIONS:

Contemporary office software systems do not provide users with sufficient support of interrupted work. Based on the analysis of interruption handling strategies and we have developed some recommendations for the user interface design for frequently interrupted work conditions:

1. Interface should give the possibility to instantaneously "freeze" the current state of the system in order to prevent occasional damages of the information while working on the interrupting tasks.

2. Interface should implement complex operations (functional threads) as single operations and provide efficient on-line assistance to the user in performing complex operations (e. g. "Select -- Cut -- Find new position -- Paste", the sequential group of actions for moving a paragraph to a new location should be organised as a single command "Move paragraph" invoking an appropriate "wizard" which guides the user in performing necessary steps).

3. Frequently interrupted work conditions require more apparent indication of the presence of information in the clipboard (e. g. in a small floating window). Otherwise users may lose clipboard information when their attention is occupied by interrupting tasks. New type of the 'Paste' operation, 'Paste and Clear Clipboard' command, purging information from the clipboard, is also suggested. This would logically symmetrize the conventional triad of 'Copy', 'Cut' and 'Paste' commands.

4. Metaphor of "cooling down text" or "drying up ink" is introduced. It is recommended to use colour coding for indicating recently changed or inserted information on the screen, colours being "cooling down" in course of document-specific time from hot colour (most recently changed information) to cold colour (old or unchanged information).

Individual differences and navigation in hypermedia

Kristina Höök
SICS
Box 1263
S-164 28 Kista, Sweden
tel +46 (0)8 752 15 00
fax +46 (0)8 751 72 30
kia@sics.se
http://www.sics.se/~kia/

Nils Dahlbäck
Dept. of Computer Science
Linköping University
583 03 Linköping, Sweden
tel +46 (0)13 28 16 64
nilda@ida.liu.se
http://www.ida.liu.se/~nilda/

Marie Sjölinder
Dept. of Computer and Systems Sciences
Stockholm Univ. / KTH
Electrum, 164 28 Kista
Sweden
+46 (0)8 16 16 00
hazze@frekvens.se

ABSTRACT
This study derives from two important observations. First, that navigation in hypermedia is a difficult task, and second, that individual cognitive differences play a role in how well users are able to efficiently use computer systems.

We found that of the cognitive abilities tested in our study, only spatial ability could be related to the time spent in completing a set of tasks in a large, hypermedia, information structure. We furthermore found that it was only certain aspects of spatial ability which were related to the ability navigate in hypermedia, namely those related to solving spatial problems mentally rather than solving spatial problems in the physical world.

We discuss how our findings may influence the design of hypermedia systems.

Keywords
hypermedia, spatial ability, human-computer interaction

INTRODUCTION
This study derives from two important observations. First, that navigation in hypermedia is a difficult task, and second, that individual differences play a role in how well individual users are able to navigate in other circumstances.

Navigation in large hypermedia information structures has long been recognised as a difficult and demanding task. The expression 'lost in hyperspace', (Conklin, 1987), describes what happens when users loose track of where they are and where to go next in the hypertext (see for example (Nielsen, 1990)). Navigation in hypermedia structures is similar but not equal to information retrieval in general.

Egan, (1988), shows that the best performer will solve tasks 20 times as quickly as the worst performer. This should be compared to "normal" tasks where a good performer does about twice as good as the worst. There are some studies on information retrieval behaviour that try to pinpoint the reasons why there are so large individual differences in terms of time spent and number of errors made. Borgman, (1989), provides an overview of such characteristics which have shown to be related to information retrieval behaviour. In most studies, we find that the users *previous experience* of the target system or computers in general, is among the most important factors that influence the users performance (Elkerton and Williges, 1984; Benyon and Murray, 1993). In addition, Egan identifies a cluster of factors which he names *technical aptitude* that is correlated with search behaviour, (1988). Technical aptitude includes spatial and reasoning abilities and background (as coursework in math and science).

Other factors mentioned by Borgman are *age*, *sex* and *personality*. Increasing age can be correlated with number of errors in information retrieval tasks (Borgman citing Greene at al. 1986). Sex is not related when experience and coursework are controlled, but as more males enter math and science education, they tend to have higher technical aptitude. Finally, Borgman shows that personality traits accounts for a large proportion of the choice of academic orientation which in turn is related to technical aptitude.

The relationships between navigation in hypermedia and the users cognitive abilities, is not clear. As indicated by the technical aptitude factor, it seems as if *spatial ability* is one crucial factor. This can be examplified with Vicente and Williges study where they found that spatial ability could be linked with whether users get lost in a hierarchical file system (1988). Also, in a study by Benyon and Murray (1993) it was found that spatial ability determined how well users performed with different interfaces to a database system. Users with low spatial ability performed better with an aided-navigation interface with a constrained dialogue, while users with high ability made better use of a non-aided navigation interface with a flexible command-based dialogue.

Spatial ability is a cognitive characteristic which offers a measure of a user's ability to conceptualise the spatial relationships between objects. It is closely allied to the notion of a cognitive map (Neisser 1976), and (following (Vicente et al. 1987) and (Vicente and Williges 1988)), may also be related to a user's ability to navigate through a complex space.

Attempts to isolate individual cognitive differences are generally based around the production of various aptitude tests and other devices designed to isolate specific factors (Dillon and Schmeck 1983, Dillon 1985). Cognitive abilities, or cognitive skills, as spatial ability, seek to describe the methods by which humans process information (van der Veer 1994).

It should be noted that spatial ability has been measured with different tests in the different studies. It is not clear that it is the same cognitive activity that is discussed in the different cases.

So, in summary, there are large differences between individual users and their ability to search for information, and a number of individual characteristics can potentially explain these huge differences. As information retrieval systems cannot be such that they only direct themselves to an exclusive population (as pointed out by Egan, 1988), there might be reasons to design the system so that it meets the needs and abilities of their users rather than the other way around. In particular, this becomes crucial if the related individual characteristics are more resistant to change than others (as suggested by van der Veer, (1989)).

Hypothesis

The present study investigates the relation between cognitive abilities, in particular spatial ability, and ability to navigate in hypermedia. Our hypothesis is that people with a low spatial ability have more difficulties in navigating to and interpreting information in a hyperspace.

The hypothesis stems from an observation in one of our previous studies (Bladh and Höök 1995) on a hypermedia documentation system, where we found that our user group could be divided into two. One found the graphs provided by the system helpful as tools for navigating to documents. The other group was annoyed by the graphs, and would have preferred other ways to navigate and search for information.

In this study, we tested the subjects cognitive abilities, and then made them solve six tasks with the hypermedia system. By choosing the navigational tasks so that they included navigation between several "pages" of information in order to find the requested information, we provoked navigation. The cognitive tests included tests of verbal ability and logical-inductive ability (apart from the tests of spatial ability and perceptual-analytic ability) in order to exclude the possibility that it was general intelligence that was related to the subjects navigational skills.

Characteristics of the hypermedia tool studied

Before describing the study, we need to provide a short introduction to the hypermedia tool studied. The tool was an on-line manual implemented in FrameMaker™. The domain described in the tool was an object-oriented software development method, named SDP, consisting of *processes* (in which project tasks were specified) and *object types* (in which the results were documented). The users were software developers, designers, project managers, etc., involved in developing large-scale telecommunications systems.

A schematic picture of the structure of the on-line manual can be found in Figure 1. The *process structure* and the *object hierarchy* function as maps to the two main structures in the domain: the process structure and the object type structure. From the process structure it is possible to reach graphical descriptions (or views) of each process, the *process graphs*. From the process graphs in turn, textual descriptions of the processes can be reached. From the object type hierarchy the *object type graphs* can be reached, and from those the textual descriptions of each object type can be reached.

So, in order to get to a textual description of a process or object type, it is necessary to navigate via a set of graphs until the bottom of the hierarchy is reached - only there a textual description will be found.

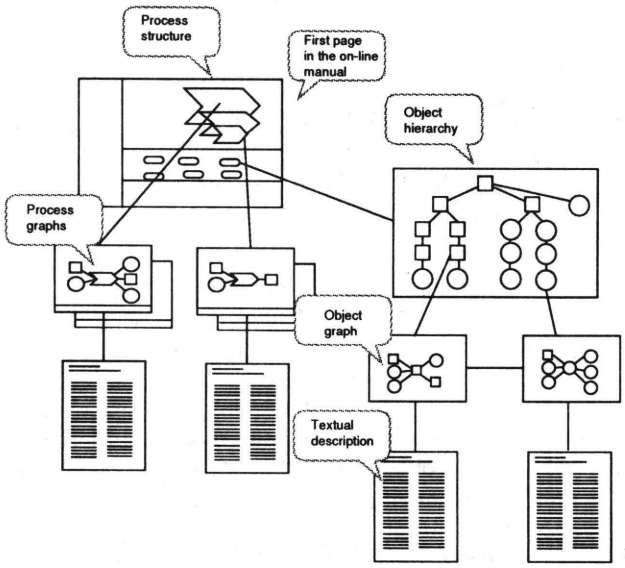


Figure 1. *The structure of the existing on-line manual.*

There are two aspects of the domain which are spatial in their nature. The domain itself is an abstract structure with relations between processes and object types. Second, the on-line manual is, as said above, structured in a set of documents with specified relations between the documents. The structure of the on-line manual is only partly based on the structure of the domain. So therefore the user has to learn two structures: the manual structure and the domain structure of processes and object types.

STUDY

Subjects

There were 23 subjects in the experiment, 19 male and 4 female, all employed at Ericsson Utvecklings AB or Ericsson. The subjects were in a range of 20 - 55 years old (m=34 years). All had some computer training, but not all had gone through higher academical training (18 had higher academic training and 5 had no academic training). All had recently gone through a four day course on the SDP method itself, but they had received little or no training on how to use the on-line manual.

Material and procedure

The experiment was divided into three parts. First, the subjects cognitive abilities were tested, followed by a questionnaire about their background (education, age, etc.) and finally, they completed six navigational tasks (while being video-taped) in the on-line manual. After solving each of the navigational tasks, the subjects answered a set of que-

ries on how difficult they had perceived it to be to find the information, etc.

Cognitive abilities
The cognitive abilities were tested using a subset of the Düremann-Sälde (Psykologiförlaget 1971) test battery:

- verbal ability (synonym test).
- logical-inductive ability (classification of images).
- perceptual analysis (drawing imitation of images).
- spatial , tested through three different tests:
- rotation of images: the subject choose the images that are identical with the image in the task.
- identification of left or right hand in pictures of hands that were turned in different ways,
- a blocks test (Koh's block test): the subject makes a pattern using blocks according to a given pattern.

Navigational tasks
Finally the subjects were asked to complete a set of six information seeking tasks and after each to evaluate their own performance, as whether they thought they had found the correct answer and whether they found it the most efficient way. The six tasks to be solved with the on-line manual and the subjects evaluation of their own performance took approximately 1/2 - 1 hour per subject. The information seeking tasks were designed to have the following properties:

- We used questions that we had collected in our previous studies from users actually working with the method and entering the on-line manual to find particular pieces of information.
- We designed two tasks, number 2 and 6, so that they asked for the same information concerning two different processes in the method - our hypothesis was that the second time around, users would more easily find the right information if they had built a good mental map of the information space.
- One task, number 5, could only be solved by looking in the textual parts of the on-line manual. That to see when the subjects made the decision to use that part of the on-line manual (immediately or after hesitation) and once in this part of the on-line manual, how they searched for information in the text.
- A criterion for some of the tasks were that they should force the subjects to navigate between the different graphs, that to see if and how the subjects used a mental model of the information space.

RESULTS

Patterns of cognitive abilities

A factor analysis of the results of the six cognitive tests revealed three underlying factors, with two tests with high loading in each factor, see table 1. Factor 1, with high loading on figure drawing test and the block test, factor 2 with high loading on the tests of synonyms and the classification of images, and factor 3 with high loading on the tests on rotation of images and of hand identification.

The factor analysis groups the tests in a different way than in the test materials. Note especially that the blocks test is more strongly associated with the figure drawing test than with the other two spatial tests.

This could indicate that there is a difference between the manipulation of spatial information in the mind and the acting in the world, even if both from a superficial point of view seem to concern the same kind of information processing.

Table 1. *A factor analysis of the cognitive tests.*

Cognitive test	Factor 1	Factor 2	Factor 3
Synonym test	.164	.843	.046
Classification of images	.284	.810	-.041
Image drawing	.958	.071	-.071
Rotation of images	.227	-.393	.654
Rotated hands	.00001	.067	.940
The blocks test	.792	-.001	.376

Cognitive ability and task completion time

The only correlation we found was that the completion time correlates with the spatial ability as measured with the two tests in factor 3, ($r = .56$, $p < .005$), see figure 2. (To be able to study the correlation between the subjects' abilities in the different factors and their task performance, we simply added the stanine points of the test scores on the two tests with the highest loading on the different factors). No other correlations achieved significance. The correlation between the completion time and the blocks test was for instance only $r = .04$.

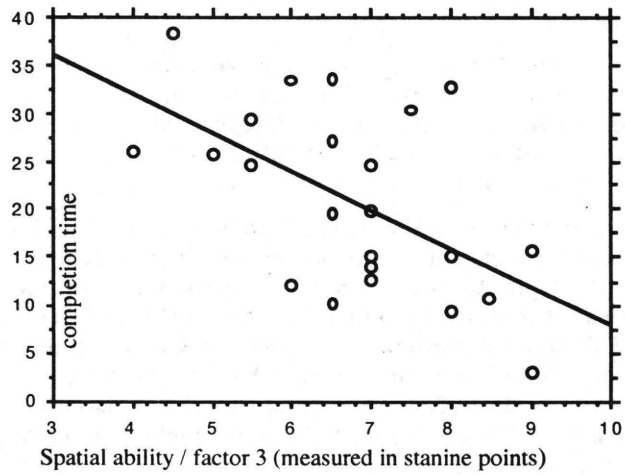


Figure 2. *The correlation between completion time and spatial ability.*

The difference between the best and worst performance in our test is 19:1, i.e. the best subject solved the tasks 19 times as fast as the subject who solved them slowest.

If we divide the subjects into two categories, those with high ability with respect to factor 3 and those with low ability, the result becomes even more evident. The group is divided into two halves of (almost) equal size: the high ability group, consisting of 12 subjects has an average stanine-score above 6.5 points, while the low ability group, consisting of 11 subjects lies below 6.5 points. In table 2, we see the results: the low ability group took about 25 minutes in average to complete the tasks, while the high ability group completed the tasks in 17 minutes. There was no significant difference in number of clicks (i.e. moves between pages of information or graphs) between the two groups. It seems as though the low-ability subjects spend more time in studying each page.

Spatial ability (factor 3)	Time	Clicks
High (12 subjects)	16.94	49.67
Low (11 subjects)	25.47	54.09

Table 2. *Total time and total number of clicks done in order to complete the six navigation tasks for the low and high ability groups on factor 3.*

Correlation with map-reading ability

In our questionnaire, we asked the subjects to estimate their own ability to read and use maps and their "sense of location". Streeter and Vitello, (1986), found that subjects self-estimated ability to read and use maps was strongly correlated with their actual map-reading ability. It is also the case that those who tend to like maps will use maps more often and thereby improve their performance. So, our subjects own estimate of the map-reading ability can be taken as a good measure of their actual (map-reading) ability.

We found that the subjects map-reading ability was correlated with factor 1 (the "external" spatial ability) in our cognitive tests, $r=.42$, $P<.05$, while there was no correlation between map-reading ability and factor 3 (the "mental" spatial ability).

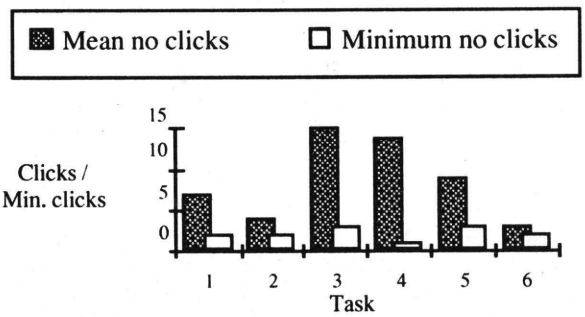
Again, this would indicate that there is a difference between spatial ability for solving problems in the world (where groundedness is possible) and spatial ability for extracting abstract structures from non-grounded domains. This may also imply that we can help one group of users to improve their performance through providing maps of the information structure. This group of users are those who are low on "mental" spatial ability, but high on "external" spatial ability.

Confidence in answers

In Figure 3 we compare the mean number of clicks the subjects took in order to complete the tasks with the optimal, lowest, number of clicks by which someone could have found the answer (as computed by us who knew where the answers were). As we can see, the amount of "unnecessary" clicking is not so bad for some of the tasks. Task 3 and 4 stand out from the rest of the tasks.

Task 3 required that the subject studied at least two different graphs and computed the answer from relating the two graphs. Most subjects studied three or more graphs (which provided different viewpoints of the same graphical relations) and looked at each graph several times in order to make sure that they had understood the relations correctly. (It should be noted that each individual graph was very simple).

Figure 3. *Mean number of clicks performed by the subjects compared with minimum number of clicks possible for each of the six navigational tasks.*



The second task which seemed to take a lot of unnecessary clicking is task 4 which was a search task in which the users were required to look through several nodes in order to make sure that they had in fact found the correct answer. The minimum number of clicks for this task was based on the fact that we knew what the answer was and we did not need to make sure that there was any information that would contradict this conclusion.

So, in summary, the users did not perform too badly in terms of visiting too many nodes in the information space.

It seems as though the hypermedia tool raises the expectation of finding the information "just a few click away". As soon as the subjects have to perform more than a few clicks, they assume that they have gone wrong somewhere in the hypertool. As we can see in Table 3, the subjects tend to be more unsure of whether they have actually found the correct answer when they have to study several pages of information (as in task 3 and 4) in order to find the answer: for task 3 eight subjects had a low confidence in whether they thought they had found the answer, and for task 4 ten subjects had a low confidence in having found the correct answer.

Tasks/ Confidence	1	2	3	4	5	6
0-2 (low)	1	5	8	10	3	3
3-5 (high)	22	18	15	12	18	19

Table 3. *The number of subjects who had a self-assessed low confidence (0-2) or high confidence (3-5) in having found the correct answer to the six different tasks.*

Another tendency, which we cannot verify in this study, but that should be dealt with in subsequent studies, is a relation between being unsure about whether the answer is correct, when the answer was to be found in a graph (or a combination of several graphs). Our subjects sometimes

stated that they would have liked to see the answer in text to make sure that the answer was correct. This tendency can explain why they did so many clicks for task 5 (where the answer could only be found in the text), and still had a high confidence in having found the correct answer. It can also explain why they were so unsure about whether the answer was correct for task 3 and 4 where the answer had to be computed from different graphs and could not be found in the test.

This confirms with the results of Hare et al. (1995) who showed that users did study figures, but when they were asked to give a definitive answer, they would rather turn to a textual description than to extract the answer from a picture.

Discussion

Our hypothesis that navigation in the on-line manual could be correlated with spatial ability turned out to be correct. There we no correlation's with other cognitive abilities.

The results with the two factors, factor 1 and factor 3, as different aspects of spatial ability should be verified in other studies, but should they hold also under closer scrutiny, this could not only have theoretical, but also practical consequences.

A practical consequence can be that we should be supporting users of hypermedia with low spatial ability with external aids that transforms an internal task into an external one - sometime we discuss below. In fact, as pointed out by Vicente and Williges (1988), such aids might also help users with high spatial ability by decreasing their cognitive load.

On the theoretical side we have argued that a distinction could, and perhaps should, be made between spatial tasks that are performed as bodily actions in the world and those that take place solely in the mind. Our analysis of the subjects' response pattern on the cognitive abilities test gave some support for that notion. Further support was gained from the fact that the only cognitive abilities tests that correlated with the performance in using the help system was those that seemed to measure internal spatial tasks. The fact that the "external" spatial ability was correlated with the subjects self-estimated ability to read maps but not correlated with the mental spatial ability, reinforces the messages that the two abilities are different.

Yet another result is that hypertools raise the expectations of finding information just a few clicks away.

Finally, we also found some interesting aspects of graphical versus textual presentation of information. It seemed as though textual presentation was considered to be a more reliable source of information, or that the subjects found it hard to interpret the graphs provided in the on-line manual. Also, compiling an answer from several graphs seemed to be a difficult task - even though each graph in itself was fairly simple.

Implications for Interface Design

Since spatial ability is resistant to training, it is necessary that we find ways to improve the interfaces instead. Then users with low spatial ability will be enabled to better solve their real tasks. The are numerous ways by which we can improve the interfaces to hypermedia.

Firstly, it has been pointed out by, among others, Vassileva that search in a hyperspace should always be complemented by the possibility to pose search questions (Vassileva, 1995, Höök et al., 1996). Our subjects sometimes said that they would have wanted to see the information written in text, but since it was so hard to navigate to a text and then find the relevant information within those text pages, they hardly ever were able to find the right information. Provided with search questions, they would (presumably) have performed much better in retrieving textual descriptions. This would furthermore have avoided the pictorial characteristics of the interface altogether, and might therefore be better suited for some users who disliked the graphs.

Secondly, hypermedia is sometimes provided together with a navigation map or dialogue history which allows the user to trace where they have been previously, or at least they are provided with an overview of the information space. This would have helped the users to know where in the information space they were currently at. On the other hand, it is not clear that all users with low spatial ability would be helped by such maps. We know that there are many users with low map-reading skills (Streeter and Vitello, 1986). Still, since map-reading skill was correlated with factor 1 and not with factor 3, we could achieve a better interface for those with high ability in factor 1 but not so high ability in factor 3.

Thirdly, it has been shown that users with low spatial ability are helped by system which achieves *visual momentum* (Vicente and Williges, 1988). Visual momentum is achieved when parts of the previous state of the interface is visible after the user has made an action at the interface. Vicente and Williges used a hierarchical file system with folders and files in folders in their studies. When the user attempts to open a folder, the content of the folder can be shown in a new window, perhaps completely covering the set of files and folders on the higher level. A way to achieve visual momentum, on the other hand, would be by inserting the file names in the folder indented inside the list of files and folders on the higher level. The previous state of the file system is thereby still visible to the user after the new action has been taken.

The reason that visual momentum helps users with low spatial ability, is that we move from requiring that the users recall where they are and where they should go next in the hyperspace, to recognising the structure and the links and basing their decisions on that recognition. So the interface provides memory support. Or phrased in terms of the results of this study: the mental spatial task is transformed into a external spatial task.

In a hypermedia system, it is harder to achieve visual momentum in a simple way since the whole idea is to move between whole pages of text (or graphics). Still, a dialogue history (either in map form or as a list of the names of visited nodes as in Netscape - a WWW browser) could be an improvement.

It should be observed that off loading the mental load of spatial cognition, might help both those users with low and high spatial ability, as discussed by (Vicente and Williges, 1988). It is not clear exactly what is improved by providing the kind of aids as outlined above.

ACKNOWLEDGEMENT

We would like to thank the anonymous subjects from Ericsson and Ellemtel who participated in our study. Some aspects of this study is also published in (Dahlbäck et al 1996) - added in this article are results from the study affecting how we should design hypermedia systems, and proposals for how to design hypermedia systems

The work was funded by NUTEK, Ellemtel and SICS jointly in the PUSH project.

REFERENCES

Allen, B. (1992) *Cognitive Differences in End User Searching of a CD-ROM Index*, 15th Annual International SIGIR'92, ACM, Denmark.

Benyon, David R., and Murray, Dianne M. (1993) *Developing Adaptive Systems to Fit Individual Aptitudes*, Proceedings of Intelligent User Interfaces'93, Florida, ACM Press.

Bladh, Malin, and Höök, Kristina (1995). Satisfying User Needs Through a Combination of Interface Design Techniques, In: K. Nordby, P.H. Helmersen, D.J. Gilmore and S.A. Arnesen (eds.), *Human-Computer Interaction INTERACT'95*, Chapman & Hall, Oxford, 1995.

Borgman, Christine L. (1989). All Users of Information Retrieval Systems are not Created Equal: An exploration into individual differences, *Information Processing & Management*, Vol. 25, No. 3, pp. 237-251.

Conklin, J. (1987). Hypertext: An Introduction and Survey, *IEEE Computer*, 20:17-41.

Dahlbäck, Nils, Höök, Kristina and Sjölinder, Marie (1996). *Spatial Cognition in the Mind and in the World: The case of hypermedia navigation*, The Eigtheenth Annual Meeting of the Cognitive Science Society, CogSci'96, University of California, San Diego, July.

Dillon, R. F. (1985). *Individual Differences in Cognition, Volume 2* Academic Press, London.

Dillon, R. F. and Schmeck, R. R. (1983). *Individual Differences in Cognition Volume 1* Academic Press, London

Egan, D. (1988). Individual Differences, In M. Helander (ed.) *Handbook of Human-Computer Interaction.* Elsvier Science,

Elkerton, J, and Williges, R. C. (1984). Information Retrieval Strategies in a File-Search Environment, *Human Factors* **26** (2), pp. 171 - 184.

Espinoza, Fredrik and Höök, Kristina, (1996). *A WWW Interface to an Adaptive Hypermedia System*, Conference on Practical Application of Agent Methodology (PAAM), London, April.

Greene, S. L., Gomez, L. M., and Devlin, S. J. (1986). A Cognitive Analysis of Database Query Production, *Proceedings of the Human Factors Society*, pp. 9 - 13.

Hare, M., Doubleday, A., Ryan, M., and Bennett, I. (1995). Intelligent Presentation of Information Retrieval from Heterogenous Multimedia Databases, In John Lee (ed.) *Proceedings of the First International Workshop on Intelligence and Multimodality in Multimedia Interfaces: Research and Applications*, Edinburgh.

Höök, Kristina, Karlgren, Jussi, Wærn, Annika, Dahlbäck, Nils, Jansson, Carl-Gustaf, Karlgren, Klas and Lemaire, Benoit (1996, in press). A Glass Box Approach to Adaptive Hypermedia, *Journal of User Modelling and User Adaptive Interaction*, UMUAI 6, in press.

Karlgren, Klas (1992) *Orientering i Hypermedia*, SISU report number 8, SISU, Stockholm, Sweden.

Neisser, U. (1976). Cognition and Reality, W. H. Freeman and Company, San Francisco.

Nielsen, Jacob (1990). *The art of navigating through hypertext*, Communications of the ACM 33(3):296-310.

Psykologiförlaget (1971) Manual till DS-batteriet, Psykologiförlaget AB, Stockholm.

Streeter, Lynn A. and Vitello, Diane (1986). *A Profile of Drivers' Map-Reading Abilities*, Humans Factors, 28(2).

Vassileva, Julita (1995). Ensuring a Task-based Individualized Context for Information Retreival from Hypermedia Information Systems for Hospitals, in IJCAI'95 Workshop on Intelligent Multimedia Information Retrieval.

van der Veer, Gerrit C. (1989). *Individual differences and the user interface.* Ergonomics, 32, pp. 1431-1449.

van der Veer, Gerrit C. (1994). *Mental models of computer systems: visual languages in the mind.* In: M.J. Tauber, D.E. Mahling, F. Arefi (eds.) Cognitive aspects of visual languages and visual interfaces. Elsevier Science B.V., Amsterdam 3-40.

Vicente, K.J., Hayes, B.C. and Williges, R.C. (1987). Assaying and isolating individual differences in searching a hierarchical file system, *Human Factors*, 29(3), 349-359.

Vicente K. J. and Williges, R. C. (1988). Accommodating individual differences in searching a hierarchical file system, *International Journal of man Machine Studies*, 29 647 - 668

Pictorial Aids in Computer Use

Raquel Navarro[1], José Juan Cañas[2] and Maria Teresa Bajo[3]
Departamento de Psicologia Experimental, Facultad de Psicología, Universidad de Granada
Campus de Cartuja, 18071 Granada, Spain
+34 58 243 767
[1]rnavarro@goliat.ugr.es, [2]delagado@goliat.ugr.es, [3]mbajo@goliat.ugr.es

ABSTRACT
The two studies reported here attempted to explore the effectiveness of different types of pictorial aids to enhance subjects' mental models and to facilitate learning. Computer naive subjects learned the operating system MSDOS under different learning conditions. We attempted to optimize subjects mental models by maximizing the congruency between the type of information implied by the command and the type of pictorial aid. Thus, in the first experiment, pictorial aids depicting change of state information were provided to the subjects, whereas relational information was provided in the second experiment. Results showed that the facilitatory effect of pictorial aids depended on the type of information represented by the pictures.

Keywords
pictorial aids, learning programming, computer use

INTRODUCTION
Part of the research efforts in computer learning has been directed toward enhancing the user's representation of the components and operating rules of the system. The main assumption is that the acquisition of procedural skills such as programming will be facilitated by a good understanding and mental representation of the computer. For example, Mayer (1981) provided mental model training to novice programmers and showed that this training increased student's performance in BASIC problems. Similarly, other studies have attempted to enhance the acquisition of good mental models by providing students with pictorial aids depicting the components of the computer involved in many operations. In a recent paper, Petre (1995) has suggested that, although pictorial information may be convenient and attractive, incorporating pictures and graphics into programming requieres a better understanding of the precise contribution of this type of representation. However, little systematic research has been directed to understand how and in which conditions the presence of pictorial aids facilitates learning to program.

Psychological research on the efficacy of pictorial aids has shown that these aids are effective provided that the information depicted by them is congruent with the type of information needed to perform a given task. Research on the role of pictorial aids in text comprehension and learning has indicated that pictures can facilitate learning only if the information provided by the picture is congruent with that contained in the text. Thus, Vicent (1992) found that learning of texts describing physical devices was facilitated by the presence of pictures, only if the pictures depicted information about the physical and functional relations among the components of the device. Also, Waddill and McDaniel (1992) showed facilitation of narrative texts with pictures that contained relational information

In the two studies reported here, computer naive subjects learned the operating system MSDOS under different learning conditions. We attempted to optimize subjects mental models by maximizing the congruency between the type of information implied by the command and the type of pictorial aid. A logical analysis of the MSDOS commands provided some insights on the type of pictorial aids needed. This analysis revealed that the commands can be classified in two categories. A first group of commands required a clear understanding of the relationship between two components of the computer to be used correctly. For example, the command "print" implies a relation between the computer screen and the printer. A second group of commands appeared not to require the understanding of this relational information, but of the changes in the state of a given component when the command was executed. For example, the use of the command "Format" implies some knowledge of the change in state (partition of the diskette into lines, sectors and so on) produced by the command in one of the components (the disk).

Our hypothesis was that the presence of pictures by itself would not be helpful in enhancing subjects' mental models and performance. Only pictures that stressed "change of state" or "relational" information would be effective in understanding the commands, and would show an advantage over verbal descriptions.

In our first experiments we explored the effectiveness of the "change of state" pictures by introducing pictures that stressed this type of information and by comparing them to pictures without this stress and with verbal information that also varied in the stress given to the change of state information. In our second experiment we explored the effectivenes of pictures depicting relational information and compared them with no-relational pictures and with verbal descriptions that varied in the type of information that tried to convey (relational or simple standard descriptions).

GENERAL METHOD

Pictures and definitions provided by Experts

The pictures and verbal descriptions of the commands used in both experiments were provided by three psychology professors with large experience on the operating system MSDOS. They were asked to write down what they thought neccessary to understand each of the commands and to use them appropriately. They were also asked to draw a picture that clearly represented the information convey by them.

An analysis of the verbal descriptions and pictures provided by the experts, revealed that the commands could in fact be classified in two categories: "Change of state commands" and "Relational commands".

Learning phases

In both experiments, students went through two learning phases. In the first one, students were provided with a short description of the main components of the computer (CPU, RAM, etc.) and with definitions of "disk", "directory" and "file". Then, they were introduced to each of the MSDOS commands. A verbal description of each command and its syntax was presented in the computer screen. Depending on the conditions, subjects might also be provided with a picture representing the command. The presentation of the description of each command was preceded by a computer menu depicting all the commands to be learned.

The commands were presented in a fixed secuence to all the students. After reading the description of each command, students performed a practical exercise showing its correct use.

To equate the number of times the students read the description, each command was presented four times for a maximum time of 60 sec (although students could go to the following screen once they had understood it).

The second phase was conducted one day later. At the begining of this phase, students were again presented descriptions of the comands (in some conditions, with pictures depicting them) and performed practical exercises. Then, students were assessed in their understanding of the commands and their ability to use them. The assessment session consisted of three parts:

Performance task

Students performed 7 practical exercises involving the use of the commands. They were provided with a piece of paper that showed the correct syntax for each command. Hence, syntactic errors were minimized

Two judges scored performance in this task by assigning one point to each successfully performed problem. Thus, students in this task had scores that ranged from 0 to 7. Independent grades for each command were not obtained because the way in which the exercises were set up implied serial dependency so that many times it was not possible to perform a given exercise if the previous one had not been successfully performed (e.g., it was not possible to copy a file within a directory, if the directory had not previously been created).

Definition task

Students were presented with a list of commands and computer components (e.g., CPU, RAM) and were asked to write down a definition for each.

Experts' descriptions and texbooks explanations were used as criteria for evaluating subjects' definitions. Two judges scored those defintions on a 5 point scale (from 0 to 4).

Relatedness taks

Students completed a relatedness rating task concerning 18 commands and computer components. They were to assign ratings (from 0 to 9) to pairs of concepts according to how related they thought the concepts were. A rating of zero indicated that the pair was unrelated, and a rating of nine indicated a high degree of relatedness. The task was also performed by five expert programmers who formed our group of experts.

This task was introduced to assess the mental representation acquired by the students in the course of learning. The task produces a matrix of proximity values in which each value represents the degree of relationship between a pair of concepts. Proximity matrices of all students in one group were averaged. Then, the five proximity matrices from the students groups and the averaged matrix of the experts were submited to the INDCLUS (Individual Differences Generalization of the Additive Clustering Model) algorithm (Carroll and Arabie, 1983). INDCLUS takes proximity matrices and performs an additive clustering analysis on them. This method produces overlapping clusters so that one concept could belong to more than one cluster. This method was chosen instead of the more traditional non-overlapping ones (e.g., Hierarchical Clustering) because clusters overlap has been shown to be appropriate for capturing mental representations (Sheppard and Arabie, 1979). Furthermore, the algorithm finds a global solution for all the proximity matrices and gives the weigths that each group matrix has in each cluster. Therefore, it is possible to analyze individual differences by comparing those weigths.

EXPERIMENT 1

Subjects

Seventy psychology students at the University of Granada with a complete lack of knowledge about computer concepts and usage participated in each of the five experimental conditions.

Design

Subjects were assigned to one of the five groups that varied in their learning conditions. Those groups were:

TextChange (TC): Students read a verbal description that stressed the state of the component affected by the command before and after it was executed.

TextNoChange (TNC): Students read verbal descriptions of the commands similar to those provided by computer manuals.

PictureChange (PC): Students saw "change of state" pictures of the commands and read neutral verbal descriptions of them.

PictureNoChange (PNC): Students saw neutral pictures and read neutral verbal descriptions of the commands.

PictureTextChange (PTC): Students saw "change of state" pictures of the commands and read verbal descriptions that stressed the state of the components affected by the command before and after it was executed.

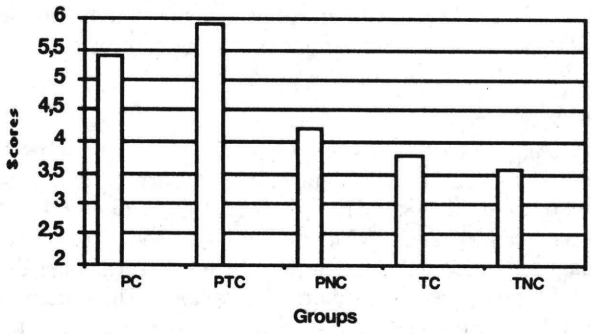
Results

Performance Task

An analysis of variance (ANOVA) in the subjects' grades showed a significant effect of learning condition, $F(4,65) = 5.90$, $p < 0.01$. As can be seen in Figure 1, subjects provided with "change of state" pictures showed better performance than subjects provided with verbal descriptions or neutral pictures. Between groups comparisons allow to reach the following conclusions:

(1) The presence of pictorial aids with change of state information in the learning phase facilitated the performance. The difference between the groups with "change of state" pictures (PC, PTC) and those without them (PNC, TC, TNC) was significant, $F(1,65) = 14.99$, $p < 0.01$. (2) There was an overall effect of stressing change of state information. Performance of the groups that had that information (PC, PTC and TC) was better than that of groups that did not have it $F(1,65) = 8.74$, $p< 0.01$. (3) Although pictures facilitated learning, they did so only when they contained change of state information.. There was a significant difference between the PC and PTC groups and the PNC group, $F(1,65) = 7.89$, $p < 0.01$. (4) When students were not provided with change of state information, having a picture made no difference as shown by the similar performace of PNC and TNC groups, $F(1,56) = 1.13$, $p = 0.29$. (5) Change of state information without a picture

Figure 1. Performance task
Change of state information

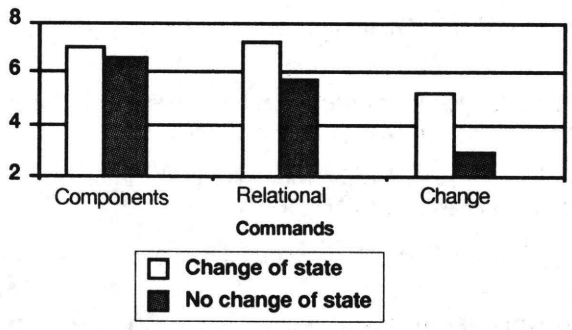


did not facilitate learning. The comparison between TC and TNC was not significant, $F(1,56) < 1$, $p = 0.72$. (6) The joint presentation of change of state information by pictures and text was not better than presenting that information by pictures only. Performance of the groups PC and PTC was similar, $F(1,65) < 1$.

Definition task

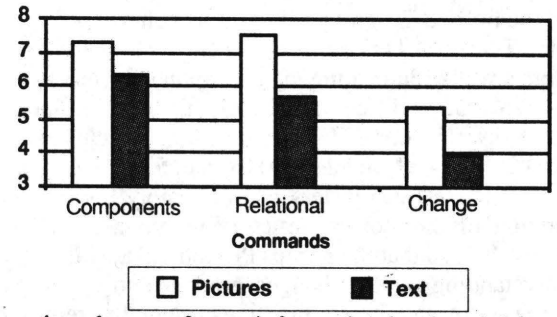
In this task the commands and computer components were independently defined, therefore we could analyzed "change of state" commands, "relational commands and componentes separately. This analysis is important because we could explore whether the observed facilitation was general or specific to the change of state commands. The results of this analysis revealed differences due to the presentation of pictures and change of state information. This facilitation was specific to the change of state commands.

Figure 2. Definition task
Change of state information



The groups that were provided with change of state information (PC, PTC and TC) learned change of state commands better than the groups that did not have that information (PNC, TNC), $F(1,68) = 8.39$, $p = .01$. There was no difference among the groups in the definition of components, $F(1,68) < 1$, and relational commands, $F(1,68) = 3.19$, $p > .05$. Figure 2 shows the mean scores for these two sets of groups in each type of definition.

Figure 3. Definition task
Change of state information



Presenting change of state information in pictorial format produced a small effect that did not reach statistical significance. The groups with change of state pictures (PC and PTC) defined the concepts better than the groups without those pictures (TC and TNC). However, the difference was not significant for the components, $F(1,68) = 2.22$, $p> .05$, relational commands, $F(1,68) = 3.46$, $p > .05$, or change of state commands, $F(1,68) < 1$. Figure 3 shows the results of these analyses.

Relatedness Task

Six proximity matrices, one for each student group and one for the expert group, were submited to the INDCLUS algorithm. The solution contained 10 clusters:

(1) CPU, RAM, Buffer Memory, Disk, File, Directory, Format, Diskcopy, Diskcomp, Dir, CD.

(2) Directory, Dir, MD, RD, CD.

(3) CPU, RAM, Buffer Memory, Peripheral Devices, Disk, File, Directory, Dir, Type, Print.

(4) Buffer Memory, Peripheral Devices, Disk, Format, Diskcopy, Diskcomp, Erase, RD.

(5) Disk, File, Directory, Format, Rename, Erase, MD, RD.

(6) CPU, RAM, Buffer Memory, Peripheral Devices, Disk, Diskcopy, Diskcomp, Print, MD.

(7) CPU, RAM, Buffer Memory, Disk, File, Directory, Rename, Erase, Dir, Type, RD, CD.

(8) Disk, File, Format, Diskcopy, Diskcomp, Rename, Dir, Type.

(9) CPU, RAM, Buffer Memory, File, Directory, Diskcopy, Type, MD, RD.

(10) Buffer Memory, File, Directory, Rename, Erase, Dir, Type, Print, MD, CD.

As we mentioned, the algorithm provides the weight that each group assigned to each cluster. Thus, 10 weights were obtained for each group. The similarity between the groups in the way they weighted the differemt clusters was analyzed through a k-means cluster analysis. The purpose of this analysis is to create group clusters so that within-cluster variability is minimized and between-cluster variability is maximized. Results of this analysis produced two group clusters. The first cluster contained the expert group and the PC and TC groups. The secong cluster grouped the PNC, TNC and TPC groups. The k-means algorithm also offers a way of indicating the conceptual clusters in which the two groups differ. Results of this analysis indicated that the two groups differ in the conceptual clusters numbered 2, $F(1, 4) = 14.82$, $p = 0.01$, and 5, $F(1,4) = 20.04$, $p = 0.01$. These two clusters seem to imply a good understanding of the storage structure within a disk. Hence, results showed that the groups PC and TC acquired a good understanding of the changes produced in a disk when saving information, so that they mentally represented those commands in a way similar to the experts.

Discussion

Results of the performance and definition tasks showed a clear effect of the type of information provided to the subjects in the learning phase. Thus, descriptions of the commands offering information about the change of state of the components produced better performance in the practical exercises and better definitions of the commands than neutral descriptions.

On the other hand, pictorial aids were clearly effective in the performance task. Although the effects of these aids were not as evident in the definition task, there was also a tendency to produce better performances. The effectiveness of these aids seem to be due to the information represented by them and not to their pictorial qualities.

Finally, subjects' mental representations, as captured by their relatedness ratings, also showed the effect of presenting change of state information in pictorial format. Results indicated that the pictorial presentation of the change of state of a component when a command is executed, improve the mental organization of those commands. This mental representation become more similar to that of the experts. Specifically, subjects who receive change of state information seem to share with the experts conceptual groupings regarding directories.

EXPERIMENT 2

Subjects

Seventy psychology students at the University of Granada with a complete lack of knowledge about computer concepts and usage participated in each of the five experimental conditions. None of them participated in the previous experiment.

Design

Subjects were assigned to five groups that varied in their learning conditions. Those groups were:

TextRelation (TR): Students read verbal descriptions that stressed the relations between the components affected by the command.

TextNoRelation (TNR): Students read neutral verbal descriptions of the commands. Those descriptions were similar to the ones provided by computer manuals.

PictureRelation (PR): Students saw pictures depicting the relation between the components affected by the commands and read neutral verbal descriptions of them.

PictureNoRelation (PNR): Students saw neutral pictures and read neutral verbal descriptions of the commands.

PictureTextRelation (PTR): Students saw relational pictures and read verbal descriptions that stressed the relation between the components affected by the command.

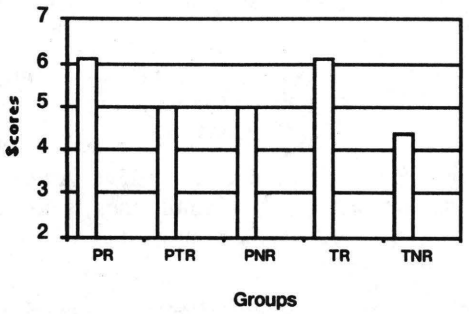
Results

Performance Task

An Analysis of Variance (ANOVA) on the subjects' scores showed a significant effect of learning condition, $F(4,65) = 3.54$, $p < 0.01$. As can be seen in Figure 4, subjects provided with "relational" information either in textual or in graphical form showed better performances than subjects provided with neutral information. The groups PR and TR were significantly better at performing the task than the PNR and TNR groups, $F(1,65) = 12.17$, $p < .01$. However, the difference between the PR and TR groups was not significant $F(1,65) < 1$. Presenting the relational information in textual or pictorial format made no difference. Finally, the groups with both textual and graphical relational information (PTR) were no better than the groups without relational information (PNR and TNR), $F(1,65) < 1$. Therefore, having relational information in

both format had a detrimental effect relative to having it in a single format.

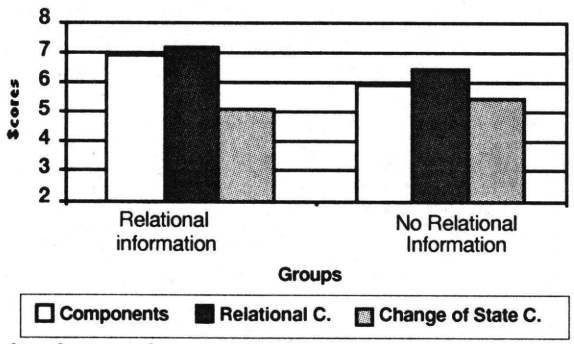
Figure 4. Performance task
Relational Information



Definition task

Figure 5 presents a summary of the main results of the definition task. The groups that were presented with relational information gave better definitions of the components and relational commands than of the change of state commands, $F(2,136) = 31.59$. In contrast, those groups that were not provided with relational information showed similar scores in their definitions of the components and the two types of commands, $F(2,136) = 1.21$, $p = .30$.

Figure 5. Definition task
Relational Information



Relatedness task

Six proximity matrices (one for each of the 5 student groups and one for the expert group) were submitted to the INDCLUS algorithm. The solution offered by the algorithm contained 10 clusters:

(1) Buffer Memory, Disk, File, Directory, Rename, Print, MD, RD, CD.

(2) Buffer Memory, Peripheral Devices, Disk, Format, Diskcopy, Diskcomp.

(3) CPU, Peripheral Devices, File, Type, Print.

(4) CPU, RAM, Buffer Memory.

(5) CPU, RAM, Buffer Memory, Disk, Peripheral Devices, File, Directory, Diskcomp, Diskcopy, Dir, Type, MD, RD, CD.

(6) Buffer Memory, File, Directory, Rename, Erase, Dir, Type.

(7) CPU, RAM, Disk, File, Directory, Format, Diskcopy, Dir.

(8) Buffer Memory, Disk, File, Diskcopy, Diskcomp Erase.

(9) Peripheral Devices, File, Format, Diskcopy, Diskcomp, Erase, Dir.

(10) Directory, Dir, MD, RD, CD.

As in Experiment 1, the clusters' weigths from the six groups were submitted to a K-means cluster analysis to assess the similarity among the groups. The pattern of results of this analysis was similar to that found in the performance and definitions tasks. The PR and TR groups were clustered with the experts. Thus, the mental representations of the two relational groups were similar to the mental representations of the experts. There was a significant difference in the weigths assigned to cluster 10, $F(1,4) = 8.05$, $p < .05$. This cluster contained the terms that were in cluster 2 in the previous experiment and that refer to the organization of information in directories.

Discussion

The results of this experiment indicated that students that were provided with information about the relations between the components of the computer were able to use the commands and define them better than those students without this information. This facilitation seem to be specific to congruent information. Thus, providing relational information produced better definitions of relational commands and components. The facilitatory effect of providing relational information did also show in the relatedness ratings, so that subjects in the relational groups showed better mental representations than subjects in the non-relational groups. In this experiment, presenting relational information in pictorial form did not improve students' perfomance. This again indicated that the facilitatory effects of pictorial aids showed by previous studies might not be due to their intrinsic pictorial qualities, but to the information provided by them.

General Discussion

Pictorial aids are used very often in computer programming to enhance learning and use of the programs. The pictures are created by the programmers, who are considered to be experts and to know what is important to include in them. Therefore, pictorial aids are usually created from experts' intuitions and little systematic research has been directed to show when or how they are really effective.

The results of these two experiments made evident the need for this systematic research. In Experiment 1, the presence of pictures facilitated learning provided that those pictures stressed the change of state of the components. However, in Experiment 2 students that were provided with pictorial information about the relations between components did not show better performance than students that received this information in textual form. Hence, presenting relational information in pictorial form did not improve learning relative to presenting it in a text.

This pattern of results seems to indicate that pictures generated by experts do not always facilitate help learners to get a better understanding of the information provided by them. Experts draw pictures following logical analyses of the information that might not be helpful to novices.

A question that might be important at this point is to understand the reasons why "change of state" pictures facilitate learning whereas pictures depicting relational information do not. This differential effectiveness might be due to the type of information or to the type of picture. Although both types of pictures were drawn by the experts their pictorial qualities were very different. Inspections of the type of pictures drawn by the experts for both types of commands showed greater variability in the pictures provided for the relational commands. Thus, experts' mental representations of relational images might not be homogeneous and might fail to facilitate learning On the other hand, comparisons between relational and change of state commands showed that the first type of information was easier to learn than the second. This differential difficulty might have reduced the effectiveness of the relational pictures.

Further research is needed into this question. However, conclusions of the two studies suggest that pictorial aids provided by experts might not be helpful to learners, so that empirical analyses of them are needed before selecting them to be included in manuals and programs.

REFERENCES

Carroll, J.D. and Arabie, P. (1983). INDCLUS: An individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika, 48*, 157-169.

Mayer, R.E. (1981). The psychology of how novices learn computer programming. *Computer Surveys, 13*, 121-141.

Petre, M. (1995). Why looking isn't always seeing: readership skills and graphical programming. *Communications of the ACM, 38*, 33-44.

Sheppard, R.N. and Arabie, P. (1979). Additive Clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review, 86*, 87-123.

Vicente, K. (1992). Memory Recall in a process control system: A measure of expertise and display effectiveness. *Memory and Cognition, 20*, 356-373.

Waddill, P. and McDaniel, M.A. (1992). Pictorial enhancement of text memory: Limitations imposed by picture type and comprehension skill. *Memory and Cognition, 20*, .472-482

Ergonomics in the Wild: The Practitioner's Perspective

Ángel Cabrera
Andersen Consulting
Torre Picasso
28020 Madrid
SPAIN
Phone: +341-5966000
Fax: +341-5966695
angel.cabrera@ac.com

Manuel Rincón
Andersen Consulting
Torre Picasso
28020 Madrid
SPAIN
Phone: +341-5966000
Fax: +341-5966695
manuel.rincon@ac.com

Julián González
Andersen Consulting
Torre Picasso
28020 Madrid
SPAIN
Phone: +341-5966000
Fax: +341-5966695
julian.gonzalez@ac.com

ABSTRACT
The relationship between the world of research (be it applied or basic) and the world of practice must be bi-directional and mutually enriching. As consultants involved in the design of information systems for organizations undergoing integral transformation, we see a continuous need for scientifically based design guidelines and impact assessment techniques. The purpose of this paper is to close the feedback gap between the world of practitioners and the world of research with a few reflections that might help identify lines of research that will satisfy some of the needs practitioners have. We discuss, among other things, the difficulties we have in reframing organizations' perceptions of information systems so as to include in them the notion of usability. We discuss also the need for reliable methods of utility analysis that would translate the benefits of ergonomic design into monetary terms, the need to connect usability goals with business objectives, and the role that usability design might play in the context of changing organizations.

Keywords
utility analysis, paradigm shift, usability goals, business goals

INTRODUCTION
Given that we, practitioners of ergonomics in the business world, take advantage of the knowledge corpus developed by Cognitive Psychology and Ergonomics when we design information systems and work environments, we cannot but try to share back with those disciplines the kinds of problems we encounter and the lessons we learn with the hope that this feedback will help identify new beneficial lines of applied research. With this objective in mind, we present here a few reflections drawn from our experience as HCI/Usability professionals.

Although the ideas that we discuss here are not confined to a particular project, we will make especial reference to a project whose size and complexity makes it a particularly interesting case study: the HCI usability activities carried out in the context of ARCO (Area Comercial), a macro project with the objective of building a complex information system to support the entire commercial processes of the Spanish telecommunication operator Telefónica.

What makes this project complex is the large number of users that it will impact (approximately 25,000), the large number of designers and developers involved (65 million US dollars in software development), the methodological heterogeneity of the teams involved (five vendors, IBM, EDS, Price Waterhouse, Eritel, and Andersen Consulting, share design and development responsibilities), and the long-term implementation of products (several implementation phases over a four year period starting in 1993).

Telefónica, as the rest of European telecommunication operators, is facing a deregulation process which is expected to be completed by 1998. In that year Telefónica will have to compete with other companies in every segment of the telecommunication market in Spain. This implies a great change for a company with nearly a century of comfortable monopoly operation. The ARCO initiative was launched as one of the keys to enable Telefónica to get ready for the new, highly competitive environment. ARCO will involve the transformation of all the (now independent) systems that currently support Telefónica's commercial operations (Marketing, Client service, Invoicing, etc.) into an integrated environment (based on a client-server platform) that will be used by several thousand users on a daily basis.

In this context, Andersen Consulting is performing a set of HCI/usability-related tasks. In particular, the focus has been placed on usability testing, plus design-related tasks in some critical application. This is due to the especial structure of this project in which, contrary to most of typical engagements, Andersen Consulting has only a little part of the systems design responsibility.

The main tasks the Usability and Prototype team is performing are the following:

- **Usability Standards Development and Maintenance**. This is a critical task in a project like ARCO. Every team in every vendor has its own criteria and style. It is crucial for the user not to perceive those differences, and consistency is maybe the first objective. The standards are the main tool to ensure that consistency.
- **Training**. Most of the systems design teams had little experience in GUI (Graphical User Interface) based design. One of the first tasks was to train them in basic GUI design.
- **Development Support**. The Usability and Prototype team acts as support for the development teams, con-

tinuing the training effort made at the beginning of the project.

- **User Interface Reviews**. As the development teams deliver new designs, our team reviews their usability and conformance to the standards. This team has reviewed roughly, 2,000 different windows with their corresponding dialogs.
- **End-user Evaluations**. This is the main task of this team. In order to evaluate the system's usability we set up a usability lab at the project office. We request some users to perform several tasks using the ARCO prototype. Then we observe and videotape their performance. This allows us to identify, in a very precise way, the most difficult parts of the design, and correct them before any further development takes place. To date, eight end-user evaluations have been performed, involving the participation of more than 75 users.

THE FUNCTIONALITY-USABILITY PARADIGM SHIFT

Before the prototype and usability team was created and this line of work started, AC had to engage in the not easy task of sharing with the client the importance of the concept of usability in a project of this magnitude. This initial step of introducing *usability* as a key objective in an initiative like ARCO was a challenge in itself for a number of reasons. The big objective of ARCO, and the design philosophy underlying the design of all of its applications, was (is) to re-engineer Telefónica's business processes along with the idea of business integration. This departing stance is easily translated by systems builders into *functional* requirements. What is not so easily grasped by system builders is the idea that functionality without usability is not functional, and that, consequently, usability and functionality need to be considered as objectives with similar priority.

Shifting the attention of system builders and system buyers from *functionality* to *usability* is a difficult, resistance prone and energy consuming process (interestingly, our experience shows that this difficulty is more marked on the builder's side than on the buyer's side). In any case, the well established community of information systems development does not very easily buy into the view of the user as a driving force of the design process, a view which, if accepted, would have large consequences in the very core of its activities. What needs to happen (is already happening) is what Kuhn referred to as *paradigm shift*.

Paradigms consist of assumptions and operating rules shared by the members of a scientific or professional community which create a common language and a common body of implicit expectations. Kuhn's studies of history of science have uncovered recurring patterns of resistance when moving forward to a new paradigm, patterns that are common to most significant scientific revolutions. Paradoxically, the acceptance of a new paradigm can turn out to be more challenging than the creation of the paradigm itself.

Understanding the way in which the usability paradigm shift is taking place in the world of information systems development is an area where practitioners may have much to say but in which researchers may be able to contribute in a more systematic way. In any case:

Exploring the mechanisms of the functionality-to-usability paradigm shift and developing educational tools to facilitate it appears as a key to the social acceptance of ergonomics and related disciplines and therefore constitutes a line of research with clear beneficial impact for both practitioners and researchers.

USABILITY VS. UTILITY

Unlike some scientific communities (especially the social sciences), organizations have well defined criteria and accounting techniques to define and assess success. This means that for any new paradigm to succeed in an organization, it will first have to prove a positive contribution to the organization's profit and loss bottom line.

In other words, a necessary (not sufficient!) condition for organizations to open up to the concepts of ergonomics and usability, is the reliable assessment of the added benefits of those concepts. Unfortunately, in the case of interventions having to do with the people side of organizations, establishing this link with the organization's bottom line is far from trivial.

Industrial/organizational psychologists have acknowledged this fact for a long time and have therefore devoted a significant research effort to solving this problem in the context of human resource interventions such as selection or training (e.g. Cabrera, 1995, Boudreau, 1995). The result of these efforts are a set of techniques, commonly referred to as *utility analysis techniques*, which more or less reliably translate the value added by a particular intervention in a given context into monetary units.

Utility analysis is not just accounting, but informed estimation of benefits. It takes into consideration the nature of the intervention at hand and builds upon it. Each of the utility analysis techniques that have been developed so far rely on assumptions that hold in certain situations and not in others.

In ARCO, as well as other projects, we have based our assessment of utility on past experiences and coarse estimates based on assumptions not yet fully tested scientifically. The difficulty of assessing utility in HCI/usability interventions stems from the fact that the benefits of ergonomic design come in a wide variety of ways. Among other things, usable computer interfaces are supposed to:

- carry savings in personnel training
- increase work performance
- incream637977423 se personnel multifunctionality
- increase work motivation
- decrease turnover
- decrease need for support
- decrease initial performance disruption, etc.

The problem that remains to be solved is twofold. First, each of those supposed effects needs to be contrasted scientifically. Second, the value of each of those effects needs to be assessed in monetary terms. A scientifically based approach to assess the utility of HCI/Usability interventions would certainly contribute to our ability to induce in organizations the functionality-usability paradigm shift required for the acceptance of ergonomic design as a valuable activity:

It is fundamental that we assess the validity of existing utility analysis techniques in the context of ergonomic interventions in organizations and that we develop new techniques if the existing ones appear not to be appropriate.

FROM BUSINESS GOALS TO USABILITY

Usability and ergonomics are context-dependent concepts. A command-based operating system interface such as a UNIX shell, which might meet the usability requirements of a highly computer literate user population would be disastrous in the context of home computing. The contrary, as it is well known, is also true: system developers tend to prefer command-based interfaces over WYSIWYG graphic interfaces for efficiency reasons.

Most usability and HCI texts emphasize the need to tailor designs to the characteristics of the user population. In other words, a good design does not only imply good functionality, it implies functionality adapted to the user's characteristics and needs.

What our experience as practitioners shows is that a third independent variable might be missing from this equation. In most business situations, it is not enough that our designs provide certain functionality and that the functionality fits the user's needs, designs must also enable whatever **business goals** are set by the organization for each particular user. It is not rare that one information system is used by users with different business objectives. In those cases, we are forced to design one or several interfaces in order to satisfy each one of those user types with respect to their respective goals.

In our design methodology, especially in projects that, like ARCO, impact a wide variety of users, we pay special attention to the task of translating business objectives into design specifications that somehow *define* what "usability" is supposed to mean in each particular context. In other words, usability is not just a function of the user, it is a function of the user and the work objectives in which the user operates.

Let us take ARCO as an example. One of the applications in ARCO, the Product Catalog, will be giving support to all the daily activities of all the commercial personnel of the company. From a learnability point of view, and given the age and prior experience of the users, we could opt for an interface with a classical text-based look-and-feel (e.g. IBM's 3270). Telephone sales personnel often need to be able to access information in the catalog while having a customer in the line, and need to be able to find specific pieces of information to attend customer requests or to offer products and services that satisfy certain customer needs. From a business perspective, whether the interface of the catalog is learnable or not is not as prioritary as the speed and accuracy at which customer requests can be satisfied. Therefore, even if efficiency implies designing a GUI interface that is less learnable for the users, the company is willing to assume the extra cost of user training given the expected benefits of speed and flexibility.

Another module of ARCO, the Executive Information System, is supposed to provide the company's managers with information summaries to help them assess the situation of the business, manage irregularities and plan future courses of action. Given the cost associated with training managers, the fact that this system is used only occasionally, and that some managers may be resistant to initially adopt this tool, learnability and subjective satisfaction (which depends highly on aesthetic factors) acquire a preferential place amongst other usability characteristics.

The following table summarizes some of the usability goals that may apply in different degrees to different products, and the types of users for which each goal is particularly important from a business point of view.

When studying and testing new design principles, it is important that those principles are not taken as universal but as context dependent guidelines whose validity is determined by a particular usability goal which may or may not be appropriate for the business objectives that guide each particular design.

Objective	Definition	Purpose
Error Prevention	the degree to which a system prevents incorrect information to be input into the system	key objective when small errors have the potential to create problems or errors in the revenue stream, or cause unnecessary downstream work to be performed in error (e.g. teller operations system in a bank system)
Flexibility	the number and variety of paths or options available in handling a logical unit of work	key objective for users who access the system in a variety of scenarios, or for systems that need to be used by different types of users (e.g. a word processor)
Learnability	the degree to which a new user with no previous training can exploit the system's functionality; (is the system self-explanatory, instructive, or straightforwardly navigable?)	key objective for occasional users and for zero-training situations
Memorability	the degree to which a user who does not access the system over an extended period of time can reach his or her initial level of mastery with it	key objective for infrequent users or for users of systems that are somewhat large and complex; supports a case for standards and consistency
Speed & Throughput	the degree to which the system allows for speedy input or inquiry of information to complete a task; tied closely to a user's productivity in performing the task	key objective when the goal is to process tasks quickly and repetitively (e.g. customer service)
User Appeal	the degree to which the system creates a pleasant subjective experience in the user	key objective when users have the option to use or not the system, in which case it is important to make a system attractive to users (e.g. a self-check-in stand at an airport)

USABILITY IN CHANGING ORGANIZATIONS

Another area which is acquiring more and more importance for practitioners is the role of ergonomics in changing organizations. Following new business paradigms such as re-engineering or business integration, information systems are no longer viewed as isolated entities, but as enablers of new business processes that go hand in hand with organizational restructuring and new human resource architectures. In other words, information systems are introduced in parallel with deep changes that affect every aspect of an organization.

Organizations are no longer stable entities neither in their structure, nor in the technology they use to accomplish their objectives, nor in their very own objectives and strategies. Some factors that have contributed to setting this new order are the globalization of markets and competition, deregulation of basic service companies (as it is the case with ARCO), speed of technological development, increased dependency of organizations on information technology, increased sophistication of customer demands, increased importance of customer service, etc.

In a scenario like this, the ability of organizations to change and adapt is not just a commendable attribute, but a real requirement for survival.

Not surprisingly, organizations are devoting more and more resources to learn how to manage change processes and to improve their ability to assimilate change (Conner, 1993). Andersen Consulting, for example, has created an entire division to help organizations assess and improve their capacity to engage in change processes (just in 1995 the number of Andersen Consulting professionals devoted to change management related projects grew at rates between 40% and 100% in different regions of the world).

Managing change successfully is an art not yet fully mastered but for which some insight has already been gained. Up to know, most of the interventions that are used to manage technological and organizational change have to do not so much with the nature of the changes themselves, but with leadership styles and communication interventions surrounding those changes.

Organizational change takes place at many different levels but it is always materialized in behavioral changes occurring at the individual level (Porras & Robertson, 1995). And it is exactly at the individual level that resistance to change becomes more visible. For this reason, we are starting to turn our attention to factors associated with the technological infrastructure that underlies most organizational changes in searching for new ways to manage change. This does not imply that we undervalue the value of leadership and communication interventions, the idea is that those interventions should be integrated and complemented by a careful (ergonomic) design of the technological infrastructure driving the change.

Andersen Consulting's view of change management, for instance, pays special attention to two concepts related to human performance and ergonomics: Human-Computer Interaction and Integrated Performance Support. The latter adds to the concept of HCI usability the idea of providing users with an integrated work environment that supports their decisions with the right kind of information, in the right form, and at the right time.

The assumption underlying this strategic wager is twofold. On one hand, it seems obvious that ergonomic design might contribute positively to users' ability to adapt to a new information system and, consequently, to their new roles within the organization. On the other (and less obvious) hand, it is assumed that good usability can improve users' attitudes towards change, i.e. their perception of the new technology as helpful rather than disrupting.

These assumptions, although intuitively sound, deserve empirical examination:

> *A line of research that would contribute to an increased understanding of change management and that could open up new routes for applied ergonomics in the business world is the study of the contribution of ergonomic design in facilitating the assimilation of change.*

In this line, we have started a study in which we are assessing people's attitudinal change towards a new information system as the result of participating in a usability test. The study is a pretest-posttest survey design in which people's attitudes are assessed prior to and immediately after participating in a usability test carried out in the context of a changing organization. Preliminary data were collected in ARCO. The results of this experience showed that people's attitudes improved as a result of their perception of an effort on the organization's side to incorporate their personal concerns and preferences into the designs. We are now in the process of gathering data from other projects and contexts that will allow us to determine the extent to which those initial effects are reliable and generalizable across contexts.

CONCLUSIONS

The benefits that applied cognitive ergonomics can provide to society at large and to the business world in particular are unlimited and, to a large extent, still unexplored. Ergonomics has entered business organizations accompanying information system development, but its effects are slowly expanding to other areas. For instance, some European financial organizations who became familiar with the concept of HCI usability while modernizing their information systems are now creating usability laboratories where all the products that are used to interact with customers (application forms, bills, automatic teller machines, branch layouts, etc.) are submitted to close usability scrutiny.

An effect of this increasing interest for ergonomics is that organizations want to understand the concept in great level of detail. They want to understand exactly what parts of the organization are impacted and how, and they want to know what benefits can be drawn from investing in ergonomics. We pointed out here some work done in the area of Industrial / Organizational Psychology as a possible start point to address these issues, but the problem remains an open (and promising) line of research.

We have also discussed the role that usability and ergonomics in general may play in managing change. There is a common (so far just intuitive) understanding that good usability might help ease the level of disruption introduced by new technology in work settings and that it might even carry some positive attitudinal changes towards organizational change. These are also issues that remain open and that we have just started addressing directly by gathering empirical data through our practice.

ACKNOWLEDGMENTS

The ideas discussed in this paper are those of the authors and by no means intend to be an institutional declaration. These ideas have been enriched throughout experiences and discussions shared with many Andersen Consulting professionals devoted to HCI design and usability. In particular, we wish to thank Xavier Trabado, Héctor Rodríguez, José Luis Arias Muerza and José Ignacio Domecq for their insightful comments on this paper.

REFERENCES

Boudreau, J. W. (1995). Utility analysis for decisions in human resource management. In M. D. Dunnette and L. M. Hough (eds.) *Handbook of Industrial and Organizational Psychology.* Palo Alto, California: Consulting Psychologists Press.

Cabrera, E. F. (1995). *Development and Application of a New Utility Analysis Model.* Unpublished doctoral dissertation, Georgia Institute of Technology.

Conner, D. R. (1993). *Managing at the Speed of Change.* NY: Villard Books.

Porras, J. I., & Robertson, P. J. (1995). Organizational development: theory, practice, and research. In M. D. Dunnette and L. M. Hough (eds.) *Handbook of Industrial and Organizational Psychology.* Palo Alto, California: Consulting Psychologists Press.

Kuhn (1962). *The Structure of Scientific Revolutions.*

Studying evaluative knowledge used by designers: towards improving critiquing systems

Nathalie Bonnardel & Laurent Joblet
CREPCO - URA CNRS 182
Université de Provence
13621 Aix en Provence, France
Tel: (+33) 42 95 37 40, Fax: (+33) 42 20 59 05
nathb@aixup.univ-aix.fr

ABSTRACT
From the perspective of developing a knowledge-based system comprising a critiquing component (which supports evaluation) for students in architectural design, we conducted an experimental study to identify and characterize their evaluative knowledge. The participants to this study had to perform two phases of evaluation of a real architectural plan and a phase of justification of the identified errors. The analysis of these different phases shows unexpected results and allow us to suggest new developments with regard to an existing support system.

Keywords
design activity, evaluation, expertise, support systems.

INTRODUCTION
Designers have to be creative as well as to express evaluative judgments. Indeed, they have to generate a design solution (i.e., to define the features of the product to be designed), which involves some level of creativity depending on the novelty of the design problem. To reach the final design, they propose and evaluate successive partial solutions, and the results of each solution evaluation step determine the following of the design activity (Bonnardel, 1992). Therefore, evaluation processes play a major role in design problem-solving. However, evaluation of design solutions is difficult for both experienced and inexperienced designers due, at least, to two reasons (Bonnardel & Sumner, in press):

- limitations in the individual competence of any designer (in complex domains, no single person can have all the relevant evaluative knowledge);
- the necessity to adopt a multidisciplinary point of view (design solutions must be evaluated from multiple, and sometimes conflicting, perspectives).

In addition, during their training, students in design are focused on developing design solutions more than on assessing them. They are provided especially with methods for constructing design solutions but with no or few methods for evaluating such solutions.

Certain computational systems, termed "critiquing systems" have been developed to support designers in constructing and evaluating design solutions. Such systems allow the users to define the solutions they wish, and support evaluation activities by analyzing design solutions for compliance with design rules or principles encoded in the system's knowledge-base (see Fischer et al. 1991). In addition to support current states of solving, we believe critiquing systems can facilitate the learning of evaluative rules or principles.

Towards the end of developing a critiquing system which could both support evaluation and facilitate learning of evaluative knowledge, we conducted a study in a school of architectural design. This study aimed at identifying and characterizing the evaluative knowledge used by novice and experienced students. First, we present this study and the results it allowed us to obtain. Then, we provide some interpretations of our findings and complement them by analyzing an existing critiquing system for architectural design in order to develop of a specific system comprising different types of modules.

EXPERIMENTAL STUDY
This study was conducted in the context of a pedagogical project with an Architectural School (Marseille-Luminy). Our basic hypothesis, logic in a teaching context, was that an evolution of expertise occurs from the beginning of the training to its end (5 years).

Subjects
Seven students participated in this study, four in their second year (referred to here as 'novices') and three in their last year of architectural school ('experienced', though not yet expert).

Experimental task
We asked them to judge an architectural plan derived from a real plan made by a famous architect: Le Corbusier. The evaluation consisted of two phases:

- first, the subjects had to globally assess the plan as correct or not;
- second, they had to point out the parts of the plan they judged as incorrect. In this case, it is a detailed evaluation since it requires to analyze in details the design solution.

These two phases were computer-monitored in order to homogenize the presentation of the experiment and to collect response times.

Once they pointed out all the parts they considered incorrect, we asked them to explain why. Verbalizations were post-experimental to avoid subjects forgetting errors in the plan due to a simultaneous verbalization.

Theoretical frame used for data analysis

The data analysis was conducted with regard to previous findings about evaluative procedures and parameters, and with regard to a characterization of expertise.

Previous research, on the design of aerospace products, showed that professional designers can use different procedures of evaluation (Bonnardel, 1992), among them:

- an analytical procedure, which is used to determine the advantages and disadvantages of a design solution with regard to various criteria or constraints (for example, designers may consider that a solution is "too heavy" or "not rigid enough ");
- a global procedure, which enables mainly expert or experienced designers to determine *a priori* and apparently intuitively (i.e., without a deep analysis) whether a solution can fit or not a given problem;
- an analogical procedure, which allows the designers to assess the current solution by transferring a judgment made for a previous solution to an analog design problem.

The same research also showed that designers refer to various evaluative parameters — these parameters are called *evaluative referents*. Different types of evaluative referents can be distinguished based on their level of abstraction and on their nature and origin (Bonnardel, 1989). Concerning the level of abstraction, we distinguish conceptual or abstract evaluative referents from the ones that are operational or concrete (linked to technical features of the design solution). We use the term "criteria" to designate the first ones (since they are used as directive principles), and the term "constraints" for the second (since operational evaluative referents constrain the choice of design options and, therefore, limit the space of possible solution elements). Distinctions concerning the nature of evaluative referents will be described in the section "results".

Among the many papers on expertise, Rasmussen's (1986) model is particularly apposite, since it describes the diagnostic activity which appears to be close to our subjects' task identifying incorrect parts of the plan. The major interest of this model, for us, is that it gives an account of the reasoning of both novices and experts. It shows that experts benefit from heuristics and shortcuts in the decision process, which allow them to avoid high level treatments. Experts can develop knowledge- and rule-based behaviour as well as "skill-based behaviour" (Rasmussen, 1988). Such a skill-based reasoning may be based on "functional knowledge" i.e. action-driven and differentiated according to the contexts of use. In this approach, if a knowledge element is used in different contexts, it will be memorized several times linked to each of the contexts (Bastien, 1988). Experts also have constructed "frames" (Minsky, 1975), which can allow them to be immediately more efficient than novices. Such frames can be instantiated with information elements linked to the current situation experts are dealing with. Therefore, problem understanding and/or action execution could be facilitated. In addition, certain research works show that experts' mental load in working memory can be alleviated with regard to novices' (e.g., Ericsson & Kintsch, 1995). For Ericsson and Kintsch, practice and training can enable individuals to use long-term memory as an efficient extension of short-term or working memory. Mnemonic elements would be stored in a "long-term working memory" and kept directly accessible by means of retrieval cues in working memory

Hypotheses

Our basic assumption was that experienced subjects have more architectural knowledge than the novices. Combining this assumption and the above previous findings allowed us to define several hypotheses:

- Reading and treatment of the problem definition will take longer for novices than for experienced subjects, who could benefit from the use of appropriate frames (facilitating the problem understanding) and could have an lightened mental load.
- During the first phase of evaluation, the global assessment time will be also higher for novices (time for analytical evaluation is relative to the number of errors identified).
- During the detailed evaluation, experienced subjects will identify more errors than novices.
- In addition, we expect that experienced and novice subjects will identify errors of different nature.
- To identify errors of different nature, experienced and novice subjects will take into account evaluative referents of different nature. These evaluative referents should underlie, during the verbalization phase, justifications of identified errors.

Results

Quantitative analysis

First, we observed that novices read the problem definition for 113 seconds (1'53") against 93 seconds (1'33") for experienced subjects. It is not significant but it is in the direction of our first hypothesis.

Contrary to our expectations, during the first phase of evaluation, the average time of global assessment wass lower for novices than for experienced subjects: respectively, 197 seconds (3'17") vs. 282 seconds (4'42"). Moreover, during the detailed evaluation, novices identify, on average, more errors than experienced subjects: respectively 10.75 vs. 8.65 errors. Though it is not an important difference, it is not in the direction we expected. Different interpretations can be invoked based on the theoretical framework previously presented:

- While performing the detailed evaluation (phase 2), the global procedure of evaluation may be so present for experienced subjects that they have difficulty in using an analytical procedure, though it would have allowed them to identify errors related to specific parts of the plan. Therefore, they identify fewer errors than novices.

  In addition, during phase 1 of evaluation, experienced subjects expect this solution to be erroneous (due to the questions asked in the experimental situation), so they persist in their attempts to identify errors, which increases the time of global assessment of the solution.

(Note to the readers: Do not confuse the global assessment which is produced during the first experimental phase, and the global procedure of evaluation which can be used by the subjects while performing the detailed evaluation).

- Experienced subjects may be focused on certain types or categories of errors they consider primary for a stage of plan analysis or with regard to their specialization in architectural design; whereas novices still have a general training comprising several points of view. Therefore, experienced subjects would look mainly for certain specific types of errors. This would explain that experienced subjects identify less errors than novices. In addition, the identification of such errors could require a deeper analysis, which would take more time than the analyses performed by novices. Moreover, the identification of specific errors could require that experienced subjects consider specific evaluative referents.

Testing the first interpretation would require another experiment. To determine whether the second interpretation can be confirmed, we needed to characterize the nature of the errors identified by the subjects as well as the nature of the evaluative referents they take into account. This motivated a qualitative analysis of their verbalizations (following subsection).

The verbalizations analysis was also quantitative since we observed that both experienced and novice subjects discovered new errors while explaining the errors they had identified during the phase of detailed evaluation. However, experienced subjects discovered, in total, much more errors than novices (see Table 1).

Phase of error identification	Novice subjects	Experienced subjects
Detailed eval°	10.75	8.65
Verbalizations	2.25	24.32
Total	13.0	32.97

Table 1: Average number of identified errors according to the experimental phases and the subjects' level of expertise.

Such new results for the experienced subjects seem to be an effect of the verbalization activity, which has been shown to modify the main activity, either positively or negatively depending on the situation (Ericsson & Simon, 1980; Ahlum-Heath & Di Vesta, 1986). It allows us to pursue the two previous interpretations:

- The verbalizations could prompt experienced subjects to change of evaluative procedure. The fact of having to explain the errors they previously found obliges them to look for justifications, which can have, at least, two effects:
  - It leads them to analyze the plan more precisely. Looking at the errors they previously found, they discover new errors topographically close to the first ones. Therefore, it would be a bottom-up process.
  - It leads them to consider explicitly criteria and/or constraints that can be used to justify the errors they previously found (though such criteria and constraints would not have been consciously taken into account during the global assessment). The criteria and/or constraints they refer to for justifying these errors can be related to other criteria or constraints stocked in memory (see Bonnardel, 1996). The criteria and constraints newly activated would lead experienced subjects to discover new errors in the plan. Therefore, it would be a top-down process.
- With regard to the second previous interpretation, the verbalizations could prompt experienced subjects to also take into account categories of errors they would have considered, otherwise, to be secondary during the phase of detailed evaluation. Based on this interpretation, during the verbalizations, experienced subjects would identify errors that are different in nature from the ones previously identified.

Qualitative analysis

To characterize the nature of the errors identified by novice and experienced subjects, we have kept a typology used by professionals in architectural design. It comprises four types of possible errors:

- *functional errors* correspond to errors which will bring discomfort to occupants of the house (e.g., needing to cross one bedroom to reach another does not respect privacy of the first one);
- *graphical errors* are drawing errors which make the reading of the plan difficult (e.g., the direction of an arrow, contrary the standard in use);
- *aesthetic errors* reflect architects' subjective opinions (e.g., a window that seems too large to a subject based on his own feeling);
- *constructive errors* correspond to errors that will really appear during the construction of the house (e.g., the fact of not having planned a place for ventilation shaft from one floor to another).

In addition, we had to add a fifth category *(misunderstood errors)* corresponding to parts of the plan subjects found difficult to understand. Since they cannot understand these parts they decide to point out them as erroneous.

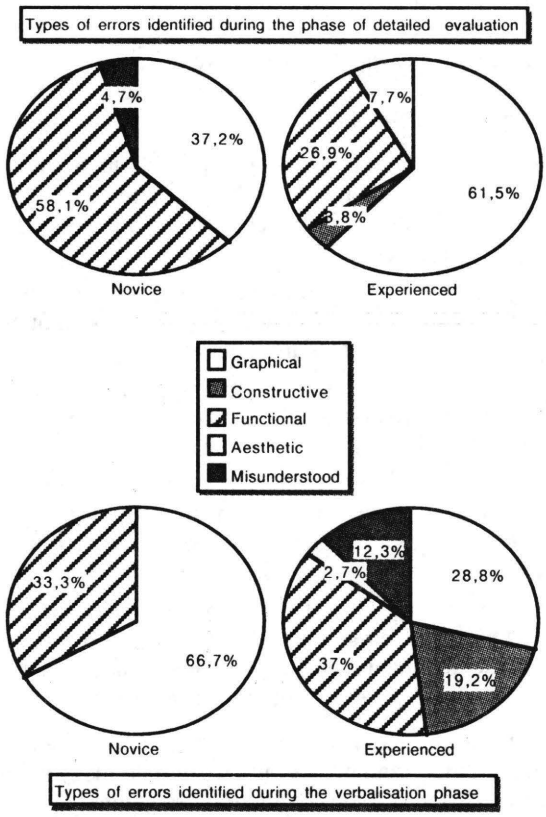
Concerning the nature of the errors, our hypothesis was that experienced and novice subjects would identify errors of different nature. To provide answer elements, we need to consider both the phases of detailed evaluation and verbalization (since the subjects discovered new errors during the verbalizations).

During the phase of detailed evaluation, novices identify mainly functional errors (58.1% of the total of identified errors - see upper part of Figure 1). It shows that they can easily imagine the way of life and moves of the occupants of the house. Moreover, they identify, on average, more functional errors than experienced subjects (6.25 vs. 2.33). Experienced subjects identify mainly graphical errors (61.55%), which shows that they are first focused on objective features of the plan: the respect of a graphical standard. However, we can see, on Table 2, that they only identify slightly more graphical errors than novices. In addition, contrary to novices, they identify a few aesthetic and constructive errors, which imply respectively the

adoption of a personal point of view and the one of the specialists who will be in charge of building the house.

During the phase of verbalization, the novices take graphical errors more into account (66.7% of the identified errors - see lower part of Figure 1). Nevertheless, it only corresponds to a few of the graphical errors (1.5), since they only identify a low number of errors during the verbalizations. In contrast, experienced subjects identify more errors than during the phase of detailed evaluation. These errors appear to be more evenly distirbuted among three categories: 37% functional , 28.8% graphical, and 19.2% constructive, the average number of aesthetic errors staying the same. It shows that experienced subjects can adopt more varied points of view than novices.

Figure 1: Nature and percentage of errors identified by novice and experienced subjects.



In addition, we observe that experienced subjects have difficulties in understanding some parts of the plan (12.3% of misunderstood errors). They may be used to judging their own plans (while drawing them) more than preexisting ones, and we can think that they develop certain own graphical habits (in addition to the graphical standard). Therefore, they can feel disconcert when they have to judge a plan provided by someone else. A complementary interpretation is that both experienced and novice subjects have to deal with misunderstood, but as novices are less secure in their feelings they prefer not to verbalize them. Such an explanation could also apply to aesthetic errors

.

Nature of errors	Novice subjects	Experienced subjects
Graphical	4	5.33
Constructive	0	0.33
Functional	6.25	2.33
Aesthetic	0	0.66
Misunderstood	0.5	0

Table 2: Average number of the different types of errors identified by novice and experienced subjects during the phase of detailed evaluation.

Nature of errors	Novice subjects	Experienced subjects
Graphical	1.5	7
Constructive	0	4.66
Functional	0.75	9
Aesthetic	0	0.66
Misunderstood	0	3

Table 3: Average number of the different types of errors identified by novice and experienced subjects during the verbalization phase.

Nature of the evaluative referents

Our hypothesis was that during the post-experimental verbalizations, experienced and novice subjects would refer to evaluative referents of different nature and level of abstraction, to justify the identified errors.

To provide information on this topic, we analyzed how the subjects justified the errors they identified. This analysis was conducted with regard to the distinction between criteria and constraints (presented in the section "theoretical frame used for data analysis") as well as a distinction among three types of constraints (Bonnardel, 1989):

- *prescribed constraints*, which correspond to constraints derived and selected from the problem's specifications (when designers have to define a product with regard to a specific request);
- *deduced constraints*, which result from an analysis of the consequences of the current state of problem-solving and/or consequences of constraints that have already been defined (e.g., "since we have insulated this side of wall, we also have to insulate this other side").
- *constructed constraints*, which result from an operationalization of criteria (since these last evaluative referents may appear too abstract to be directly used at certain stages of problem-solving). The criteria being acquired by the designers as their experience grows, constructed constraints reflect their own preferences (e.g., "there is a lake of *luminosity* in this room, so we should add windows" [constructed constraint that operationalizes the previous *criterion*]).

In addition to these three types of constraints, we added a fourth category of evaluative referents to give an account of all the subjects' justifications. It consists of *conventional*, which corresponds to rules learnt by students in architectural schools (e.g., "in France, the direction of the arrow [associated with stairs] should be in the direction of climbing the stairs").

We analyzed separately the subjects' justifications; then, we compared our findings and reached a satisfactory degree of agreement.

Contrary to our hypothesis, we observed that both novice and experienced subjects refer globally to the same proportion of criteria, constructed constraints and conventions (see Table 4), so we cannot conclude to the existence of differences in the level of abstraction of the evaluative referents they take into account. Nevertheless, experienced designers express about fourth times more each type of these evaluative referents.

However, and in accordance with our hypothesis, experienced subjects also formulate deduced constraints (which is not the case for novices). Such constraints show that they are involved in improving the design problem-solving: deduced constraints allow them to define beneficial changes to perform to the current plan (see above example). Therefore, they do not only assess the plan, they try to modify it, contrary to novices.

If we relativize the number of produced justifications with regard to the number of identified errors, we notice that experienced subjects justify more each error than novices (respectively 2.1 vs. 1.3 justifications by error).

Types of justifications	Novice subjects	Experienced subjects
Criteria	5.5	24.33
Constructed constraints	6	24
Deduced constraints	0	5
Conventional	5.25	17

Table 4: Average number of the different types of justifications produced by novice and experienced subjects during the verbalization phase.

DISCUSSION

Interpretations of the findings

Our first result is that, contrary to our expectations, experienced subjects identified fewer errors than novices during the phase of detailed evaluation (moreover, experienced subjects needed more time to judge the plan as incorrect). While justifying these errors during the phase of verbalization, the subjects discovered new errors, and at this time we observed opposite results to the previous ones: experienced subjects identified much more errors than novices.

Such a double result requires some interpretations. Some possible elements of explanation have already been suggested but they needed some additional analyses:

According to our first interpretation, experienced subjects would have difficulties in changing the global procedure of evaluation for an analytical one (which would have allowed the identification of errors with regard to criteria and constraints). In the following experimental phase, verbalizations would have the effect of prompting experienced subjects to use the analytical procedure in order to justify the errors they previously found through the global procedure. The adoption of the analytical procedure would, then, lead them not only to provide justifications but also to discover new errors, either through a bottom-up process (when it is based on topographical links among errors) or a top-down process (when it is based on thematic or influential relationships among criteria and constraints).

The analysis of the place of the newly discovered errors tend to show that the links among the errors depend on the subjects' level of expertise:

- though a few novices discovered new errors during the verbalization phase, when they did it, all of them used a "topographical search strategy" (identifying errors that are closed to the ones they are justifying);
- experienced subjects used both bottom-up and top-down processes of error search; each for about half of the newly discovered errors (57.7% of them result from a topographical search strategy).

A similar topographical search strategy has been described by Rasmussen, in a different area (plant process control), as "context free" and conducted with regard to a model of normal system conditions (Rasmussen, 1988). In our study, this strategy does not seem context free since it leads to different errors depending on the subjects' points of view (e.g., functional vs. constructive points of view).

Anyway, we may think that students in architectural design may progressively change of error search strategy, going from a bottom-up to a top-down approach as they acquire expertise. Therefore, a complementary study would have to be performed with experts in order to characterize the "final" search strategy(ies) they use.

• Another interpretation was that, during the phase of detailed evaluation, experienced subjects would have been mainly focused on categories of errors they consider primary for a stage of plan analysis and/or with regard to their specialization in architectural design (contrary to novices who still would adopt a diverse approach though less precise). Then, the verbalizations would prompt experienced subjects to consider also "secondary" categories.

Results from the qualitative analysis of the identified errors show that, in accordance to the previous interpretation, experienced subjects identified errors belonging to a supplementary category of errors (constructive errors, which require to adopt the point of view of specialists who will be in charge of house building).

Towards developing a critiquing system for students in architectural design

Among the various critiquing systems that exist, the KID system (Nakakoji, 1993) – an extension of JANUS (Fischer et al., 1991) – has been developed to support the evaluation of kitchen floor plans. Therefore, it constitutes

our main reference in architectural design. This system integrates five modules (Nakakoji, ibid.):

1) a *construction kit* , which allows the user to construct plans by selecting design units from a palette of items and placing them into a work area (in a direct manipulation style);
2) a *specification component,* which allows the user to specify design intentions linked to the customers' characteristics;
3) a *critiquing module,* which analyzes the design in the work area for compliance with a critic-rule-base; when problematic design unit configurations are identified, the user is provided with critiquing messages (e.g., "the stove is not away from a door");
4) an *argumentation-base,* which contains design rationale that is used to explain to the user why and how the critiquing message is related to his/her current plan and/or customers' characteristics (e.g., safety arguments: "by placing the stove too close to a door, it will be a fire and burn hazard to unsuspected passersby (such as small children).");
5) a *catalog-base* composed of completed plans together with their associated specifications, to allow the user to reuse previous design solutions (or parts of them).

We believe such a system is very interesting and can serve as a model for developing a support system for novice and experienced students in architecture. However, our findings allow us to suggest some developments to add to such an approach to suit the users' levels of expertise better.

Concerning novices, our findings tend to show that they need a traditional critiquing support to reach a more complete evaluation of plans (the final number of errors identified by novices being less than half of the one of experienced subjects). In addition, contrary to experienced subjects, they do not try to improve the presented plan, so they could also benefit from a support for improving current design solutions.

Concerning experienced designers, we observed that they adopt a "secondary" point of view only lately, as a consequence of verbalizations. Therefore, they could benefit from a support system that would directly provide them with "constructive information", i.e., critiquing messages based on the point of view of the specialists who will be in charge of house building.

Therefore, we suggest the development of three complementary functions:

- The system should not only support evaluation but also provide the user with propositions of modifications for the current design solution, when he/she asks for them.
- The critic-rule-base should not only include standard about architectural design itself but also standard about house building. In this case, the supplementary critiquing messages could be presented when the user asks for them, or in a separate window on the screen, or else after a presentation of primary critiquing messages (reflecting an order of priority).
- The specification component should not only allow the user to specify customers' characteristics but also include a procedure to identify the user's preferences (or architectural styles). Such an identification could be derived from a presentation of a few "typical" solutions that the user would have to assess before really using the system as a computational support. These solutions would be built in order to satisfy, at different degrees (including "not at all"), certain criteria and/or constraints representative of main architectural styles.

ACKNOWLEDGMENTS

We wish to thank very much J.-C. Lebahar and his students of the Ecole d'Architecture (Marseille-Luminy) for their collaboration to this study.

REFERENCES

Ahlum-Heath, M.E. & Di Vesta, F.J. (1986). The effect of conscious controlled verbalisation of a cognitive strategy on transfer in problem solving. *Memory and Cognition*, 14(3), 281-285.

Bastien, C. (1988). Architecture du système cognitif et individualisation des conduites. *Archives de Psychologie*, 56, 245-250.

Bonnardel, N. (1996). Les topoï: un mode de représentation des connaissances en conception de produits aérospatiaux. In P.-Y. Raccah (Ed.). *Topoï et Gestion de la Connaissance*, Collection Sciences Cognitives, Paris: Masson, 41-56.

Bonnardel, N. (1992). *Le rôle de l'évaluation dans les activités de conception.* Thèse de Doctorat de l'Université de Provence.

Bonnardel, N. (1989). *L'évaluation de solutions dans la résolution de problèmes de conception.* (Rapport de Recherche INRIA n°1072). Rocquencourt : INRIA.

Bonnardel, N. & Sumner, T. (in press). Supporting evaluation in design. *Acta Psychologica.*

Ericsson, K.A. & Kintsch W. (1995). Long-term working memory. *Psychological review*, 102(2), 211-245.

Ericsson, K.A. & Simon, H.A. (1980). Verbal report as data. *Psychological Review*, 87, 215-251.

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.

Minsky, M. (1975). A framework for representing knowledge. In P.M. Winston (Ed.), *The psychology of computer vision.* New-York: Mc Green Hill.

Nakakoji, K. (1993). *Increasing Shared Knowledge Between Humans and Design Environments: The Role of a Specification Component.* Ph.D. thesis, Department of Computer Science, University of Colorado.

Rasmussen, J. (1986). *Information processing and human-machine interaction.* Amsterdam: North-Holland.

Rasmussen, J. & Goodstein, L.P. (1988). Information technology and work. In M. Helander (ed.), *Handbook of Human-Computer Interaction.* North-Holland: Elsevier Science Publishers, 175- 201.

Cognitive Function Analysis

Guy A. Boy
European Institute of Cognitive Sciences and Engineering (EURISCO)
4, Avenue Edouard Belin, 31400 Toulouse, France
Tel: (33) 62.17.38.38
E-Mail: boy@onecert.fr

ABSTRACT
This paper presents the Cognitive Function Analysis (CFA), an analytical method that supporrts participatory design and usability testing. It enables design teams to elicit and visualize cognitive functions involved in dialogues between users and complex dynamic systems. CFA has been successfully used in the redesign of the user interface of a flight management system onboard aircraft. This short paper discusses important properties of such a method and the lessons learned so far.

Keywords
cognitive modeling, knowledge elicitation, participatory design, evaluation, methodology.

DEFINITION
According to French tradition in ergonomics, a distinction is made between *task* (prescribed task) and *activity* (effective task). When a job needs to be performed, it can be decomposed into tasks. Thus, tasks (at any level) are the requirements of the job. When someone performs a task, he/she produces an activity. This activity may differ from the original requirements, i.e., the effective task differs from the prescribed task. This is due to the fact that it is very difficult to anticipate all possible situations when tasks are designed. A human operator facing a situation produces a specific activity according to the requirements, but may need to *adapt* the task to this situation. A *cognitive function* is the representation of the mechanisms that produce an activity as the result of the excution of a task. Examples of cognitive functions in aeronautics are: clearance transmission; frequency change; rolling authorization; parking space allocation; flight level allocation; etc.

PROPERTIES
A cognitive function (CF) can be decomposed into smaller grain cognitive functions. In particular, a CF invoked for performing a specific task can be decomposed into two main CFs: the CF devoted to the interaction (I-CF): this is related to the syntax used to perform the task; the CF devoted to the content of the task itself (TC-CF): this is related to the semantics of the task. An easy-to-use user interface results in simple CF interaction components. A counter example is the current interface called the multifunction command and display unit (MCDU) used to program a flight plan. Many pilots do not like to use the MCDU because they do not like its lateral location, its alphanumerical keyboard, its overloaded screen, its complex menu system, etc. A study was carried out at EURISCO to improve the current concept of MCDU. The pilot needs to push keys (e.g., a first interaction sub-CF is devoted to locating the right key and checking what the computer response is), browse menus that can be more or less complicated due to the depth or recursion of these menus, etc. The interaction component of the overall CF can be complicated. CFs devoted to the content of the task are concerned with processes such as minimizing the distance between two points, satisfying a constraint imposed by Air Traffic Control (ATC), etc. A complex cognitive function such as programming a flight plan using a MCDU can usually be decomposed into a CF graph. The first layer of blocks mainly represents TC-CFs. For instance, the *Preflight* TC-CF is decomposed into three TC-CFs *Setting up*, *Flight plan preparation*, and *Performance*. The *Setting up* TC-CF is then decomposed into two TC-CFs *System status check*, and *Navaids deselection*. The *System status check* TC-CF is conditionally (*If A/C Status page is not displayed*) decomposed into four I-CFs *Depress **'DATA'** key*, *Select **'A/C STATUS'***, *Check Database period of validity*, and *Check Clock/Date*. A cognitive analysis of the flight programming task enabled the elicitation of 8 primitive cognitive functions: *browse*, *check*, *depress*, *enter*, *insert*, *modify*, *select* and *set*. Each more complex cognitive function is a composition of these primitives.

USER-CENTERED AUTOMATION
Note that TC-CF can also be transferred to the machine. I-CF transfer defines an automation that transforms a physical interaction into a more cognitive interaction. TC-CF transfer defines an automation that distributes the responsibility of the task between the human and the machine. The way TC-CF are understood by designers is crucial because a transferred TC-CF has necessarily a limited autonomy. The way this autonomy is perceived by end-users is an important issue. Sometimes users may not anticipate this limited autonomy because the system (using the transferred TC-CF) functions well most of the time.

In the MCDU experiment, it was decided to remove the alphanumerical keyboard, and to introduce a larger navigation display (ND) screen and a direct manipulation facility such as a trackball to manage information on the screen, as an alternative to the current one. We have consequently developed a set of cognitive functions that are used by pilots to program flight plans. The efficiency of these functions was assessed using a GOMS analysis

(Card et al., 1983). This analysis was performed with the help of operational people. It showed that the alternative to the current setup is much better, i.e., it takes much less time is needed to access planning functions with this alternative design.

Transferring cognitive functions from humans to machines is also called automation. For instance, technology-mediated human-human communication can be greatly enhanced by directing tedious and time-consuming I-CFs towards the machine. *Human-centered automation* (HCA) from a cognitive function viewpoint consists in eliciting and constructing the cognitive functions that are involved in a specific task, facilitating the understanding of what should be automated according to a set of principles and guidelines. First of all, the transfer of cognitive functions from the human operator to the machine requires the definition of decision criteria expressed in terms of attributes of automation. Psychological factors such as boredom, complacency, competency erosion, fatigue or workload can be used to assess the relevance of CF transfer. According to Billings's attributes of automation (Billings, 1991), a categorization of I-CF and TC-CF transfer has be developed. I-CF tranfser can be associated with accountability, predictability, comprehensibility, dependability and error-resistance. TC-CF transfer can be associated with subordinatability, adaptability, flexibility, informativity and error-tolerance.

THE CFA METHOD

Cognitive simulation enables *domain experts to assess the relevance and correctness of the cognitive function analysis* (CFA). In particular, the level of *granularity* of the representation is essential in order to capture relevant and necessary cognitive functions useful in air-ground communications. The CFA method consists of the following steps:

- define a first sub optimal set of contexts using classical elicitation methods such as interviews, brainstorming, field studies;
- develop the knowledge blocks that describe processes responsible for both pilots' and controllers' activities within these contexts;
- run a cognitive simulation and assess the relevance and correctness of it;
- refine current contexts and knowledge blocks;
- run a new cognitive simulation and assess the relevance and correctness of it.

In the MCDU example, each cognitive function interface artifact was designed on a Hypercard card that gave the user the illusion of a "real" navigation display screen. The use of such a hypermedia tool enables the generation of active documents that can both simulate real-world behavior and provide on-demand explanations. This example has demonstrated that the use of such a cognitive simulation for rationalizing design decisions can be extremely useful in explaining why a part of a device has been specifically designed. This approach can be understood by both designers and end users. Cognitive functions represented as knowledge blocks provide a powerful tool for mediating interaction between these practitioners. Furthermore, blocks enable the incremental storage of design rationale in context, and easy retrieval and better interpretation of design rationale in context.

FIRST LESSONS LEARNED

The design of CFs consists in writing dialogue specifications. CFs are incrementally generated as cases that can be considered as active documents. Document management is already a field of investigation that includes indexing and information retrieval, creation of new documents, modification of existing documents, organization of libraries of documents, etc. In general people tend to facilitate document management by analogy. Memorizing cases of successes or failures enables better management of future analogous cases.

Cognitive functions can be grouped into categories or classes for easier storage and retrieval. In this book, CFs are categorized and structured according to context. Context structures play a significant role in the management of large CF bases. If cognitive functions are considered as active documents, they can be grouped into libraries. The larger a library is, the more a librarian is needed to help store and retrieve documents. This type of librarian agent is itself a cognitive function. It can be considered as a meta-CF. The librarian CF includes several CFs addressing various uses of other CFs as documents. This perspective enables anyone to access and use CFs developed by others. A CF base is an *external memory* useful for analysis, design and evaluation in the field of human-machine interaction.

CF elicitation is intrinsically an educative process. We often learn better from stories (episodes). However, these stories need to be further structured to be retrieved at the right time in the right format. CFs are first elicited in a natural language form and subsequently indexed in context. The main problem is to find the right level of detail that guarantees good understandability for reuse.

ACKNOWLEDGEMENTS

Helen Wilson provided astute advice towards improving the quality of this paper.

REFERENCES

Billings, C.E. (1991). *Human-centered aircraft automation: A concept and guidelines.* NASA Technical Memorandum 103885. Moffett Field, CA.

Boy, G.A. & L'Ebraly, H. (1994). Modèle de Tâche et Représentation de l'Expertise pour la Conception de Dispositifs d'Interaction Homme-Machine. Exemple d'Application en Aéronautique. EURISCO Technical Report no. T-94-009-VI.

Card, S.K., Moran, T.P. & Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Lawrence Erlbaum Ass., Hillsdale, NJ.

SimulACRUM: A cognitive architecture for modelling complex task performance and error

Simon Grant

EuropeanCommission
Joint Research Centre
Institute for Systems, Informatics and Safety,
Ispra (VA):
Italy

City University
Northampton Square
London EC1V 0HB
simon@city.ac.uk
http://www.city.ac.uk/~sd351/home.html

ABSTRACT

Three examples from the literature demonstrate the kind of phenomena that need to be covered by cognitive models of error. A novel cognitive architecture for the modelling of human behaviour and error is here outlined. It is based on classes, and instances, of entities, and of situations, represented in a contextually modular way as task units. Crucially, there are two kinds of transition between task unit instances: in known situations it is rule-based, while in unusual cases it is governed by matching the current state representation against normal state envelopes of 'usual', expected states.

Keywords

cognitive simulation models; cognitive architecture; error; complex systems; cognitive ergonomics.

INTRODUCTION

Analysis of currently available cognitive architectures shows that there are gaps in the treatment of human errors in complex task performance. On the one hand, well-known 'architectures' such as Soar (Newell 1990) and ACT-R (Anderson 1993) have been predominantly used to model learning or skill acquisition, which is not often highly salient in errors found in complex systems. Rather, we see mode errors, fixation errors and other kinds of phenomena that are often clearer to the domain practitioner than to the cognitive modeller. There has been extensive treatment of such practical themes in the literature, for instance by Reason (1990), Hollnagel (1993), and Woods, Johannesen, Cook, & Sarter (1994). Decision making under time constraints is important, and the concept of situation awareness is prominent, notably in tasks such as aviation.

On the other hand, the foremost work in modelling and simulating complex tasks, e.g., (Cacciabue *et al.* 1992), though it takes theory from cognitive science, draws the underlying architecture more from computer science, and thus lacks detailed correspondence between the mechanisms of the architecture and plausible mechanisms implementing effects recognised by cognitive psychology. The philosophical point of view also gets little attention.

The challenge is to devise a more realistic cognitive architecture that can provide the basis for meeting the practical modelling requirements of cognitive ergonomics. We shall first consider some examples from the literature of the kind of human error that needs to be modelled.

EXAMPLES OF ERRORS FOR MODELLING

Brief discussion of three examples of human error taken from the literature indicates issues which are important for the cognitive architecture underlying any adequate model of the cognitive processes responsible for these errors.

An Unexpected Maintenance Event

Woods (1994, p.40) describes the incident of the missing O-rings in the Eastern L1011 in May 1983 in more detail than here. Mechanics usually replaced the aircraft engines' 'master chip detectors' complete with O-rings that had been fitted by someone else—helpfully, but against the exact definition of the procedure. When the usual place to get them was empty, they got others, but these did not have O-rings ready fitted. They did not notice that the O-rings were missing, and checks were inadequate to spot the mistake. This led to all the engines malfunctioning during the next flight.

This suggests that the relevant skill of an operator is not identical to the procedure manual, and that an analysis of error that is limited to deviations from written procedures is inadequate. The operator's error in this case (and many others) is not a deficiency in routine knowledge, for the situation in which the mechanics found themselves normally was adequately covered by their skill. The error arises from the combination of a particular level of skilled knowledge combined with some kind of unexpected or unusual situation.

Three Mile Island

Reason (1990, p.251; pp.189-191) gives a useful summary covering the main chain of events, active errors, contributing conditions, and latent failures. The accident is so well-known that there is no need to describe the scenario here.

A cognitive explanation of the accident depends critically on a model of how the operators' limited knowledge of, and limited access to information led them to suspect that the problem was other than it actually was. The relief valve (PORV) was stuck open, but the indicator said closed: we need to model the fact that based on the available observations that were likely to be made at the time (which itself is based on how the situation was being perceived) the diagnosis they made was plausible. We should also model the fact that the observed evidence did not produce a timely shift of view—in this case, to a more realistic

recognition of the plant state including a misleading indicator.

There were important failures of management, regulation, design, training and maintenance, pointed out by Reason, without which the accident could have been avoided. However, from the point of view of analysing the immediate causes of the accident, these wider organisational (or latent) factors can be seen as either affecting the availability of information to the operators, or affecting (or limiting) their operationally effective task skill to respond in appropriate ways to the information provided.

A Well-Studied Air Accident

The official report of the Mont Sainte-Odile (approaching Strasbourg, France) aircraft accident was by the commission of enquiry (Monnier *et al.* 1993). One clearly possible, and commonly acknowledged, interpretation of the accident is that one of the pilots mistook the mode of the flight management system, entering a descent rate of 3300 feet per minute instead of a descent angle of 3.3 degrees, which were possibly confusable. (The design has since been changed to be less confusing in this respect.) Perhaps the surprising aspect is that neither pilot noticed the effects of the mistake, which we can easily imagine as fairly obvious. This was probably helped by the fact that the flight was in the dark, and that complications in the approach demanded abnormally high attention and workload.

To get the error in perspective, it is helpful to remember that no more than about one minute elapsed between the presumed wrong setting and the crash itself. This minute was one in which they were concentrating on another task, presumably not suspecting that they had any need to check on the progress of what was being taken care of automatically by the flight management system.

A cognitive model of this accident needs to take into account the effect of high workload, and relative inexperience, on which variables would be noticed, and the inferences drawn from those variables in terms of whether there was anything amiss with the progress of the flight. One common factor in air accidents of the type "controlled flight into terrain" (CFIT) is that there is no awareness, at least until it is too late, that there is anything wrong with the flight path. The situation awareness has diverged from what it should be — consistent with reality.

General points

A key feature needed to model this kind of accident (or similar incidents or events) is the combination of a rule-based approach to normal situations, together with the ability to predict what will happen when the situation is changed in some unexpected way, in terms of the humans' situation awareness (that is, how the human understands what is happening). In order to do this, one needs to go below the level of the complete rule. The approach that emerges from the examples given is to consider what the human is likely to observe, and what known situation will be inferred by the human as being the most likely one given what is observed, and given the other factors in the situation.

The easier step from there is to make a reasonable explanation of any particular accident. This needs an intuitive understanding of how human cognition is likely to work in the given situation. But the harder step of making a predictive model, or even a cognitive simulation of the behaviour noted, needs those insights on the workings of human cognition to be formulated and implemented.

CONSTRUCTING A COGNITIVE MODEL

The features that are necessary for a model of this kind focus around the types of situation in terms of which the events are interpreted by the operator, and the variables noticed at different times, which both guide the interpretation of what the situation is, and also any action that may be called for in the situation.

In a complex system, where there are many more variables involved than can be constantly monitored at all times, the selection of variables to monitor at any time is largely dictated by the particular situation. Thus, in a modern, glass-cockpit aircraft, the pilots will select the particular information display on the screens that most closely matches their needs. However, it is also clear that some other variables, not necessarily recognised as directly required, are also salient. Thus, had the Mont Sainte-Odile scenario arisen in clear daylight, the pilots would more likely have noticed their approach to the ground, which could have interrupted their concentration on the lesser task of reaching the runway centre-line.

As in discussion above, the organisational and latent factors come into this cognitive analysis by affecting either the content and extent of the pilot's skill, or the availability of information. A mismatch of the task, the operator's skill, and the information provided can then often be seen as a responsibility of the management. Thus a cognitive model may provide another view of, and approach to, organisational factors.

The importance of the cognitive factors is common to the accidents described, and to many other examples of human error in complex tasks with complex systems, including air traffic control, the medical profession and the financial industry. This gives an incentive for building a modelling system which has built-in mechanisms capable of handling these phenomena, rather than trying to build, from scratch, a cognitive model of each different domain and each different accident. The common mechanisms amount to what is called the cognitive architecture.

DEVELOPING AN ARCHITECTURE

When considering developing architecture, it is as well to remember that there is no way that an architecture itself can be proved or disproved. This is because any normally rich computational architecture can in principle be made to model any of the same range of phenomena that is possible with any other architecture. The relative merits of different architectures must be discussed on more practical grounds, such as the convenience for practical modelling, or the size of model content code that must be written to represent certain observed phenomena. Thus, for instance, Anderson (1993) discusses the relative numbers of produc-

tions necessary to get ACT-R and Soar to do similar things.

The justification of proposing a new architecture here is that there needs to be a clearer mapping between the mechanisms suggested in the architecture and the phenomena that have been pointed out in the examples above. Other architectures do address some of these issues, and developments should therefore consider previous insights, to ensure that a new architecture is at least as useful as former ones. To discuss the merits of other architectures is, however, a long study, and this paper will confine itself to proposing a new one, hoping that the points of influence will be apparent.

As well as accommodating insights from other architectures, a new one should be philosophically reasonable, and internally coherent. One example of a philosophical discussion would be about the way in which commonly used words mapped onto concepts of the architecture. This is quite demanding for any architecture with a large scope.

However difficult the philosophy is, to be of practical worth, the architecture must be used for making models of real-life skill, and in particular, 'human error' arising directly from the exercise of that same skill. This may be expected to follow on from what is presented below.

SimulACRUM

SimulACRUM stands for Simulation Architecture for Cognition-Related User Modelling. The acronym is rather broad in meaning, and this reflects the general-purpose intention behind the architecture.

Background

SimulACRUM starts from the concept of contextual modularity (Grant 1994) in being based on cognitive task units. These are in the tradition starting with Bartlett's (1932) schemata and going through frames, scripts, and other such modular divisions of the knowledge underlying competence in a task or activity.

A task unit is a very general concept, not limited to situations we normally call tasks. It is not just the observed human behaviour, or 'activity', that defines the task unit, but the relationship of that unit of behaviour with other similar units, whether they be parts of actual tasks, games, pastimes, habits of mind, or any pattern of response to the world that is identified by the cogniser as relevantly similar to another pattern of response. Calling them 'task units' helps to focus on the relatively clear case of a task, where there is likely to be a goal. However, a goal is not necessary for a task unit in SimulACRUM: it can more simply be a pattern of response to conditions.

Most cognitive architecture focuses on some kind of task unit, be it a production rule system like ACT-R (Anderson 1993), where the unit of procedural knowledge is the production, or a model such as COSIMO (Cacciabue et al. 1992), which uses a frame system to represent task knowledge. Unlike the case in other architectures, in SimulACRUM, task units are seen as similar to, but complementary to entities, which are the things that are taken to exist, physically or conceptually. The interplay between task unit and entity is central to the architecture, and differs, for example, from ACT-R's relationship between the productions and the 'chunks' of declarative memory.

Theory

We are all familiar with classes of entity, such as animals, dogs, humans, or aircraft, and also particular individual members, called instances, of those classes: you, me, perhaps someone's dog called Spot, or the aircraft F-GGED (the one in the Mont Sainte-Odile accident). SimulACRUM adopts these straightforward ideas of entity class and instance, which have similarities with (and some differences from) object-oriented ideas. Classes of task unit are perhaps not so obvious, but in SimulACRUM, a task unit class is somewhat like a script, in that it holds the general methods for observation, evaluation, decision and action etc. for a particular kind of situation, like what to do, driving a car, at traffic lights. The instances of these task units are the episodic representations of particular events, such as the times when one has encountered traffic lights.

Thus, procedural knowledge is represented as task unit classes; episodic knowledge as task unit instances; entity classes represent much of declarative knowledge (thinking of the European verbs *savoir*, *wissen*, *sapere*) about things (with an indefinite article) and entity instances represent knowledge (*connaître*, *kennen*, *conoscere*) of particulars (with the definite article).

Behaviour

For a cognitive model built with the SimulACRUM architecture, observation of the world situation leads, by a pattern-matching process, to identification of 'what is happening', that is, the current task unit, which is instantiated from its class into a current instance (which will later become part of episodic memory). Once a task unit has been identified, that governs action as long as the situation remains within the bounds, recorded with the task unit, representing the normal range of situations appropriate to the task unit. If some part of the environment is noticed going outside these bounds, it will trigger a rematching of the changed conditions to a new task unit, if there is a more suitable one available.

These pattern-matching processes in some way resemble Reason's (1990) similarity matching and frequency gambling, embodied in COSIMO (Cacciabue *et al.* 1992). This needs a mechanism equivalent to an envelope defining where the task unit state vector (that is, the set of relevant variables with values) is considered 'normal'. It is impossible to take into account all possible observables in a real world context, but the state vector is likely to have more variables than the ones needed explicitly to feed decision-making and action rules. Any attribute can be added that has been associated with the usual situations in which the task has been practised, without the extra computational cost associated with adding extra rules.

SimulACRUM's approach to stress and emotional factors is to allow 'tension' associated with a task unit, and 'charge' with an entity. This may be related to the degree of risk perceived in the situation or object. Task units with

a high 'tension' will be instantiated more readily than ones with a low 'tension': this reflects the common experience of being preoccupied with a worrying situation. More attention will be given to monitoring the variable attributes of charged entities—the position of a wasp, or of a conflicting aircraft, for example.

In SimulACRUM, this pattern-matching form of task unit change occurs when the conditions are unusual in the sense of not having been practised, and not if a task is running according to usual, well-known conditions. In well-practised cases, the human will have learned that certain conditions may be taken to indicate reliably the appropriateness of the next task unit. This means that in task units we have rules for 'moving' to another task unit, given certain conditions—or, to be more precise, rules for instantiating a new task unit.

Structures of the architecture

The basic components of the SimulACRUM architecture are the entity and task unit classes and instances. The details will be tuned as a result of the experience of applying the architecture in the modelling of real-life cognitive skill and error.

The way that the architectural components are used as modelling elements will depend on the particular human's skill at the particular task and their degree of expertise. Thus for illustration only, the example of an ordinary motor car is given for the entity, and making dough (for bread) for the task unit. Where practical, one example case is given in parentheses along with the component: many are often possible. Entity classes may have the following components, each of which also has a variable strength of connection with the entity class.

- supertypes (road vehicle), subtypes (automatic transmission);
- superparts, subparts (engine);
- evidence rules for variable attributes (for speed, look at speedometer);
- execution rules for affordances (to slow, press foot harder on brake pedal);
- normal state envelope and dynamic (must have at least three wheels, an engine, etc.; must work);
- charge as a function of entity state (high when water temperature is near maximum).

Entity instances (e.g., my own car) then may have:

- parent entity class (as above);
- part links to other entity instances;
- evidence and execution rules which may supersede the ones from the parent class;
- current state, with uncertainties for each variable;
- current charge.

Task unit classes may have at least:

- supertypes, subtypes (making a particular kind of dough);
- superparts (making bread; pizza), subparts (kneading dough);
- task-action rules, linked to evidence and execution rules of entity classes (when the dough has risen, knead it);
- task unit instantiation rules, linked to evidence rules of entity classes (when the dough is ready, continue with making bread or pizza);
- normal state envelope and dynamic (suitable temperature of dough mixture);
- tension as a function of task unit state.

The rules and links are also connected with a variable strength to the task unit. Task unit instances (such as the time I made dough for pizza on 14th June 1996) while they are being executed, may have:

- parent task unit class;
- part links to other task unit instances;
- task-action rules, linked to evidence and execution rules of entity instances;
- task unit instantiation rules, linked to evidence rules of entity instances;
- current state, with uncertainties for each variable;
- current tension.

The normal state envelopes are perhaps the most difficult component to describe directly. For an entity class, the envelope concerns the question of what changes in the entity would make it be a member of a different class, or at least a dubious member of the class in question. The envelope may have some hard edges, linked to attributes that are considered (by the individual human) to be defining of the class. Something that doesn't have wheels simply isn't a car. But the envelope can also have less clear boundaries: would something without seats be a car?

For task units, the normal state envelope determines when the situation is no longer appropriate for that unit of task knowledge. Again, this can have sharp or hazy boundaries. The hazy, or disputed, boundaries of normal state envelopes go along with the idea that they are acquired generally by experience and not by explicit rules, and that they are used in pattern-matching rather than rule manipulating mechanisms.

The size of the classes and instances, in terms of number of rules, links, etc., is limited in some way by the capacity of what is often called working memory, and SimulACRUM will have a working memory mechanism to do this.

Another important feature is the global activation threshold, connected to general stress, or arousal, which works, together with the strength of connection of rules, and the tension and charge, in the selection of rules for firing at every level.

APPLICATION OF SIMULACRUM TO EXAMPLES

Having sketched some basics of the architecture, here it is outlined how models could be made, of the examples presented earlier, in terms of SimulACRUM. This is intended to give the beginning of a feeling for the elegance and practicality of the match between the architectural mecha-

nisms and the modelled phenomena, which is a key factor in the evaluation of an architecture.

The O-ring incident could be modelled in terms of normal state envelopes of entity and task unit classes. The usual task units would refer to the entity class 'master chip detector', with a rule about where these were to be found. Since O-rings were always fitted, the entity class 'O-ring' would not feature in the usual task units, and the detectors would be recognised perhaps using only a few characteristics. When there were none in the usual place, the question of locating some would have naturally arisen. A default rule was to get them from the stockroom, but this was not a usual procedure and therefore it would not be clear exactly what had to be done in this case. The obvious assumption for the mechanics was that the detectors obtained from the store were the same as the detectors normally found in the foreman's cabinet. There was clearly no explicit active rule to check the O-rings presence, and since they had not come across detectors without O-rings, the normal state envelope would not be crossed.

The normal task unit state envelope can be illustrated further by a feature which would be more likely to be part of it: the feel that a mechanic gets when fitting a component into place. In unexceptional conditions, one might expect that if a part fitted badly, or if it made an unusual noise when being fitted, or something of that nature, that this would have crossed the normal state envelope for the task unit of fitting the part, provoking at least a moment's reflection of "that's strange". This presumably did not happen in the case cited.

The case of Three Mile Island could be represented firstly as one of evidence rules. When the PORV stuck open, and the indicator showed closed, that information was mistakenly believed by the operators, when there could have been more detailed rules for ascertaining the actual state of the PORV. One could say that there was a deficient evidence rule. However, even if the operators had a more sophisticated rule, one would have to take into account the possibility that under stressing condition, such as alarms, the global activation threshold will rise, and short, strongly connected methods of evaluation will take priority. Thus the information may be less reliable than would be ideal, and the behaviour less predictable.

The operators were presumably already in the process of diagnosing the fault. Diagnosis, where the symptoms are not identical from one incident to the next, seems to consist of a particular kind of task unit, where those task units that are instantiated have not been experienced widely before, and where therefore there is little detail in the expectations of what to do. Operator reasoning may follow basic or primitive patterns, or use 'weak' methods. A SimulACRUM model, through built-in architectural limitations on attention range and memory capacity, would be able to suggest what problem-solving strategies would be plausible, but unlike architectures such as Soar (Newell 1990) it would not rely on just one normative problem-solving method, as these can vary greatly between individuals. Instead, the problem-solving method would be an emergent characteristic of the task unit rules.

The Mont Sainte-Odile accident could be modelled in a number of steps. The initial supposed error could be a lack of care with verification of a procedure, due to time pressure. Here again the strengths of connection of the rules come into play, that when there is time pressure, the less strongly connected ones are left out—common in everyday experience of time pressure. Perhaps more importantly, the fact that this was a real danger appears not to have been known by the pilots. Knowing about the dangers could be modelled either by stronger connections of the feedback evidence rule with the execution rule, or with a higher 'tension' of the task unit. Higher tension of the task unit, or higher charge of the entity, is what one may expect when a pilot has had near-accident experiences (possibly in simulator training) connected with that task unit or entity.

Then there was the failure to notice the abnormal indications that showed that the aircraft was not behaving as it should. This loss of situation awareness could be attributed to task pressure, and needs to be modelled in terms of attentional resources, salience of cues, etc. It seems likely that a ground proximity warning device would have succeeded in calling the attention of the crew to the loss of situation awareness.

Some practical points can be made in response to the analysis of these kinds of problem. They all could be helped by more operator experience of the problem itself. The difficulty here is to know what experience to give in training; and to select the problems to be worked on in simulation. One factor that emerges from the modelling architecture is that it is important matter for training to get the operators' normal state envelopes right—that is, they must be alert to the changes of situation variables which signify something being amiss. SimulACRUM suggests that this is different from training in the rules themselves, or even explicit rules for diagnosis of problems, because it uses a different mechanism. This could be investigated experimentally.

For practical applications, very detailed modelling is needed if all these mechanisms are to be represented. A full-scale study of a practical task would need very substantial resources, and this would be necessary to develop the architecture in a valuable way.

WHAT SIMULACRUM IS NOT

SimulACRUM at present does not have a model of learning, as do Soar and ACT-R. While learning is vital to complex task training, it is rarely central to operation. A model of learning needs an underlying model of knowledge, and so developments of SimulACRUM will provide the basis for a modelling learning in a more realistic cognitive way.

Also by way of disclaimer, SimulACRUM does not deal directly with interpersonal systems, but it represents a basis in individual cognition for doing so.

CONCLUDING COMMENTS

Architectural mechanisms emphasised in the applications above include particularly those ones that are not simply rule-based in a straightforward way: the normal state enve-

lopes; the variable strength of connection of rules, modulated by the global activation threshold; tension and charge.

The rule-based aspects of the architecture are needed to give the observed highly predictable behaviour in usual situations. It is the way in which these two separate aspects are articulated that is the basis of the outline of the SimulACRUM architecture that has been presented, and will remain central to future developments.

One of the great difficulties in research into cognitive architecture is that one cannot definitively prove one architecture to be definitively better than another. What has been presented here needs developing and using in the practical modelling of complex tasks, and only after that has been done, also using other architectures, will it become apparent which architecture is more useful. The considerations given in this paper are indications of potential value.

It is rather like new products on the marketplace, or new programming languages. One has to use a number of them before it becomes apparent which is most useful for what. The value of presenting architectures in development is to stimulate and prepare the ground for future modelling projects, in which models and architecture can be developed together.

ACKNOWLEDGEMENTS

Anne-Laure Amat's comments have been very helpful.

REFERENCES

Anderson, J. R. (1993). *Rules of the Mind.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Bartlett, F. C. (1932). *Remembering.* Cambridge, England: Cambridge University Press.

Cacciabue, P. C., Decortis, F., Drozdowicz, B., Masson, M., & Nordvik, J.-P. (1992). COSIMO: a cognitive simulation model of human decision making and behavior in accident management of complex plants. *IEEE Trans. Systems, Man, and Cybernetics*, 22 (5), 1058-1074.

Grant, S. (1994). Modeling complex cognition: Contextual modularity and transitions. In: *Proceedings of the 4th International Conference on User Modeling*, pp 157-162, Hyannis, MA, USA. The MITRE Corporation.

Hollnagel, E. (1993). *Human Reliability Analysis, Context and Control.* London: Academic Press.

Monnier, A *et al.* (1993). *Rapport relatif à l'accident survenu le 20 janvier 1992 près du Mont Sainte-Odile (Bas-Rhin) à l'AIRBUS A320 immatriculé F-GGED exploité par la Compagnie Air Inter.* Ministère de l'equipement, des transports et du tourisme, France.

Newell, A. (1990). *Unified Theories of Cognition.* Cambridge, MA: Harvard University Press.

Reason, J. (1990). *Human Error.* Cambridge, England: Cambridge University Press.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *Behind Human Error: Cognitive Systems, Computers and Hindsight.* CSERIAC, Wright-Patterson Air Force Base. Ohio.

Cognition in the Workplace: Integrating Experientialism into Activity Theory

Manuel Imaz
Computing Department
Open University,
Milton Keynes,
MK7 6AA, UK
100420.3621@CompuServe.com

David Benyon
Computing Department
Open University,
Milton Keynes,
MK7 6AA, UK
D.R.Benyon@Open.ac.uk

ABSTRACT
When system developers design a computer system (or other information artifact), they must inevitably make judgements as to how to abstract the worksystem and how to represent this abstraction in their designs. In the past such abstractions have been based either on a 'traditional' philosophy of cognition, or have been based on intuitive 'spontaneous' philosophies. A number of recent developments in cognitive psychology have raised questions about the legitimacy of such philosophies. In this paper we discuss where the abstractions which designers employ can come from and how such abstractions are related to the concepts which the users of these systems have.

Keywords
cognitive semantics, activity theory, cognitive ergonomics, system design, experientialism

COGNITIVE ERGONOMICS AND DESIGN
The complex interaction of artifacts and people in their workplace requires that system designers produce a variety of computer-based and non computer-based information artifacts in order to enable people to undertake their work. The process of system development is informed by a number of disciplines. For example, the emphasis of Human-Computer Interaction (HCI) is primarily pragmatic; it is concerned with developing computer-based systems which demonstrate a high degree of usability (Preece, et al., 1994). Software Engineering focuses on developing robust, efficient and maintainable software systems. Cognitive Ergonomics places a greater emphasis on understanding the nature of cognition and on using such knowledge to inform the design of information artefacts.

When system developers design a computer system (or other information artifact), they must inevitably make judgements as to how to abstract the worksystem and how to represent this abstraction in their designs. In the past such abstractions have been based either on a 'traditional' philosophy of cognition, or have been based on intuitive 'spontaneous' philosophies. A number of recent developments in cognitive psychology (Hutchins, 1995; Nardi, 1996c; Lakoff, 1987) have raised questions about the legitimacy of such philosophies. In this paper we discuss where the abstractions which designers employ can come from and how such abstractions are related to the concepts which the users of these systems have.

We look briefly at the 'objectivist' tradition which has dominated software engineering and HCI methods. We also examine the 'contextual' tradition which considers the individual *in situ*; in a social setting working with other people and artifacts in order to accomplish work. We then introduce an approach known as 'experiential cognition' or 'cognitive semantics' which offers a non-objectivist alternative to traditional cognitive psychology by re-introducing the 'subjectiv' into a rational model. We illustrate the potential of such an approach with an example of 'pilot error'; looking at the cognitive demands made by modern aeroplane instruments on the working environment of a flight.

WHAT MODELS REALLY RE-PRESENT
All approaches to systems development employ some form of (conceptual) modelling. It is, therefore, worth taking the trouble to look briefly at how models function. First we need to consider the purpose and use of the model. Second we need to consider whether the model possesses the necessary structure and processing capability to fulfil its purpose and third we need to consider the physical characteristics of the model - its notation and its usability. The analytic, explanatory and communicative power of a conceptual model arises from the structure, operations and constraints which that model is able to capture (Kangassalo, 1983).

For example, consider maps as conceptual models of a terrain. Maps are abstractions from the domain (the terrain) which emphasise some features whilst suppressing others. If one wants to find a path from one village to another then a relief map (one that shows only the height of the land) is unlikely to be of much help. It's purpose is not appropriate for the purpose at hand. If we have a map of the scale 1:50000 this may be more suitable for the purpose of finding a path. The map may have been designed to show fields, fences, rivers and paths and so contains the required functions (the ability to follow a path on the map) and the necessary structure (the concepts of path, etc.) to fulfil its purpose. However, the map may be so poorly designed at the physical level (e.g. it uses black lines to show boundaries and paths and rivers) that it is not usable for its purpose (Benyon, 1995).

This example of maps is extremely important as it is a clear example of bringing *objectivity* into play. When we say that a map is poorly designed, we are saying at the

same time that it would has been possible to design it *correctly*, including the necessary information to obtain all desired purposes. But at this stage it is necessary to take into account what some authors like Denis Wood says about maps: "maps work by serving interests" (Wood, 1992. p. 4).

The choice of things represented on the map is not purely a *technical* problem, but a more general economical, sociological and political one. "Obviously, [maps] do so in all of the choices made, more and less explicitly, about how a particular territory will be represented. A map drawn for the purpose of asserting property boundaries will not generally indicate the presence of specific plants and animals within a given area" (Suchman, 1995).

This example is intended to show how the very notion of objectivity is always associated to a given context, and asserts a certain point of view into this context. The conclusion could be that maps (conceptual models) can not serve all interests at the same time: they show some aspects of reality hiding at the same time others. There is always a choice, a perspective from which we *see* reality, and this perspective is the location from which the observer is perceiving reality.

METHODS FOR DESIGN

Methods for interactive systems design have developed over the years with little thought for why they utilise the constructs which they have nor really how designers should attempt to represent the required work system as a computer-based artifact. This has occurred the same way in nearly all sciences: it is only after a considerable evolution that a discipline would be able of getting some insight or "consciousness" about its own objects and methods it uses.

In the case of computing, it has been necessary to wait some 50 years before starting to think about how software engineers have produced their theoretical artifacts, in particular those used in order to create models of different domains of activities. Software engineers and HCI designers have typically adopted one of three generic approaches: data-centred, object-oriented or task-based.

Object Oriented Methods

The objectivist approach to systems development is mostly clearly seen in object-oriented approaches to system design. Objects are the basic concept in OO approaches. They are defined as 'an encapsulation of attributes and exclusive services [behaviours]; an abstraction of the something in the problem space,...' (Coad and Yourdon, 1992).

Davis (1993), like many OO theorists stresses that objects correspond to real world entities. In OO design analysts/ designers are encouraged to identify objects in the domain and use these as the basis for the conceptual model. Thus the underlying philosophy is that there is an objective world just waiting to be represented. 'The designer looks for classes of objects...In this phase the major tasks are to *discover* the classes required to model the application and to *determine* what behaviour the system is responsible for and assign these specific responsibilities to classes'. (CACM, 1990, our emphasis).

Task-based methods

The objectivist philosophy is also apparent in many task-based methods which have proved popular in HCI. Task-based models seek to represent the domain in terms of tasks. The ontology of task-based approaches includes the basic concept of a user 'task'; 'a goal together with some procedure or ordered set of actions that will achieve that goal' (Amodeus, 1994). Other basic concepts such as 'goal', (a state of the environment or agent which is desired by the agent) 'operation' 'simple task', 'unit task' or 'action' (a task which involves no problem solving or control component), 'plan' or 'procedure' (a sequence of tasks, subtasks and/or actions) are defined in terms of this basic concept.

Task-based approaches date back to the earliest exposition of HCI theory (Card, Moran and Newell, 1980) and continue to be popular today (e.g. Browne, 1994). The emphasis of task-based approaches is primarily on human tasks and the need to understand what people (rather than computers) have to do (Benyon, 1996). Developing task-based models requires the designer to study what users do and to abstract from this. Although the grounding of task-based approaches in the study of 'actual' use introduces some element of context into the description, there is still a strong element of the existence of objective behaviours.

The problem with most classical task-based models like GOMS (Card, *et al.*, 1980) is that do not take into account the social aspects. However, the objectivist position adopted by task-based models has also been shown to be problematic. For example, Draper (1993) found a variety of interpretations of a simple word processing task and notes 'If the notion of task is so unstable and indeterminate in small scale word processing...then what hope is there of using it as a theoretical concept...? (p. 207).

Even much higher level task-based approaches such as cooperative design (Greenbaum and Kyng, 1991) or OSTA (Eason, 1988) rely on the notion of a task. Often these methods are more heavily grounded in ethnographic approaches to understanding. One of the best known of the ethnographic approaches is Situated Action (Suchman, 1987) which arose as a reaction to the domination of 'classical' cognitive science that 'overemphasised the importance of plans in shaping behaviour' (Nardi, 1996a).

The focus of Situated Action is the study of human activity in its emergent, contingent aspects. The unit of analysis is the concept of setting, which is defined as the 'relation between acting persons and the arenas in relation with which they act" (Lave, cited by Nardi, 1996a). The problem with this conceptualisation of human endeavour is that it is difficult to generalise. "Situated action models make it difficult to go beyond the particularities of the immediate situation for purposes of generalisation and comparison. One immerses in the minutiae of a particular situation" (Nardi, 1996).

Data-centred methods

Data-centred models have existed in the world of database design since 1976 when Peter Chen published his paper on the entity-relationship model (Chen, 1976). Shortly afterwards, in 1979 the dataflow diagram was established (e.g. DeMarco, 1979). Typically details of the entities, dataflows and processes are recorded in a data dictionary. Other representations associated with the data-centred approach include state transition diagrams (which come in a variety of guises). These focus on the data which is necessary to move the system from one state to another.

The basic concept of a data model is a data element (or data item). A data item consists of one or more symbols, a name (and usually a more comprehensive description of the meaning of the data item) and a context. The name, description and context ascribe the semantics to the data item. Data items are generalisations of the actual and potential values which that data item can take. These values are known as the domain of the data item (Benyon, 1996).

Another level of abstraction in data-centred approaches is provided by something which is usually known as an *entity*. Although some authors argue that entities are things which exist in the 'real world' (e.g. Chen, 1976), others prefer to focus on an entity as a purer, subjective data-centred concept (Benyon, 1996) and see an entity is an aggregation of data items, expressing the semantics that certain data items belong together. The implications of these different views are discussed in (Hirschheim, Klein and Lyytinen, 1995).

Whether or not relationships between entities and a real world are considered as an 'objetivist' conception (entities are things which exist in the 'real world') or we use a 'formalist' conception (an entity is an aggregation of data items, expressing the semantics that certain data items belong together), the problem of categorising reality remains.

THE ANALYSIS OF CONTEXT

Two related, but different approaches to systems design which take a broader cognitive position are distributed cognition (Hutchins, 1995) and activity theory (e.g. Bødker, 1990). When the emphasis is put on 'a cognitive system composed of individuals and the artifacts they use' (Nardi, 1993. p. 56), we are speaking about distributed cognition, in which knowledge is distributed between people and artifacts. When the emphasis is mainly put on a group of people, working and communicating through artifacts, we are speaking about activity theory.

Activity Theory has its origins in the works of the Russian psychologist Vygostky beginning in the 1920s, but has only recently been recognised in the western scientific community (e.g. Bødker, 1990, Bannon 1991). These approaches take into account the context of the worksystem and concentrate on the placing of users within a worksystem. Distributed cognition offers a more radical development of cognitive psychology which Hutchins refers to as 'cultural cognition' (Hutchins, 1995, p. 354). Distributed cognition seeks to provide a cognitive and social perspective.

Nardi (1996a) sees distributed cognition and activity theory (and to a lesser extent situated actions) to be very close. 'Activity theory seems the richest framework for studies of context in its comprehensiveness and engagement with difficult issues of consciousness, intentionality and history...A creative synthesis of activity theory...leavened by the focus on representations of distributed cognition....would seem a likely path to success' (Nardi, 1996a. p. 96).

Both activity theory and distributed cognition take a holistic, systems view of the world. An activity is ...'a system that has structure, its own internal transitions and transformations, its own development (Leont'ev, 1978,p. 50). Similarly for distributed cognition it is the overall system (which includes many people and many information artifacts) which has a goal; the goal of this activity of flying a plane is the "successful completion of a flight" (Hutchins, 1991). In terms of activity theory this would be expressed as the *object* (more or less equivalent to the system's purpose or goal) of the activity is the successful completion of a flight.

There are many details to these two approaches which we do not have space to discuss here, and interested readers should pursue distributed cognition through the highly interesting and readable Cognition in the Wild (Hutchins, 1995). Rogers and Ellis (1994) and Nardi (1996b) provide a useful comparisons. '...activity theorists argue that consciousness is not a set of disembodied cognitive acts...but rather is located in everyday practice: you are what you do' (Nardi, 1996a, p. 7). In a similar vein, Hutchins warns that believing that systems design is simply a question of taking the human out of the system and replacing the human with a computer is to misunderstand the nature of cognition. In computer implementations it is only *some* of the cognitive processes which are automated.

EXPERIENTIALISM

A rather different and radical alternative to traditional objectivist cognition comes from George Lakoff, Mark Johnson and their colleagues. *Experientialism* (or *cognitive semantics* as it is also known) tries to show that there is no direct relationship between the categories we build and an objective world. It is a complex web of categorisation processes that are grounded in repetitive patterns of action we use in everyday life (particularly images, or 'image-schemas' and metaphors) that allow us to conceptualise a given domain. '...the structure inherent in our experience makes conceptual understanding possible and constrains the range of possible conceptual and rational structures' (Lakoff, 1988, p. 120).

Experientialism is a reaction against the classical cognitive science tradition of the 1960s and 1970s (much as distributed cognition is a different reaction). Lakoff characterises this 'objectivist' psychology as seeing the mind as the algorithmic manipulation of abstract symbols. These symbols somehow get their meaning by being associated with *things* in the world. In contrast, experiential cognition

emphasises the role of the body in characterising concepts and in the human imaginative capacity for creating concepts and modes of rationality that go well beyond any mind-free external reality (Lakoff, 1988). Indeed the title of Lakoff's book *Women, Fire and Dangerous Things* (Lakoff, 1987) is chosen because these seemingly diverse things are classified as belonging to the same category by the Australian aboriginal language, Dyirbal.

Experientialism, or the theory of experiential cognition, employs five main concepts;

- basic level concepts which are formed at a level of abstraction 'at which humans interact with their world most effectively' (Lakoff, 1988, p. 133). These concepts are 'basic' in terms of; *perception* (they are easy to identify and correspond to a single mental image), *function* (at both a motor and cultural level), *communication* (the words for the concepts are typically common, root words in the language which children learn first) and *knowledge* (most attributes of the category members are stored at this level)

- image schemas (or images) which are basic notions of spatial organisation which humans possess. Image schemas form the basis of most metaphors and are again embedded in and structure our direct experience of the world.

- imaginative processes which act on the basic level concepts and image schemas and which allow us to form abstract conceptual models. These processes are: schematisation, metaphor, metonymy (when one member of a class stands in for the whole class) and categorisation.

- basic cognitive processes which act on the basic concepts. These processes, however, are 'high level' compared to traditional cognitive processes and include things such as scanning, focusing, vantage point shifting and so on.

- 'mental spaces' in which such cognitive activities can take place.

In terms of experientialism, rational thought is the application of the imaginative processes and basic cognitive processes to the basic concepts and image schemas. Meaningful structures arise from the structured nature of bodily and social experience and our innate capacity to imaginatively project from the bodily, social or other interactional experiences to abstract conceptual structures. For our present purpose, we will concentrate on the experientialist notions of image-schema and metaphor.

Image schemas are 'abstract patterns in our experience and understanding that...are central to meaning and to the inferences we make' (Johnson, 1987). Image schemas are described as being 'nonfinitary' which means that they are continuous, analogue representations and thus facilitate certain cognitive operations. A collection of image-schemas has been defined (Johnson 1987). The CONTAINER schemas is perhaps the most common, but others include PART-WHOLE (something can be seen as a whole or as its constituent parts), LINK (two or more things have a link between them), SOURCE-PATH-GOAL (or sometimes, just a PATH, which goes from a source along a path to a destination). There is an UP-DOWN schema, a BACK-FRONT schema and so on.

"Take for example, the CONTAINER schema - a schema consisting of a boundary distinguishing an interior from an exterior. The CONTAINER schema defines the most basic distinction between IN and OUT. We understand our own bodies as containers - perhaps the most basic things we do are ingest and excrete, take air into our lungs and breathe it out. But our understanding of our own bodies as containers seems small compared with all the daily experiences we understand in CONTAINER terms" (Lakoff.1988. p.140).

Another main aspect of experimentalism is the central role that is assigned to metaphor. The notion of metaphor here is not simply a linguistic expression. On the contrary, the locus of metaphor is the conceptualisation of a domain. Metaphor is a cross-domain mapping - conceptualising one domain in terms of another - and is central to our thought processes (Lakoff, 1993). As an example of metaphor, we may consider a love relationship as a journey.

Our relationship has hit a dead-end street.

In this example we have defined two different domains; the source domain (travels) and the target domain (love). We can refer to this metaphor in terms of names which gives the following structure: TARGET-DOMAIN IS SOURCE-DOMAIN. In this case we have LOVE IS A JOURNEY. The name is used for the mapping; the set of correspondences that characterise it. For the LOVE IS A JOURNEY metaphor we have:

The lovers correspond to the traveller

The love relationship corresponds to the vehicle

The lovers' common goals correspond to their common destinations on the journey.

When dealing with metaphor it is important to realise that, whilst the names of the mappings take a propositional form, the mappings, themselves are not propositions. Metaphors are sets of conceptual correspondences.

In the next section we present a cockpit example. In the example we show how the concept of metaphor can be useful in helping to understand to a breakdown on a human-computer system.

THE COCKPIT EXAMPLE

Our example is based on a real cockpit scenario. On 27th November 1983 there was a Boeing 747 accident near the Barajas airport. It was an Avianca (Colombian Airlines) regular flight from Paris to Bogota with a stop at Madrid.

Even if the total scenario is quite complex, it could be said that the main cause of the accident was the misuse of the Ground Proximity Warning System (GPWS). According to the Official Report on the Accident, the GPWS was insistently warning about the fact that the plane was below a minimum altitude.

The captain is reported to have reacted saying: "OK, OK" meaning "I know it is a GPWS malfunction, so I will continue the approach procedure". The official reports says:

Captain answers "Bueno, bueno" (OK, OK).

However some people very close to this Official Commission (and some magazines of that time) say that the real answer that the captain gave - registered in the CVR (Cockpit Voice Recorder) - was:

"Calla gringo" (Shut up gringo).

'Gringo' is an expression used by some Latin-American people referring to (north) American people.

In this context the captain was Colombian - (south) American - the aircraft (north) American, and the warning artifact (GPWS) *speaks* English. Additionally, there is a known fault in the artifact that in certain circumstances it gives a false warning and this leads to a mistrust reaction by the captain. We suggest that the captain has used a way of anthropomorphising the artifact by using the metaphor: THE DEVICE IS A STUPID PERSON.

This particular use of the metaphor shows also that the relationship the captain establishes with the artifact is not neutral. It shows some type of mistrust, a pejorative attitude regarding the artifact. The metaphor determines that the relationship with the artifact produces some unconscious reactions as if it were a 'real' person. As this GPWS was an device that in certain circumstances gave an erroneous information, this malfunction determined that in a critical moment the captain did not accept the warning as a real one, but as another malfunction indication.

CONCLUSION: THE ROLE FOR EXPERIENTIALISM IN DESIGN

If cognitive ergonomics is to make a significant contribution to the design of information artefacts, then we must take a broad view of cognition; one that recognises the social construction of knowledge, the distribution of information throughout the worksystem and the ways in which language and thought are employed in real work settings. System designers similarly need to think carefully about the concepts which they use in designing systems; such concepts need to reflect an appropriate view of cognition.

In terms of system design one of the authors (Imaz 1995) has shown that what has been called a *paradigm shift* between a data-centred, process method and OO methods corresponds to a change in the underlying metaphor. In the case of data flow diagrams, the underlying metaphor is PROCESSES ARE MACHINES as any DFD can be considered as a collection of machines in which the raw material is data. Each machine has raw data as input and produces some semi-elaborate data which is the input for other machines.

The underlying OO metaphor is OBJECTS ARE PEOPLE. This can be deduced from the concepts used in the discourse of OO methodologies: behaviour, responsibility, collaboration, acquaintance, visibility, messages, etc. This is mainly a sociological discourse which differentiates OO methods from a machinist discourse used by structured methodologies (Imaz 1995).

We can also see how the LINK image-schema is the preconceptual grounding of relationships used in an ER diagram (there exists a relation - link - between a company and the products and services it sells). The SOURCE-PATH-GOAL image-schema is the basis for other software engineering construct: the state transition diagram. An ATM is an initial state - source - waiting for the introduction of a magnetic card, in which case it moves - path - to another state - goal.

Applying these concepts to systems design techniques is one avenue of our research. Another concerns the choice of an effective and well-grounded view of cognition which can be used by designers.

A stubborn problem within systems development is the representation of context (e.g. Cockton *et al.*, 1996). However, activity theory provides a conceptual grounding for the notion. "Activity theory, then, proposes a very specific notion of context: the activity itself is the context [...] Context is both internal to people -involving specific objects and goals- and, at the same time, external to people, involving artifacts, other people, specific settings (Nardi, 1996b. p. 76).

The problem with this view is how to understand and characterise the ideas of internalisation and externalisation. How can we design artifacts which represent information in a way that will be both understood and dealt with appropriately by different people? As we have seen from the cockpit example, failing to achieve this can be disastrous.

We see internalisation/ externalisation as perhaps the weakest issue in the activity theory and the concepts experientialism provides (image-schemas, categorisation, metaphor, cognitive processes -focusing, scanning, etc.-) as potentially complementing this theory. Image-schemas, for example, are the pre-conceptual artifacts by which we internalise (as concepts) some of our external recurrent everyday pattern of activities. We have already seen some examples of image-schemas CONTAINER, PART-WHOLE and so on. The concepts once produced are applied (externalisation) to our activities.

Experientialism, then, provides the beginnings of a basic ontology which can be applied to interactive system design. Image-schemas and metaphors are the main constructs that cognitive semantics offers us in order to create a cognitive point of view that is compatible with sociological or ethnomethodological approaches. However, whilst experientialism builds its concepts based on a general framework like *experience* other approaches like distributed cognition or activity theory have developed a framework based on the concept of an *activity*. The relationship between these concepts needs further exploration.

BIBLIOGRAPHY

Amodeus (1994) Glossary of terms. At http://www-lgi.imag.fr/Les.Groups/IHM/AMODEUSGlossary/GlossaryEntries/task.html

Bannon, L. (1991) From human factors to human actors in Greenbaum, J. and Kyng, M. (eds.) (1991) *Design at Work: Cooperative Design of Computer Systems*. Lawrence Erlbaum Associates: Hillsdale, NJ

Benyon, D. R. (1996) Domains Models for User Interface

Design. In Benyon, D. R. and Palanque, P. (1996) *Critical Issues in User Interface Systems Engineering.* Springer-Verlag

Bødker, S. (1990) *Through the Interface.* Lawrence Erlbaum; Hillsdale, NJ

Browne, D. (1993) *STUDIO* Prentice Hall

CACM (1990) *Communications of the ACM* 33(9)

Card, S. Moran, T. P. and Newell, A. (1980) *The Psychology of Human-Computer Interaction.* Lawrence Erlbaum Associates, Hillsdale, NJ

Chen, P. P-s (1976) Towards a unified theory of data *ACM Transactions on database systems* vol 1 (1) 9 - 36

Coad, P. and Yourdon, E. (1992) Object-Oriented Analysis. 2nd edition Prentice-Hall

Cockton, G., Clarke, S., Gray, P. and Johnson, C. (1996) Literate Development: Weaving Human Context into Design Specifications. In Benyon, D. R. and Palanque, P. (1996) *Critical Issues in User Interface Systems Engineering.* Springer-Verlag

Davis, A. M. (1993) Software Requirements; Objects, Functions and States. Prentice Hall

DeMarco, T. (1979) *Structured Analysis, Systems Specification.* Yourdon Press; Englewood Cliffs, NJ

Draper, S. (1993) The notion of task in HCI. In Ashlund, S., Mullet, K., Hendersen, A., Hollnagel, E. and White, T. (eds)*Proc of Interchi '93 Adjunct proceedings* ACM press

Eason, K. D. (1988) *Information Technology and Organisational Change.* Taylor and Francis: London

Greenbaum, J. and Kyng, M. (eds.) (1991) *Design at Work: Cooperative Design of Computer Systems.* Lawrence Erlbaum Associates: Hillsdale, NJ

Hirschheim, R. Klein, H. K. and Lyytinen, K. (1995) *Information Systems Development and Data Modelling: Conceptual and Philosophical Foundations.* Cambridge University Press; Cambridge, UK

Hutchins, E. (1991). How a cockpit remembers its speeds. Manuscript. La Jolla. University of California., Department of Cognitive Science. *Cited in* (Nardi, 1996a)

Hutchins, E. (1995) *Cognition in the Wild.* MIT press

Imaz, M. (1995). *Object Oriented Methods: An Epistemological Approach.* MSc dissertation. Open University.

Johnson, M. (1987) *The Body in the Mind. The Bodily Basis of Reason and Imagination.* University of Chicago Press; Chicago.

Kangassalo, H, 1983. Structuring Principles of Conceptual Schemas and Conceptual Models. In Bubenko, J. (ed). *Information Modelling* . Chartwell-Bratt

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What Categories Reveal About the Mind.* Chicago University Press.

Lakoff, G. (1988). Cognitive Semantics. In Eco, U., Santambrogio, M. & Violi, P. (eds.). *Meaning and Mental Representations.* Indiana University Press.

Lakoff, G. (1993). The contemporary theory of metaphor. In Ortony, A (ed.). *Metaphor and Thought* 2nd. Edition. Cambridge University Press.

Leont'ev, A. N. (1978) *Activity, Consciousness and Personality* Prentice Hall; Englewood Cliffs, NJ

Nardi, B. (1993). *A Small Matter of Programming. Perspectives on End User Computing.* The MIT Press.

Nardi, B. (1996a). Studying Context. In Nardi (Ed) *Context and Consciousness. Activity Theory and Human-Computer Interaction.* The MIT Press.

Nardi, B. (1996b). Activity Theory and Human-Computer Interaction. In Nardi (Ed) *Context and Consciousness. Activity Theory and Human-Computer Interaction.* The MIT Press.

Nardi, B. (Ed.). (1996c). Context and Consciousness. Activity Theory and Human-Computer Interaction. The MIT Press.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. and Carey, T. (1994) *Human-Computer Interaction.* Addison-Wesley.

Rogers, Y. and Ellis, J. (1994) Distributed cognition: an alternative framework for analysing and explaining collaborative working. *Journal of Information Technology* 9, 119 - 128

Suchman, L. (1987) *Plans and Situated Actions.* Cambridge University Press.

Suchman, L., 1995. Making Work Visible. in *Communications of the ACM* . Vol.38. No. 9.

Wood, D. (1992) *The Power of Maps.* Routledge; London

Goal-oriented categorisation: the role of task, daily activity, expertise and problem attributes

Willemien Visser & Laurence Perron
Ergonomic Psychology Project
INRIA - Rocquencourt
78153 LE CHESNAY Cedex - France
+ (33) (1) 39 63 52 09 / 59 34
Willemien.Visser@inria.fr / perron@nuri.inria.fr

ABSTRACT
The context of the study presented is the development of a reuse-support tool for telephone-traffic monitoring operators. On the basis of previous studies, goal orientation was supposed to be a crucial factor in use of knowledge in professional activities. A categorisation experiment has been conducted on 17 operators, in order to examine if this variable plays also a role in the interpretation of traffic incidents, and if so, to analyse the form it takes. We conclude, with all proper reservationss, that goal orientation plays indeed a role in operators' interpretation of traffic-monitoring incidents. The variables whose role have been identified are: operators' tasks and daily activities and their expertise acquired through their professional career, i.e. their type and level of expertise. We also observed that, according to their goal orientation, the operators consider differently the characteristics of the incident to be handled. The implications for a reuse-support system will be discussed.

Keywords
categorisation, goal orientation, reuse-support tool, telephone-traffic monitoring, CBR (case-based reasoning)

INTRODUCTION AND RATIONALE

The knowledge representation in knowledge-based support systems should be compatible with the way in which the supposed users of the system construct mental representations of their knowledge —at least, that is how cognitive ergonomists consider this issue.

This paper focuses on a study conducted in order to identify telephone-traffic monitoring operators' organisation of their knowledge on past monitoring incidents. The study was carried out in the framework of a research project aiming at the development of a tool to support these operators in their reuse of such knowledge. The type of system intended is a case-based reasoning (CBR) system, i.e. an A.I. system in which the main knowledge units are "cases", that is —in this situation of a system for traffic operators— representations of specific incidents that have been handled in the past. In CBR, one or more cases are the main knowledge units used in reasoning tasks (problem solving or interpretation): to summarise the different stages in CBR, cases are to be retrieved and adapted (see Kolodner, 1993).

The purpose of this research was therefore to specify traffic operators' organisation of their knowledge on past monitoring incidents, in order for this organisation to constitute the basis for the specifications of the intended CBR support system. Indeed, for the operators to be able to retrieve appropriate cases from the system, the case organisation should be compatible with the operators' mental knowledge organisation, especially, the internal and external organisation of the knowledge elements corresponding to the traffic incidents. The internal organisation of an incident representation refers to the way in which its attributes characterise it. Its external organisation refers to the way in which it is related to the other incident representations. An aspect of this external organisation particularly relevant in the context of retrieval of incidents for reuse is the relationship "usefulness (that is, for handling the incident)": what is it that makes an incident handled in the past useful for another incident, to be handled in the present. We suppose that this depends on an incidents' attributes. In order to identify these attributes, a categorisation experiment has been designed.

The crucial relation between mutual reusable cases is often supposed to be "similarity". This relation can be operationalised in various ways, from strict structural analogy to a resemblance on idiosyncratic, semantic attributes. Kolodner (1989) proposes to proceed by "selection of the best [case]", using "usefulness" rather than similarity. The adoption of the "usefulness" criterion means that one takes into account the reasoner's goals: "it only makes sense to focus on relative goodness of correspondence [based on similarity] after we know that the reasoner's goals are being attended to" (p. 161).

In the psychological research on analogical reasoning, the role played by subjects' goals is one of the hot topics, which opposes e.g. researchers such as Gentner to Holyoak or Keane. Holyoak and colleagues (Thagard, Holyoak, Nelson & Gochfield, 1990) in their ARCS model (analog retrieval by constraint satisfaction) insist on the importance of pragmatic constraints in analog retrieval ("pragmatic centrality"), in addition to structural and semantic constraints. "Stored structures that are potentially important to a system's goals are more likely to be retrieved than irrelevant ones" (p. 267).

CATEGORISATION STUDIES: THE ROLE OF GOAL ORIENTATION

Categorisation studies have indeed been considered to provide data on the way in which people organise their knowledge. Generally, categorisation studies have examined

grouping of entities independently of any task -let alone, tasks related to operators' activities in a work situation. A few studies, however, have taken into account tasks in which the entities to be categorised are processed (Dubois, Bourgine, & Resche-Rigon, 1992/3). These studies tend to show that categorisation may be oriented by a particular task limited in time, but also by a longer professional career (Visser & Falzon, 1992/3). In a study conducted on two expert designers with different professional careers, Visser and Falzon proposed to distinguish the influence exerted by each one of these two variables. The different categorisations constructed by the two designers were explained using the concept of "operative categorisation".

According to Ochanine's theory of the "operative image" (Ochanine, 1978), depending on their task objectives, subjects develop a restricted and distorted model of the object to be handled; this model is more suited to the task than a "true", "objective" representation.

The "operative categorisation" concept was introduced to refer to a rather stable system of categories elaborated throughout an entire professional career. We differentiated the concept from the one of "task-dependent categorisation", also a system of categories elaborated through a goal-oriented activity, but in a particular, more time-limited task.

In the present study, conducted in the context of a traffic-monitoring support system, the knowledge organisation we were after, i.e. the categorisation we wanted to examine, was supposed to be goal oriented. The different traffic operators in the centre we analysed, especially the experts, had a different professional background and —in spite of a intended generalised polyvalence of operators— a different specialism. We wanted to examine if goal orientation plays indeed a role in the interpretation of traffic incidents, and if so, to analyse the form it takes.

TRAFFIC MONITORING: TWO TASKS, REAL-TIME DIAGNOSIS (RT), AND OFF-LINE (OL)

Traffic-monitoring consists of two rather different tasks, real-time diagnosis (RT), and off-line analysis (OL) of incidents. Off-line analysis is conducted *a posteriori*. As noted above, traffic operators in the centre studied are supposed to be not specialised in one of the two tasks, but to be polyvalent. Every operator knows both tasks, but, as they are working longer in the centre, they nevertheless become specialised. The operators' managers are "experts": even if they are specialised, they are also expert in both tasks, compared to the other operators[1], i.e. the specialists and the relative beginners, who are still polyvalent. Both in RT and in OL, experts e.g., often do the incident processing properly speaking, whereas the preparation of this processing has generally been done by the polyvalent subjects.

In RT incident handling, one of the operators' subtasks is to consign the incidents in a computerised incident base, the Incidents' Log (IL). This log is exploited in OL incident handling, for statistical analysis of the incidents. Incidents are consigned in the IL, once they have been handled, but operators also make descriptions of the incidents while they are establishing their diagnosis. An important difference between these manuscript descriptions and the IL files is that schematic representations of an incident — which operators commonly draw— cannot been included in the IL.

Previous studies on the traffic operators had shown already that these two tasks involve rather different viewpoints on the incidents: through analysis both of traffic operators' prescribed tasks and their actual activities, and of traces of these activities (incident classifications in the IL), Perron (1994) showed that RT and OL operators have different approaches to the incidents. She noticed, e.g., that, for RT operators, classification of an incident according to its cause —as required by the IL, due to its organisation in classes on this basis— was not compatible with their task —whereas it corresponds to one of the goals of the OL operators.

Two types of considerations, thus, led us to design a categorisation experiment, focusing on the variable of goal orientation: the hypothesis of goal-oriented categorisation formulated on the basis of earlier studies, and the previous observations on the traffic operators.

The support system should *a priori* be useful for both types of task, RT and OL. The data we were after should be relevant in a "reuse" situation, that is in a situation in which the past-incident case to be retrieved be useful to the processing of the current incident. Thus, the task-dependent aspect of goal orientation has been manipulated in the following way:

- all subjects were given the experimental task to put together incidents that could be useful for each other's processing;
- for about half of the subjects this task was to be executed in a RT situation, for the other half in an OL situation.

Operators were put in the RT or OL situation corresponding to their prevalent daily task (this leads to confound experimental task, and daily task). They were given the documents normally used during their activity. This leads to giving to the experimental task a rather realistic character.

In addition to this experimental variable supposed to introduce different goal orientations through different tasks (incident grouping in a RT or in a OL situation), other possibly relevant variables were related to the operators' professional career (cf. the "operative" aspect):

- the operators' level of expertise (three levels: experts, specialists and polyvalent subjects);
- mainly for the experts (to a lesser degree, for the specialists), the operators' type of expertise (variations with respect to the operators' previous and present experience in traffic monitoring).

We had information about the operators' type of expertise through the previously conducted analysis, and also on the

---

1. OTES

"Operators" is the general term used in this text to refer to all subjects, polyvalent, specialist, or expert.

basis of a questionnaire about their professional career and present experience which operators were asked to complete at the end of the experiment.

Examples of previous experience and/or present specialist tasks are planning of actions in emergency situations, or in case of special events (e.g. T.V. shows, Mother's day), or writing a guide for diagnosis.

A CATEGORISATION EXPERIMENT

Method

Subjects

Participants were 17 operators: 10 in the RT, seven in the OL condition, according to their prevalent daily task (this leads to confound experimental task, and daily task). Table 1 presents the distribution of subjects over the level-of-expertise conditions.

	RT	OL
experts	2	2
specialists	2	3
polyvalent subjects	3	5

Table 1. Number of subjects in each level-of-expertise condition

Because the availability of the operators for participation in the experiment depended on variables which were unforeseeable —also by the operators themselves— (mainly the occurrence of incidents), we were not able to obtain two groups of equal size: there were 7 subjects in the RT situation, and 10 in the OL situation. Because of the importance of qualitative analysis, we decided to not exclude the data of three OL operators.

Material

There were 31 incident descriptions (see the skeleton in Figure 1). These were incidents which had been handled, two or three weeks ago.

Incident n° xxx - area
ALARM: tool - date / hour / duration
code SGQS - alarm label
direction of flow - values of indicators
ACTIONS:
hour - action xxx
action yyy
.....
CAUSE: .....
IMPACT on TRAFFIC: .....
COMMENT: .....
(schema)

Figure 1. Skeleton of incident description used in the experiment

Key. "tool": tool used in order to detect the incident
"code SGQS": circuit-group code

N.B. This does not imply that the traffic operators, or even several traffic operators, had to group "their own" incidents:

- the experiment was conducted by day; among the incidents collected, several had been handled, at least in part, by night (different operators are working under the conditions of day and night);
- many incidents are being handled by several operators, consecutively or in parallel;
- incidents may be consigned by another operator than the one / those who handled it.

In order to collect the incidents, we chose randomly four days during which we accumulated all incidents consigned in the IL. With respect to their attributes, incidents were presented with all attributes used by the operators in their manuscript and their IL description (so, schematic representations made during incident handling, but which could not be recorded in the IL, were also included).

For the experimental conditions to be as realistic as possible, the traffic operators were given, in addition to the incident descriptions, a certain number of documents — different in the RT and OL conditions— which they normally use for their task, thus reproducing the operators' daily work situation as much as possible.

Procedure

Subjects were requested to group incidents according to their mutual utility for their handling in a RT or an OL situation (an example was presented). After subjects had constituted as many groups as they wanted to, they were asked to make explicit the reasons, both for putting together incidents, and for distinguishing groupings of incidents. There were no other constraints. Two or three possible false assumptions were made eliminated. They could take as much time as they wanted. They were told explicitly that they could put a same incident in different groupings. They could modify their groupings as often as they wanted. When they were satisfied —inasmuch as possible— with their groupings, these were considered their categorisation: they were not asked to proceed to other groupings.

Data analysis

The experimental data were analysed using a statistical method, which calculates similarities between items, on the basis of —all items token two by two— their having been put together, and their not having been put together (Barthélémy & Luong, 1986). Data presented in this paper concern categories of incidents, i.e. incidents frequently put together by a reference group of subjects, i.e. especially

- by all subjects;
- by the RT subjects;
- by the OL subjects.

N.B. The data concerning categories of subjects (i.e. subjects with close categorisation behaviours) were not clear, and will not be presented in this paper.

RESULTS

The results —i.e. concerning categories of incidents constructed by the subjects— can be summarised as follows:

categories by RT subjects ≠ categories by OL subjects. Analysis of the incident categories, guided by

- their attributes, and
- subjects' justifications of their groupings[2],

shows that all variables that we distinguished, play a role in the composition of the incident categories:

- subjects' experimental task & daily activity;
- subjects' level of expertise;
- subjects' type of expertise;
- incident attributes.

The influence of the subjects' experimental task and that of their daily activity cannot been distinguished: subjects were put into the context of their prevailing daily task.

Role of subjects' experimental task & daily activity
The main grouping criterion used by the subjects is an incident's alarm, followed by its cause combined with other incident characteristics (e.g., type of object concerned by the incident: circuit group or switch; actions executed; temporal characteristics: e.g., date and hour of occurrence, duration of the incident, fluctuating or stable character). This results holds for all subjects, for RT subjects and for OL subjects. RT subjects refer, however, to different alarms than OL subjects:

- the alarms privileged by the RT subjects are "inefficacy" and "saturation / degradation";
- OL subjects make finer distinctions, referring to "low efficacy", "saturation / degradation of 1st choice circuit groups" and "saturation / degradation of last choice circuit groups" (" zero efficacy" has been combined with other characteristics leading to several smaller incident groupings).

This difference can be explained by reference to the subjects' daily activities. RT diagnosis is guided by incident processing using procedures: "inefficacy" and "saturation / degradation" refer to two chief groups of procedures proposed for alarm handling. The three groups constituted by the OL subjects refer to the main organisation of the documents used in order to analyse incidents in OL.

Role of subjects' level of expertise
This factor intervenes in various ways.

Different categorisation behaviours
The "finest" categories are constructed by the experts, relative to those made by the other subjects. Experts construct between 7 and 28 categories (m=18,5); specialists between 6 and 17 (m=10,4); and polyvalent subjects between 3 and 14 categories (m=8,3).

Different categories
Several categories constructed by experts are different from those made by the other subjects, specialists and polyvalent subjects. They do not put together incidents which have been put together by all other subjects (incidents n° 1 and n° 11).

Different approaches to the incidents
Experts have a different interpretation of the experimental task than the specialists and, especially, polyvalent subjects: experts' categories are more guided by incident processing properly speaking —compared to the preparation of this processing, executed by, especially, the polyvalent subjects. This translates their respective daily activities, the proper incident processing versus its preparation.

Concerning many individual incidents also, experts make a different analysis than their colleagues: e.g. incident n° 24 has been categorised by a RT expert according to its alarm, whereas several polyvalent OL subjects categorised it according to its cause. The RT expert referred to a particular tool used for its diagnosis; the polyvalent OL subjects favoured its analysis. Experts are also the only subjects who "dare" to ascribe another-than-the-declared cause to an incident (i.e. as declared on the incident presentation).

Experts may make a "completer" analysis of the incidents: they are e.g. the only subjects who ascribe a cause to an incident whose cause has not been declared (on the incident presentation). They may take into account more elements when they examine the situation in which the incident occurred (e.g. available means, organisation of labour).

Role of subjects' type of expertise
This factor has been identified particularly clearly for the experts, but it plays also a role in the specialists' categorisation. The four experts differ from each other with respect to this variable. Most differences can be traced back to the experts' professional career. Particular examples will not be presented because of the technical nature of the knowledge involved.

Role of incident attributes
All subjects use the same types of attribute as first and second criterion to constitute their groupings on, i.e. alarm, and cause combined with other incident characteristics.

The first and second groupings of particular incidents most frequently put together by the RT subjects are also the same as the first two groupings made by the OL subjects (but not by the experts).

In order to justify a grouping, operators often referred to an incident's prognosis (evolution of the state of the object involved in the incident, and of the global situation in which the incident occurs; periodicity and/or permanence of the incident; possible actions; predictability of events). Note that these attributes are not mentioned by the operators in their manuscript description, and are not foreseen in an incident's description in the IL.

Discussion

The present paper has examined if goal orientation plays a role in operators' interpretation of traffic-monitoring incidents, and if so, to analyse the form it takes. In a previous categorisation study (Visser & Falzon, 1992/3) we distin-

2. Justifications of distinctions between groupings have only been analysed for the experts.

guished two variables intervening in goal orientation: the operators' entire professional career (cf. "operative categorisation"), and particular, more time-limited tasks (cf. "task-dependent categorisation").

The conclusions of the present study are still rather provisional. First, part of the data —those concerning categories of subjects, not presented in this paper— was not clear. Second, most of our conclusions are based on qualitative analysis guided by the knowledge we acquired through previous analysis of the operators' tasks and activities.

With all proper reserves, we then formulate the following provisional conclusion: goal orientation plays a role in operators' interpretation of traffic-monitoring incidents. The following intervening variables have been identified:

- an operator's task and daily activity (the task-dependent aspect);
- an operator's expertise acquired through their career (the operative aspect).

These two aspects are indeed those proposed in the Visser & Falzon (1992/3) study.

The role of an operators' entire professional career has been decomposed into the role of their level of expertise and their type of expertise —even if the role of the type of expertise was difficult to evaluate precisely. In the Visser & Falzon (1992/3) study, the differences between the experts were attributed to their different types of expertise; both experts had the same level of expertise (inasmuch as this can be determined): their colleagues considered them both as "the" experts of the department.

The incidents' attributes played a role in as far as operators with different goals processed them differently: they paid attention to different attributes, and used a same attribute differently.

The results of the present experiment specify those obtained through the previously conducted analysis of traffic operators' tasks and activities (Perron, 1994). These analysis led to attribute to the RT and the OL operators different viewpoints on the incidents.

With respect to implications for the planned reuse-support system, the translation of the results does not seem immediate. Attributes and their role in different tasks have been identified. Certain results are clear. Alarms, e.g. seem to be the first selection criterion for all operators, in all tasks. The reference to causes is, however, less evident.

For example, different operators may ascribe different causes to particular incidents —a result which strengthens observations made during the previous analysis of the operators' traffic-monitoring activity. Sometimes an operator —moreover an expert— categorises an incident in terms of a cause non mentioned in its description —or goes against its declared cause.

The identification of an incident's cause is one of the most difficult activities in traffic monitoring[3].

All incident characteristics do not have the same status in the incident description as it is presently proposed to the operators (in the computerised IL):

- alarms are always mentioned;
- causes are generally inscribed, but not always;
- the same observation holds for most other characteristics which the operator, in order to categorise an incident, uses in combination with the cause (type of object, actions, certain temporal characteristics, such as date and hour of occurrence, duration);
- as noted already, attributes related to an incident's prognosis are, however, not inscribed in its description.

Nevertheless, these unrecorded characteristics related to an incident's prognosis are important for its handling: e.g. they are referred to in order to justify incident groupings. Therefore, one may suppose that they are also important for handling a new, similar incident, which one plans to handle on the basis of an incident handled in the past (i.e. referring to reuse of the incident).

For these reasons, we judge that the way in which traffic operators characterise (describe) incidents requires further investigation. This research could progressively use and develop a tool (language, formalism) providing support for the characterisation of incidents, i.e. for their analysis and diagnosis —and so, indirectly, for the description and organisation necessary for their storage in the case base of a possible CBR system.

ACKNOWLEDGEMENTS

This research was supported by a grant from CNET N° 94 5B 008.

The author thanks Françoise Détienne for her critical remarks concerning a previous version of this paper.

REFERENCES

Barthélémy, J. P., & Luong, N. X. (1986). Représentations arborées de mesures de dissimilarité. *Statistiques et Analyses de Données, 11* (1), 20-41.

Dubois, D., Bourgine, R., & Resche-Rigon, P. (1992/3). Connaissances et expertises de divers acteurs économiques dans la catégorisation d'un objet perceptif. *INTELLECTICA, n° 15*, 241-271.

Kolodner, J. (1989). Selecting the best case for a case-based reasoner. *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society.* University of Michigan, Ann Arbor, Michigan, August 16-19, 1989. Hillsdale: LEA.

Kolodner, J. L. (1993). *Case-based reasoning.* San Mateo, Calif.: Morgan Kaufmann.

Ochanine, D. (1978). Le rôle des images opératives dans la régulation des activités de travail. *Psychologie et Education, 3*, 63-79.

Perron, L. (1994). *La supervision de réseaux de télécommunications: analyse de l'activité centrée sur la réuti-*

3. We don't discuss in this paper the supplementary problem that what constitutes one incident for an operator can be built up by several component events —with consequences becoming causes, for example.

lisation et la catégorisation d'incidents (Rapport interne). Rocquencourt: INRIA.

Thagard, P. R., Holyoak, K. J., Nelson, G., & Gochfield, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence, 46*, 259-310.

Visser, W., & Falzon, P. (1992/3). Catégorisation et types d'expertise. Une étude empirique dans le domaine de la conception industrielle. *INTELLECTICA, n° 15*, 27-53.

Modeling Complex Work Systems - Method meets Reality

Gerrit C. van der Veer
Dep. of Computer Science
Vrije Universiteit, Amsterdam, NL
gerrit@cs.vu.nl

Machteld Hoeve
Dep. of Computer Science
Vrije Universiteit, Amsterdam, NL
mhoeve@cs.vu.nl

Bert F. Lenting
CTIT
Twente University, Enschede, NL
lenting@ctit.utwente.nl

ABSTRACT
Modeling an existing task situation is often a first phase in the (re)design of information systems. For complex systems design, this model should consider both the people and the organization involved, the work, and situational aspects. Groupware Task Analysis (GTA) as part of a method for the design of complex systems, has been applied in a situation of redesign of a Dutch public administration system. The most feasible method to collect information in this case was ethnography, the resulting task model needed the GTA formalism to be adjusted to situational aspects of the work. The study shows that GTA as an approach is feasible in complex design cases, and that the formalism allows adjustment to cover situational aspects, while still keeping its cognitive ergonomic value for design.

Keywords
ethnography, groupware, GTA, task analysis

INTRODUCTION
Various design approaches currently exist in Cognitive Ergonomics. The common characteristic is attention to users and user relevant aspects of the system to be developed. The implementation of this attention, however, is very divers. In this introductory section we will briefly mention some viewpoints. In current user interface design practice, different viewpoints and approaches are often combined or integrated.

First of all, some approaches are based on user participation (Mumford, 1987). Attention to the user and his task situation is not based on systematic analysis and representation of task knowledge, but on the availability of the users during the design process. These methods allow users, who are not professional designers, to influ-ence design decisions in a very direct and explicit way.

A different viewpoint can be found in design methods that stress the importance of systematic knowledge of existing task situations. Many approaches are based on cognitive psychological theory. TKS (Johnson, 1989), the Hierarchical Planning method (Sebillotte, 1988), and MAD (Scapin & Pierret-Golbreich 1989) present a method where design starts mainly with investigating current task knowledge. Methods of this type apply a conceptual framework that is derived from psychological notions on human memory and human planning. The prospective system users provide the knowledge that is subsequently represented in task models that feature as the main source for subsequent design activities. The design requirements and decisions are strongly based on these task models, where the (professional) users only indirectly contribute to this.

Again, a different perspective may be found in the mainstream of CSCW design methods. Here, again, the existing task situation is the primary source for design. However, not the psychological basis of expert knowledge is considered, but the task-knowledge that can be experienced in current work situations and in "communities of practice". Activity Theory (Nardi, 1995) and ethnographic approaches (Jordan, 1996) are the main theoretical foundations for methods that aim at acquiring insight in current situated task domains. Design is based on this insight. Different from the previous type of approaches, however, the analysis of the current task situation is not guided by an a priory conceptual framework, nor is it followed by a formal representation of the task structure.

Many approaches (though mainly the knowledge oriented views) apply representation-techniques, conceptual frameworks or models to accommodate user-oriented viewpoints: CLG (Moran, 1981), GOMS (Card, Moran & Newell, 1983), CCT (Kieras & Polson, 1985), TAG (Payne & Green, 1986) and ETAG (Tauber, 1990) each provide methods to explicitly consider user related characteristics in design, mostly based on cognitive psychological theories. In these types of approaches, it is the designer who evaluates human aspects as part of design specifications, systematically considered with the help of the model or framework, in order to optimize the design. The framework provides a model of cognitive aspects of the user, in the form of a competence model or a performance model (de Haan, van der Veer & van Vliet, 1991).

The approach that we will present in this contribution is an attempt to integrate several viewpoints: we start our design by investigating knowledge of the current task situation. We apply different methods for this, related to the complexity of the human-machine system to be designed - both knowledge elicitation and ethnographic methods may be needed depending on the current task situation. We apply a conceptual framework that is based on a priory notions of human cognition and on our vision on situated tasks, as well as on our vision of the further design of complex systems. Finally, we apply a formal modeling approach. In this way we interpret and model the task world, as far as it is relevant for the designer, NOT as a complete account of the world as it exists for the people living in it.

Part of our approach will be illustrated in this paper. Our complete view on design considers several phases of design decisions, guided by explicit design rationale, which will not be discussed in detail (Van der Veer, van Vliet and Lenting, 1995). In the next section we elaborate our task analysis method GTA, which may be considered the first phase of design in this view (other phases regard technology specification, early evaluation, and implementation - which will in practice be elaborated in an iterative structure of design activities). The remaining sections will provide a brief account of an actual design situation, regarding the reorganization of the Dutch social security system. We will illustrate our approach in this case and provide an account of what happens when applying GTA in a real life situation. This lead us to reconsider the conceptual formalism of GTA, though the conceptual framework as such is still feasible.

GTA - MODELING COMPLEXITY

Groupware task analysis (van der Veer, Lenting, and Bergevoet, 1996) is an eclectic method. Analysis of the task situation may consist of (a) collecting information on current task situations; (b) modeling the current task situation including conflicts, problems, and needs of prospective users; (c) negotiation with the client of design on requirements; (d) investigating physical and technological possibilities and constraints; and (e) specification of the prospective task situation, based on b, c, and d, and often including the redesign of organizational and procedural aspects of the task situation. At this stage there will often be a need for early evaluation, even though no actual artifact has been specified in any detail.

It should be noticed that not all types of methods will always be feasible. E.g., sometimes, the actual design situation may prohibit activity a (the users may be unknown, the task situation may be inaccessible because of military strategic rules), or the current task situation may already be completely modeled, and the design may be restricted to system details that are supposed or required not to influence the task situation at all, in which case the whole set of task analysis activities will be superfluous.

We will provide a detailed account of some methods from categories a and b. For an account of the whole structure of methods in our view on task analysis, see van der Veer, Lenting & Bergevoet (1996).

The intended end-product of GTA is a formal task model that specifies the intended task situation, intended to be the input for the specification of technology details. This model (generally referred to as "Task Model 2") is build only after the current task situation is sufficiently investigated. The current task situation, modeled in "Task Model 1", may be based on knowledge acquisition methods like the ones proposed by Sebilotte (1988) or Johnson (1989). These methods apply mainly as far as relevant task knowledge is explicitly available from "professionals". In many cases this is partly true. Professionals, however, sometimes apply implicit knowledge (especially in cases where their task activities are triggered by situational aspects like the presence and state of certain task relevant objects - e.g., driving in a well known town).

Another aspect of task knowledge seems to be available mainly in the form of "group-knowledge" (Jordan, 1996), where anecdotes sometimes turn out to be vehicles of relevant strategies, or where teams are able to provide information for performance only by combining role specific elements from different perspectives. Group knowledge may also be only implicitly available, in which case only trained ethnographers may detect relevant aspects of task behavior, about which none of the players in the situation seems to be aware.

In complex task domains, many different methods may be needed to collect all relevant information, from knowledge acquisition through interviews and psychological elicitation techniques, via behavior observations in situation, via study of documents, to ethnographic methods. GTA proposes to combine these methods, listing them mainly in two categories: collecting informa-tion that is of a cognitive nature (explicit knowledge from professionals and from documents), and informa-tion that stems from (mostly participatory) observation and registration in situe, which requires hermeneutic interpretation, like ethnographic interaction analysis.

In our practical experience with the GTA method we found these sources to be highly complementary. Information collected with the various methods turned not to be contradictory most of the time, or the contra-dictions could readily be interpreted (documents loose their validity over time, knowledge is role-specific and different roles may relate to conflicting goals). Conflict in this sense should not be removed but should be represented in the task model, as an important aspect of the current task situation.

The core of GTA is a conceptual framework that serves two purposes: to be some guideline for the original collection of information on the current task situation, and to be the base for modeling both the old and the new task situation. Although for the knowledge elicitation type of data collection a framework is quit common, for the ethnographic approach this is not obvious. The framework we apply is based on analysis of ethnographic methods in many design cases, and has up to now shown to be applicable, though we certainly have no theoretical proof for its validity.

The GTA framework consists of three viewpoints: people, work, and situation. Mainly the second viewpoint is quit common in the knowledge elicitation type of task analysis approaches, and the first is sometimes present as well (Johnson, 1989). All three seem to be valid in the ethnographic examples we have been able to analyze.

People

This viewpoint focuses on human actors, both as individuals and in groups. In GTA we consider the following concepts: (a) actors as individual persons. Important is mainly to collect information about relevant characteristics and about variability of these, e.g. typing skills or education; (b) roles as a categorization of people that perform certain sets of tasks; and (c) organization of people, revealing how and in what time structure roles are allocated to actors.

Work
Work refers to structures of tasks and activities. At least we need to collect information on (a) tasks, indicating meaningful packages of work that intend to fulfill a certain goal; (b) task structures; (c) actions, defined as simple activities that do not have an individual goal but derive their meaning from the task as performed in a certain situation; (d) protocols and (e) strategies, which indicate structures of tasks that are commonly performed in certain situations, in general resp. in the case of experts.

Situation
Analyzing a task world from the viewpoint of the situation means detecting and describing (a) the environment as a structure of physical, conceptual, and social relations as well as the relevant history of these relations; (b) describing the relevant things (we hesitate to coin these "objects" which might cause confusion with our object oriented formalism for other concepts in GTA), where things may be physical as well as conceptual, like passwords or messages; and (c) object structures, referring to semantic relations between objects, which may indicate type relations ("is a") as well as relations of other semantic character ("contains", "part of").

In performing data collection in GTA we try to apply all three viewpoints and collect information of all concepts from these viewpoints as well as the relation each has to other concepts (e.g. roles are described in terms of tasks that belong to the role, in terms of history of the situation in relation to the allocation of the role, as well as in terms of characteristics of the individuals that may take the role, and in terms of the rights that the role allows towards relevant objects).

In representing task knowledge in a task model, GTA proposes three different types of representation: object structures, relational graphs, and structural graphs.

Object structures
For some concepts in GTA it may be useful to define an object class, where the instances of that type are characterized by relations to objects of the same and other types. Thus far we regularly use object classes for roles, for tasks and for things, but we experimented with other classes as well (actors). Each object class is characterized by the above mentioned relevant relations.

Figure 1 presents an example of a thing object for a calendar situation.

Relational graphs
Relations between objects of the same class are represented in the object structure, e.g., see the subordinate and superordinate slots and the themes and places slots in figure 1. For purposes of design decisions, however, a graphic representation will provide a more easy overview. We will benefit from a tree representation of type-hierarchy, as well as from a graphic representation of the themes-places relation where themes are objects that may find place in or on the current object, and the places are other objects that have the current object as a theme. In the same way relational graphs may show the semantic relation between roles and between tasks.

Figure 1. Object structure of thing object.

Object: timeslot	
superordinate: memo	**subordinate:** day_slot hour_slot quarter_hour_slot
themes: meeting holiday business travel	
places: month calendar week calendar day calendar	
relevant tasks: cancel meeting initiate meeting postpone meeting receive meeting cancellation forward meeting cancellation	
actors/roles; competence meeting participant; initiate cancellation meeting initiator; prohibit cancellation	
passive/active: passive	
attributes: time, date	

Structural graphs
A last graphical representation that we found beneficial for representation of the task structure refers to relations between objects of different classes. E.g. we regularly will need to represent the relation between task structures and roles over time for which a certain kind of workflow diagram may be suitable. Likewise the relation between task structures and things is often needed to be represented.

THE SOCIAL SECURITY CASE

GTA has been applied mainly to design exercises that were explored in the course of educational activities. Though the task analyses considered actual work in real life situations, the scope of the designs were tuned to the amount of time available in the design classes (ranging from 1000 to 3000 person-hours). The current example stems from a much larger design enterprise, where many person-years were spend on the specification phase alone. Our report considers the first phase of task analysis only, from initial inventarisation and collection of information to the formal representation of the resulting task model of the current situation, task model 1.

The Dutch social security system provides a large number of support facilities for citizens in divers situations of financial and other needs. In part of these cases the main reason for applying for support is a situation of very low or no income. Reasons may vary from loss of job to situations of divorce or loss of family member, forced change of housing situation, to other types of personal disasters.

A special type of tasks that is frequently performed by the social security systems in Dutch municipalities is the settlement of requests for support of primary subsistence. This task requires several levels of information collection and subsequent decisions.

The Dutch social security system as a whole is currently in a transition from very traditional office organization supported mainly by some database facilities and much stand

alone simple personal computing facilities to a work structure that should be much more efficient, well structured, and requiring less overhead. Integrated implementation of workflow management tools and other forms of cooperation technology are being considered. Some prototyping systems are being constructed and installed in certain offices. Some divisions of the Amsterdam branche have been appointed as test bed for certain aspects of the new development. In this situation, one of the authors was contracted to perform a task analysis with the aim to describe the current situation, as a start for a systematic design effort. A period of 6 months full time could be devoted to this goal.

Colleagues had already studied other parts of the situation, and an overall sketch of the office structure and a collection of characteristics of the actors in relation to the identified roles was available.

A METHOD MEETS REALITY

At the start of our task analysis of the handling of requests for primary support of subsistence in the social security case we profited from the fact that a group of our colleagues already had been working in the same offices. A detailed analysis was available of the characteristics of actors and regular roles, and, on the other hand, the office workers were aware of the status and possible obtrusive presence of ergonomic analysts.

Our main data collection started with some interviews of the type proposed by Sebilotte (1988). One of the first parts of our task model 1 was a structural graph of the process, of which figure 2 shows only the general structure.

During the phase of knowledge elicitation, however, we found out that an enormous variety of strategies existed and that workers evidently were unable to provide a generic account of the details of their work.

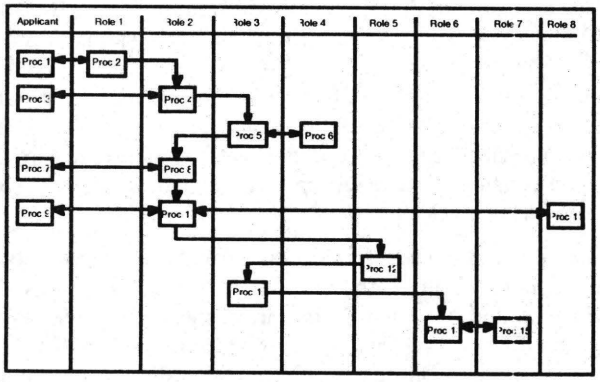


Figure 2. Process structure of work flow.

Ethnography: Hanging Around

We decided to embark on an ethnographic study. One of the authors spend 2 months in the office, "hanging around" in order to get a feeling of what was going on in the complex process, and, on the other hand, to make her role clear to the people. All employees knew that the whole social security system was being redesigned, and that the researchers were participating in this process as part of an outside consulting project. During these first months the employees learned to trust the analyst in relation of what would not be communicated to (other) staff members, and the analyst learned to identify the "hot spots", instances of events and interactions that seemed relevant for systematic analysis.

The analyst also learned to present herself to the applicants for financial support, which were the clients of the service. These people are often trying different ways to "earn" as much support as possible from the system by presenting their case in a way they consider most promising for this. The analyst learned how to communicate with these applicants, in order to establish a relation of trust in this case as well, meaning that she will not be, and will not be considered, a possible source of information or influence in any direction.

Participant observation

This development of a basis of trust turns out to be very important in order to be allowed witness to important aspects of interactions. It needs to be established by a strong commitment of not involvement with the process. On the other hand, this commitment could not always been kept: e.g., the analyst was present at a certain moment when a document was placed in the pigeon hole of a worker who was not expected to be present for the next 4 weeks. She was convinced that this would seriously affect the applicant's expected financial conditions for the next period, so she decided to interfere with the event (and make a record of this action). In the light of the commitment to all actors involved, not reacting to this type of observation would not have been appropriate. However, the analyst performed no action that violated any commitment on which the trust was based, i.e., that was perceived not to be in the interest of any actor that had allowed the analyst to participate in private and personal aspects of the situation.

Interaction Analysis

After 2 months the analyst felt she had acquired enough insight in the situation to be able to identify "hot spots". In addition, the analyst increased insight in the character of the roles in this organization, both by collecting a more complete overview of the task structures related to the roles that were already identified, and by establishing the notion of some additional roles in certain situation, where a second person was needed to conform important decisions. This second person normally performed identical roles, but was in this case chosen because there was no relation to the applicant concerned. Hence, in the actual situation of a certain request this person acted in a different role.

In order to systematically collect information, we choose to apply interaction analysis (Jordan, 1996), which consists of recording instances of hot spots on video, transcribing these into content logs, interpreting them, and combining interpretations of equivalent hot spots into general descriptions. Jordan proposes different possible views for orientation. We decided, based on an analysis of the situation (including relevance of locations, differences between actors, complexity of the process, etc.) to collect both setting-oriented records for those settings, where the

internal structure seemed to be situation and actor dependent, and process-oriented records for the total complex process.

For setting oriented records, the analyst collaborated with several employees in identical roles, who were willing to cooperate, and who differed on relevant aspects of their strategies in the setting concerned or on their views on the roles involved. For the process-oriented records, we asked collaboration of actors who were participating in the very first parts of the complicated process, in order to acquire approval from a set of applicants to videotape and analyze the complete process of their case. At the start it was, of course, unpredictable in which direction cases would develop. We had to discard certain parts of cases after it turned out the continuation developed into a different process in the service. Each relevant record was transcribed into a content log.

After collecting a sufficient number of records for the relevant subprocesses and complete process records, the tapes and the content logs were interpreted. In this activity we asked the participation of some of our colleagues, who were at that time also working in the same offices in the course of different design cases, as part of the same external consultancy project. The records were interpreted in terms of the conceptual framework.

GTA FORMALISM REVISED

Although the interpretation of the records in terms of the GTA conceptual framework was relatively straight-forward, serious problems arose when trying to construct models from different accounts of the same event, both when combining observations from different actors in the same role, and from combining observations from a single actor for different occurrences of the event. This is in contrast with the situation that is normally reported from task modeling approaches of the knowledge type (MAD, GTA), where professionals are found to be consistent and in agreement between each other relating the task structure. The behavior in the case of our complex system turned out to strongly influenced by situational factors as well as by attributes of the partners in the interaction.

The first addition we had to make to the GTA formalism concerned the way to represent the structure of a certain task. In the past, we used constructors in the sense of Scapin and Pierret-Golbreich (1989), where notations like LOOP, PAR, OPT, indicated that certain subtasks are performed in a loop, in parallel, or optional, though we normally adjusted the actual set to the actors' perception of the task domain. Constructors are used both in the representation of task object structures, and in representations of relational graphs that show task decompositions. In the current case, however, we needed several much more elaborate constructors, especially in cases where an employees was interacting with the applicant for support:

CHOICE (<LIST>) and OPT (<LIST>) ,

where LIST can be any ordered subset of A, Cb, Cs, S. CHOICE indicates that one of the subtasks is chosen based on a condition that is a function of the elements of the list, evaluated in the order stated, OPT indicates that the following task is optional, based on the evaluation of the ordered list. See figure 3, for part of a task relational graph.

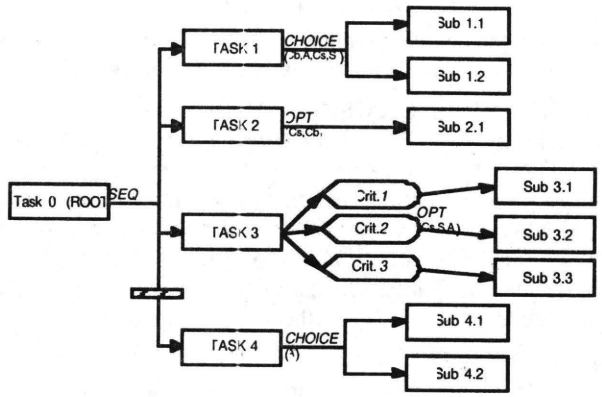


Figure 3. Relational task graph with new constructors, interrupted sequence and multiple criteria.

In this process the elements of the list were: preference/ strategy of the individual actor (A), behavior of the client (Cb), situation of the client (Cs), other situational aspects (S).

Moreover, in the formal representation we still found the need to frequently add notes to certain constructors where a single verbal label was insufficient to make the formal representation interpretable for members of the design team.

Another addition we needed to define for the task relational graphs indicated the (frequently occurring) interruption of a sequence of related tasks by periods of events that were unrelated to the current process.

A last addition to the task graph concerns the frequent situation where a task is decomposed into a set of subtasks, each of which is only performed based on a criterion that is valid for that subtask only. In order to provide a readable and valid representation, we had to explicitly introduce this set of combined criteria and subtasks as part of the task hierarchy.

CONCLUSIONS

The GTA framework showed to be valid even for the highly complex and highly situated type of process that analyzed. The methods of data collection in this case were mainly of ethnographic nature, where a long informal introduction and subsequent interaction analysis turned out to be feasible. The GTA formalism needed to be adjusted to allow a representation of situated and person-dependent substructures of tasks and events.

The task analysis as a whole was in fact checked for validity by the subsequent phases of design. Based on the model of the current task situation (task model 1) an analysis of the detected conflicts and problems, a new task structure of the process was specified. This specification was formalized with the same extended GTA formalism, which allowed the development of scenarios for the process and especially for the subprocesses that were proposed to be redesigned. The scenarios were implemented by developing video fragments via role-playing. Both the

original task model 1 and the scenarios that represented the proposed re-design were judged by the workers as realistic descrip-tions, resp., sensible proposals to solve the conflicts and problems. The GTA method seems to abstract the rich task insight that the analyst acquired via ethnography in such a way that the important aspects were saved, and thus were available for systematic redesign.

ACKNOWLEDGEMENTS

Thanks to CEC, DG 13 for support under project COST 14; thanks to the members of COST 14 project IMPACT, and to Bonnie Nardi, Brigitte Jordan, Dominique Scapin and Susanne Sebillotte, for stimulating discussions; and many thanks to clients and employees of the Amsterdam branche of the Dutch social security system who trusted us enough to let us be part of their world.

REFERENCES

Card, S. K., Moran, T. P. and Newell, A. (1983) *The Psychology of Human-Computer Interaction.* Lawrence Erbaum Ass., Hillsdale NJ.

De Haan, G., van der Veer, G.C. and van Vliet, J.C. (1991) Formal modelling techniques in human-computer interaction. *Acta Psychologica* 78, 27-67.

Johnson, P. (1989) Supporting system design by analyzing current task knowledge. In D. Diaper (ed) *Task Analysis For Human-computer Interaction.* Ellis Horwood, Chichester.

Jordan, B. (1996) Ethnographic Workplace Studies and CSCW. In D. Shapiro, M.J. Tauber and R. Traunmueller (eds). *The Design of Computer Supported Cooperative Work and Groupware Systems.* North-Holland, Amsterdam, 17-42.

Kieras, D. and Polson, P. G. (1985) An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies* 22, 365-394.

Moran, T. P. (1981) The command language grammar: a representation for the user interface of interactive computer systems. *International Journal of Man-Machine Studies* 15, 3-50.

Mumford, E. (1987) Socio-technical systems design: evolving theory and practice. In G. Bjerknes, P. Ehn and M. Kyng (eds). *Computers and Democracy.* Aldershot, Avebury, 59-77.

Nardi, B. (1995) *Context and Consciousness: Activity Theory and Human Computer Interaction.* MIT Press, Cambridge MS.

Payne, S. J. and Green T. R. G. (1986) Task-action grammars: A model of the mental representations of task languages. *Human-Computer-Interaction* 2, 93-133.

Scapin, D. and Pierret-Golbreich C. (1989) Towards a method for Task Description: MAD. In L. Berlinguet and D. Berthelette (eds). *Work With Display Units 89.* Elsevier, Amsterdam.

Sebillotte, S. (1988) Hierarchical planning as a method for task-analysis: the example of office task analysis. *Behaviour and Information Technology* 7 (3), 275-293.

Tauber, M. J. (1990) ETAG: Extended Task Action Grammar - a languagefor the description of the user's task language. In D. Diaper, D. Gilmore, G. Cockton and B. Shackel (eds). *Proceedings INTERACT 1990.* Elsevier, Amsterdam.

van der Veer, G. C., Van Vliet, J. C. and Lenting, B.F. (1995) Designing complex systems - a structured activity. *DIS 95.* Ann Arbor, MI: ACM Press, 207-217.

van der Veer, G. C., Lenting, B. F. and Bergevoet, B. A. J. (1996) GTA: Groupware Task Analysis - modeling complexity. *Acta Psychologica* 91, 297-322.

Quality Models in Usability Evaluation of Business Management Software

Marcin Sikorski
Faculty of Management and Economics, Ergonomics Dept.
Technical University of Gdansk, ul. Narutowicza 11/12, 80-952 Gdansk, Poland
e-mail: msik@sunrise.pg.gda.pl

ABSTRACT
This paper presents the characteristics of quality models which are developed by evaluators while evaluating usability of a software product (management support). General differences among quality models for users, HCI experts, company managers and software developers are presented. Initial results of a small-scale research confirm that quality models of the users are mostly task-based, while quality models of experts are mostly document-based.

Keywords
software quality, usability evaluation

INTRODUCTION
One of the current problems in software engineering practice is the assurance of software product quality. This is a difficult subject because:

- any commercial product should match requirements of various users and also should be suitable for a variety of task situations in real work environments,
- the same product can come from various environments and technologies, which have differing impacts on the quality of final product,
- software developers and project managers often understand product quality in a much different way than end users,
- the number of available quality assurance and evaluation methods makes a difficult selection for a specific case.

Both practitioners and researchers agree that all types of software quality metrics must include:

(a) quality of engineering design (quality of construction) and

(b) quality of product usage in real tasks (usability).

The question remains open how to evaluate their impact on the quality of a final product and how the quality of development process affects both components.

SOFTWARE QUALITY EVALUATION
Reliability, safety, modifiability, robustness and reusability are common attributes of software product quality used by software developers (Pressman. 1993). These attributes however can be evaluated only by analysis of the code and experimental measurements.

The standard ISO9126 Software Product Evaluation Characteristics specifies six quality attributes: functionality, reliability, usability, efficiency, maintainability and portability. This standard provides high-level evaluation guidance but does not attempt to build quality into software product or to address practical usability issues in a substantial way.

Both mentioned sets of attributes represent the point of view of developers, who tend to put the quality of construction first, and to assume that high quality of design will naturally result in high quality of usage for a given product.

Whether this assumption is true will be verified by:

- end-users encountering usability problems when implementing the product to real work settings,
- company managers who will make pre- or post-purchase evaluations of cost/quality trade-off for the product.

Both groups base judgements of the quality of the considered product mostly on practical experiences from typical tasks; in other words -- on perceived product usability.

Usability-oriented estimations can be made by:

- knowledge-based evaluation: referring to models, guidelines or standards on usability /quality; this type of evaluation is made by HCI experts, so it is highly subjective but often correct if typical tasks of the users are appropriately represented;
- experimental user testing: using observations, video recording of interactions or post-hoc questionnaires and interviews; objective and subjective data are thus combined in order to detect and analyse difficulties in product usage in actual tasks.

QUALITY MODELS
Knowledge-based evaluation and user testing are usually performed by developers or experts, who sometimes are employed as consultants for purchase-responsible managers. In all types of evaluation the usability (as task-related quality) is estimated by comparing the actual product with a hypothetical, ideal one, assumed to posses all recommended task- and user-centred characteristics. To make this comparison possible, a person performing an evaluation must rely on a reference quality model, com-

posed of various attributes and characteristics affecting product usability.

Experts, users, managers and developers have different imaginations and expectations about product quality. It is reflected in different preferences as to desired product functionality. These preferences and quality attributes are presumed to be classified together into a hierarchical structure which forms a quality model of the product. This quality model (or a family of quality models) is specific for each individual involved and also has some characteristic features across the evaluators group:

- Experts base quality evaluation on their knowledge, design and application experiences, as well as on available documents (user interface guidelines, standards) they are familiar with. Abstract evaluation criteria, their relative hierarchy and presumed dependence make up a quality model which the expert uses in evaluation process. Conformance testing, documentation analysis and formal modelling are popular evaluation techniques which support the development of knowledge-based quality model for experts.
- Users to be involved in an experimental testing also use a quality model, but of a different kind: their model is usually less structured, often composed of isolated functions important only in typical tasks the user knows, neglecting at the same time aspects important for other users and system administrators. The user's quality model is often intuition-based, benefit- and performance- focused, and almost never refers to a human factors knowledge base or documented user interface recommendations. Previous experiences with similar type of software also contribute to the development of the user's quality model which is essentially task-based.
- Managers use a quality model which reflects their expectations as to the product quality, viewed from the perspective of the whole company or department they are responsible for. Data security, maintenance problems or costing factors play a much more important role than usability issues. Benefits the product offers for the company demand learning and adaptation which are often considered as natural costs the users (employees) have to bear. The manager's quality model is mostly utility-based but sometimes refers also to documents (e.g. ISO9000-3)..
- Developers build their quality models on formal software evaluation methods, based on analysis of the code. Sophistication of the code is said to represent the technical quality of the product which is not likely to be impaired, even if confronted with frustrating implementation experiences. Blaming the users for product misuse, low performance or making errors is a common indication that a code-based software developer's quality model was used. In this type of quality model usability issues appear in lower areas of the model, if they are present at all.

Observed differences in quality models of experts, users, managers and developers are the result of different types of knowledge and experience they engage into evaluation. Also differing personal attitudes to the evaluation process and product have serious impact. As a result, views of the product and relevant quality models of various evaluators differ substantially. Moreover, because various evaluators focus attention on different aspects of the same product, evaluation results are often hard to compare or combine. These differences seem to be especially important for business management software which involves at least four parties at various stages of its life cycle: software developers, HCI experts, users and company managers.

Because current deficiencies in evaluation methodologies stem in a big part from different and inconsistent quality models the evaluators apply to the analysis, it seems reasonable :

- to investigate the impact of quality model on evaluation results, and possibly also logical relationships among various quality models;
- if possible, to develop an evaluation methodology in which a quality model would be an integral part of the method.

If both goals could be accomplished and a family of quality models could be standardised to some degree, then the results of different evaluations would be easier to compare and transfer. As a result, evaluation could become more objective and verifiable because it would be apparent which quality model was incorporated as referenced in the evaluation method.

QUALITY EVALUATION FRAMEWORK

In order to check the possibilities for incorporating quality models into evaluation methods, we modified an existing usability testing procedure, consisting of the following steps:

1. Gathering functionality recommendations from the users.
2. Developing a generic quality model.
3. User experimental testing of the package (in training and a series of typical tasks).
4. Filling in a Software Evaluation Questionnaire.
5. Discussion with the subjects.
6. Psychological tests (optional).
7. Experimental data analysis.

We tested this approach when evaluating usability of payroll packages for small and medium companies. The evaluation was made by user testing and by heuristic evaluation by experts, according to specified evaluation criteria (Nielsen, 1993). This research was limited to only two groups of evaluators: users and HCI experts, which seem however sufficient to detect differences in the quality models in both groups. Six experienced users of payroll packages took part in the experiment, and three HCI experts.

Quality model extraction: Users

We assumed, that if we extract the most often expressed quality preferences and cluster them into related quality

(usability) attributes, then a generic quality model could be developed, representative for the interviewed group of users.

First, we interviewed users about the functions and features the ideal product should possess in order to be an efficient tool in their tasks (and to be primarily considered for purchase for their company).

The users placed on the list more than 20 different requirements relevant to various areas of system use, from simple data entry to guidance and support. Next, the most frequent requirements were clustered into related groups which formed quality attributes important for the users. These groups of attributes were next treated as main components of a generic quality model representing users view of considered software product..

Table 1. Components in the quality model for users

Requirements #	*User's expectation-Product feature*	*Quality criteria*
1, 2, 5, 6	Functionality	PRODUCTIVITY
1, 2	Speed	
1, 4, 5, 7	Time saved	
1, 2, 3	Reliability	
11	Good documentation	EASE OF USE & LEARNING
19	Screen readability	
6	Simple work procedures	
6, 12, 13	Intuitiveness and support	
10, 14	Easy communication	
17	Easy learning, support	
9, 15	Simple correction of errors	
16	Flexible installation	FLEXIBILITY
8	Customizable formulas	
19	Customizable screen	

During the interview session the users could easily list functions they find important, but they had some problems in:

- specifying quality attributes (which should later serve as evaluation criteria) relevant to functions they had listed and to desired features of user interface;
- specifying relationships among these quality attributes (importance ranking and embedding problems);
- use of terminology -- the same descriptors (attributes) had different meaning for various users.

To make this task easier, we suggested a free recall and listing of the recommended functions, but linking them only with quality attributes that were selected from available menu.

The subjects were however not able to generate by themselves relationships between these attributes, so guidance and help of moderators-researchers who conducted the experiment was needed again. The researchers tried to be objective and not to influence the users in specifying the relationships between the quality attributes; they rather offered technical help in clustering them into related groups.

The quality model from Table 1 can be also presented in a form of tree similar to the one in Fig. 1. The tree structure much better represents hierarchical nature of the quality model. Although it is much simplified, it remains usually sufficient for practical usability testing.

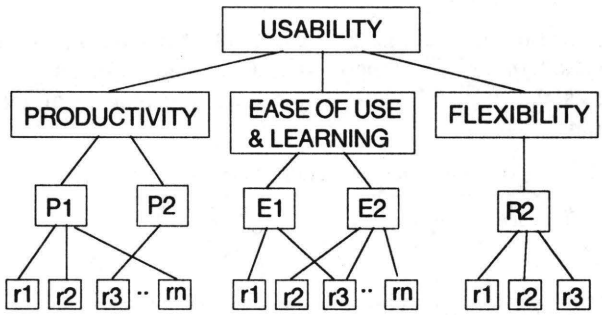


Fig. 1. Quality model as a tree of quality attributes

(r_i - requirements).

This stage of experiment produced a balanced, generic quality model as a tree which had three levels of detail:

- lower level - specific recommendations as to functions and features (r_i) and lower level quality attributes (P_i, E_i, R_i);
- upper level - higher level quality attributes, representing evaluation criteria and main components of user-perceived quality,
- top level - type of evaluation: usability (which in fact means task-related quality for the subjects).

After the generic quality model was developed in order to specify evaluation criteria, the users started testing the payroll products in a series of typical tasks. The next stages of usability testing were performed according to the plan presented on the previous page but the remaining part of this paper concentrates only on the quality models.

Quality model extraction: Experts

The HCI experts were also asked to specify their preferences as to quality attributes of the same ideal payroll package. Because the experts did not have a thorough knowledge of the task domain, they skipped the stage of listing the functions needed and started from specifying

quality attributes and higher-level evaluation criteria. They were more systematic than in the case of users, covering also a broader area and more relevant to usability criteria known from literature and documents (e.g. ISO9241).

The basic components of quality model for HCI experts are shown in Table 2.

The experts had some problems in pointing out all the lower-level components which build up specified quality attributes in the payroll product. Instead of listing the elements of the interface or functions available (as the users did), they rather focused on various types of task- and learning- oriented support the product should offer. It shows a substantial difference in how the experts and users perceive the same product and which application area they find as primary.

The experts were much more interested in specifying possible relationships among available quality attributes, and in verifying the consistency of their quality models. In contrast to users, they expressed much lower interest in precise performance data relevant to the product (like task execution time).

By comparing and discussing individual quality models a generic model for experts was developed, which represents the experts' view of the expected quality of the product.

Table 2. Components of the quality model of experts.

Requirements	*Quality attributes Evaluation criteria*	*Evaluation metric*
*	Suitability for the task	
*	Data security	
*	User control	
*	Low maintenance costs	USABILITY
*	Error tolerance	
*	Suitability for learning	
*	Modifiability	

The quality model of experts turned out to be not as hierarchic as the one of the users, probably because the quality attributes seem to be much broader in scope and overlapping in many areas. Due to limited number and broader scope, quality attributes from Table 2 did not present any special problems in their ranking or embedding.

An interesting observation is shown in Table 3, which shows the appropriate percentage of all quality attributes (from users and experts) which can be connected or to task experience or to professional documentation as their source of origin. The sum in rows is not equal to 100% because in some cases the classification of attributes was doubtful.

Table 3. Task- and document-based quality attributes

attributes:	*task-based*	*documentation-based*
users	71%	11%
experts	18%	64%

These rough data however seem to confirm that:

- the quality expectations and relevant quality model for users is based on task experience and practical knowledge of task domain,
- the quality model for experts is built upon:
  - -understanding the interactions between user and software,
  - -knowledge of standards and other documents which suggest evaluation methodologies,
  - -consciousness that complex relationships among quality attributes exist, which are affected by both imperfect modelling of the task context.

Statistical analysis of these relationships could be performed, but the main difficulty here lies in imprecise classifying of many of quality attributes. Some of them may be both task- and documentation- related, as well as dependent on specific programming techniques applied for the developing the code, what affects for instance the performance and speed of computing operations.

DISCUSSION

Observed differences in quality models for users and experts presented in this paper seem to have a big impact on quality and usability evaluation of software products.

The users and experts view the same product from different perspectives and they focus attention on different aspects of the product. Users assess product quality referring to the practical experiences from typical tasks they know, and for them usability seems to be a sufficient metric for quality.

Experts in turn attempt to include the quality of interaction as a big part of product quality, but also quality of engineering design (esp. user interface) plays an important role, as well as product maintenance and administration issues.

Similar and additional differences would be probably found if two other groups of evaluators were studied: managers and software developers, who would highlight other quality aspects of the same payroll packages.

Producing appropriate quality models for all involved evaluators may be not easy in many cases. This activity seems also to be close to knowledge acquisition because a reflection model has to be made of the attributes, associations and beliefs involved in the human assessment procedures. For this reason simplicity should be viewed as a value, even if the analysis remains partly incomplete.

Specifying quality models and finding relationships between quality preferences of developers, users, managers and experts may give an important insight for developing products that offer balanced quality on all stages of life cycle, for various groups of users.

It should be also noted that a quality model is likely to change as the user develops his/her skill, explores new tasks or discovers new opportunities in teamwork. Regular analysing of quality models could potentially discover discrepancy areas between users expectations and what product has to offer; possible improvement areas could be then easily specified for the particular product.

Additionally, if quality models could be combined within one methodology for software quality evaluation, then the differences in evaluations made by various groups of evaluators could be easier explained and even predicted in advance.

SUMMARY

This paper attempted to present some initial observations on the quality models characteristic for various evaluators of software products quality. These observations were made as a preparatory study for a research project which aimed also to develop of a methodology and software tools to support the process of capturing quality attributes for quality models, marking relationships among them and developing a family of quality models relevant to the same software product.

REFERENCES

ISO 9000-3 (1991). Quality Management and Quality Assurance Standards. Part 3: Guidelines for the application of ISO9001 to the development, supply and maintenance of software.

ISO9126(1991). Software Product Evaluation Characteristics and Guidelines for Their Use.

ISO9241(1992). Ergonomic Requirements for Office Work with Display Terminals (VDTs).

Nielsen J.(1993). *Usability Engineering*. Academic Press, New York.

Pressman R.S.(1993).*Software Engineering*. McGraw Hill, New York.

Cognitive and Organizational Dimensions of Task Appropriateness

Chris Stary
Department for Business Computing
University of Linz
Freistädterstraße 315
A-4040 Linz
+43 732 2468 7102
office@ce.uni-linz.ac.at

Alexandra Totter
Working Life Research Centre
Aspernbrückengasse 4/5
A-1020 Wien
+43 1 21 24 700
forba@xpoint.at

ABSTRACT
Task appropriateness is considered to be a key design principle for the usability of interactive software. Although it is related to the tasks users are performing, it rather addresses the functionality of software instead of the organizational and cognitive context of an user interface. In this paper we examine critically the interpretation of task appropriateness and its dimensions in design frameworks, standards, and evaluation methodologies. We propose a strategy to overcome the epistemological deficiencies through a theoretical framework. This framework enables further research for the integration of cognitive and organizational dimensions of task appropriateness.

Keywords
task appropriateness, usability engineering, epistemological analysis, framework revision.

INTRODUCTION
Task appropriateness is considered to influence the usability of user interfaces, and thus, their acceptance essentially, e.g. (Rosson et al., 1995). User interfaces are those elements that relate the technical part (i.e. the computer system) with the social one (i.e. the human). Hence, interactive systems are socio-technical systems (Stary, 1994).

Task appropriateness is one of the design principles that is part of most of the current techniques for the development and validation of user interfaces. Task appropriateness is also part of standards, such as the DIN-Norm 66 234 (part 8), as well as federal directives, such as the EU-Directive for Human-Computer Interaction (EC, 1990). The major concern of task appropriateness is the elimination of hindrances through interaction devices or styles for the effective and efficient accomplishment of user tasks. User interface development should focus on the individual organization instead of optimizing the execution of functions (system tasks) for a particular platform, hardware or operating system. Hence, the influence of events and factors not related to end user tasks has to be minimized.

Although a widespread agreement exists concerning the goal to be achieved through task appropriateness, i.e. how users should benefit from task-appropriate interactive systems, in the course of implementing task appropriateness for the design or the evaluation of user interfaces there exist terminological and conceptual differences. These differences concern the level of detail describing task appropriateness (granularity) as well as the involved dimensions and key factors establishing task appropriateness. For instance, in approaches dominated by standards task appropriateness is explained through a single general statement that is illustrated through particular examples, whereas in approaches stemming from psychology of work several factors are detailed that concern task appropriate user interfaces.

In addition, differences are based on different dimensions of task appropriateness. For instance, standards focus on technological issues, namely functional design, whereas occupational measurements focus on the organizational dimension of task appropriateness, namely, the individual work place design.

The paper gives an overview of the concept of task appropriateness in the various approaches, including the different layers of abstraction as well as particular factors that have been considered to be relevant in cognitive and organizational ergonomics. It continues with outlining the terminological and conceptual differences as well as similarities between the various approaches for design and evaluation. Finally, a strategy is proposed to overcome epistemological differences, in order to achieve an unifying concept of task appropriateness for user interface design and evaluation.

TASK APPROPRIATENESS
In this section we review standards and styleguides for user interface, as well as methodologies for evaluation in order to demonstrate the

- variety of explanations
- relative importance

of task appropriateness for designing user interfaces. Due to the space limits we give prototypical examples from each of the sources.

There exist several standards for the design of user interfaces: ISO 9241 (part 10,11), VDI 5005, DIN 66234 (part 8), etc. For instance, in VDI 5005, the standard for office communication task appropriateness is the third issue besides enhancing competence and activity spaces for the design and evaluation of user interfaces. Organizational issues are not integrated in the design principles in this standard. They are referred to in VDI 5001. Hence, there is no refinement of task appropriateness for organizational

issues, such as task allocation between human and computers, within VDI 5005.

In DIN 66234 part 8 task appropriateness is one of 5 basic principles for ergonomic dialogue design. It is described as effective support of end user task accomplishment, without stressing the user due to interaction features, in particular due to inherent properties of interaction devices. The derived principles concern the adaptation of the input of control information and data, the support of standard tasks and default values, and more. Even when user characteristics are addressed, such as the capabilities of human information processing, the standard advises the adjustment of software features, e.g. the splitting of windows to avoid information overload.

In ISO 9241 (part 10) the following explanation is given: 'Wherever it is consistent with the task requirement there should be possibilities offered to reduce complexity by either grouping, zooming or equivalent techniques. The presentation of output should match the user's task requirements. It should be possible to condense input, minimize repetitive work, and control the data according to the task specification. As an overall goal of the designer's aim to achieve suitability for the task, tradeoffs between consistency and complexity may be necessary while still maintaining a task-oriented view.' In accordance with the other standards the measurements are also oriented towards software features. From this short extraction it also becomes evident that the national and international standards are almost identical in targeting improvements at the user interface.

In order to establish an industry standard styleguides have been developed: SAA-CUA by IBM, Human Interface Guidelines by Apple, Motif Style Guide by OSF, etc. For instance, the SAA-CUA Guide is based on 3 principles:

- user controllability,
- reduction of memory load, and
- consistency.

Task appropriateness is a feature of controllability. It is explained as the adaptation of dialogue steps to the steps of task accomplishment. Task appropriateness requires transparency of the user interface concerning the non-technical aspects. Technical aspects should be hidden from the user. The interaction should be prompted and thus, reduce conscious mental processes for the handling of interface features. Similarly in the standards task appropriateness is also oriented towards the adaptation of software functions.

In evaluation methodologies, such as in (Ravden et al., 1989) task appropriateness (termed as appropriate functionality) is explained as the requirement for user interfaces 'to meet the needs and requirements of users when carrying out tasks' (p.36). Task appropriateness is measured through 11 questions for a selected task. The questions involve different dimensions in a non-systematic way. As a consequence it has to be clarified to which element of an interactive system (user, task, interaction features) each of the questions refers to.

In EVADIS II (Oppermann et al. 1992) two clusters of criteria have been identified: 9 criteria to evaluate the organization, and 12 criteria to evaluate the software. Task appropriateness is part of the cluster for software evaluation and denoted with usability: Usability contains those aspects of task appropriateness that are related towards the proper functionality of the software for task accomplishment (p. 31, 64). In addition, comfort comprises those aspects of task appropriateness that are set up in DIN 66234 part 8 (p. 31f, 64). The measurement is performed through 3 and 15 questions, respectively after preparing the questions according to the tasks and users at hand.

The standards and styleguides as well as the methodologies mentioned above are dedicated to the design and evaluation of user interfaces. Actually, they restrict the activity space of the developers to the set of functions provided by the software. As a consequence, developers of interactive systems do not take into account the second part of an interactive system, namely the human, and the organization each user is operating in. The concept of task appropriateness should consequently address organizational and cognitive issues.

Another reason to develop a holistic concept of task appropriateness is the nature of design and design processes. According to Hacker (1994) effective design of software corresponds to the design of organizations and work. Hence, the elements of design are tasks. They have to be considered initially in the process of software development, i.e. before the set of software functions is designed (p. 55).

THE ORGANIZATIONAL DIMENSION

In a similar way to Hacker Volpert et al. (1993) identify the acquisition of user tasks and the behaviour of users in the course of task accomplishment as a basic requirement for the design of software that meets human needs. In addition the global goals of human-oriented work, such as job satisfaction should be achieved. Hence, the design of software becomes a subtask of designing work. This relationship between tasks and software enables the development of software designed appropriately for task accomplishment.

When efficient design of software has to start with the design of global tasks (that should be supported through software) the organization and allocation of tasks should be the starting point for system design. The organizational dimension of task appropriateness is considered to be prior to software design.

Ulich (1994) also interprets task appropriateness in the context of work: In the course of software development the tasks have to be allocated to humans and the computer system. Task appropriateness denotes the adequate support of tasks that fulfil the criteria of task design as defined in occupational psychology. These criteria have to be met before design principles for software development are considered. Meeting this requirement will support the definition of tasks that allow holistic activities, job enrichment, social interaction, autonomy in time and activity, individual development and qualification, and self-regulation.

Methodologies for evaluating interfaces take into account the principles of occupational psychology. For instance, in KABA (Dunckel et al., 1993) the following task-related criteria are defined: decision space, communication requirements, time invariants, individual organization, and mental load. High mental load concern the consequences of events or situations that hinder task accomplishment. These hindrances are extrinsic factors of the work process. Hence, their removal require additional activities. They may also result in overload. Hindrances may either concern

- the access to information,
- the navigation through information spaces, such as the WWW,
- motor problems, or
- interrupts.

Overload usually results in time pressure and/or monotony of work.

Cognitive processes are affected through problems with information access and navigation, interrupts that may require a long phase of re-entry into the process of task accomplishment, and mental overload that usually results in unreliable results of work.

All of the criteria are measured in a quantitative and/or qualitative way. Users are observed and/or interviewed. In addition, documents should be analysed and questionnaires including some evaluation procedures should be used. Although KABA is based on human-oriented criteria, the element of work that is under evaluation is the task. Actually, the evaluation focuses neither on the user (how users perceive and accomplish tasks) nor on the process of task accomplishment from the perspective of the user. It rather checks the conditions and constraints that are given or may occur for the completion of tasks.

We find a similar situation for RHIA/VERA (Volpert et al., 1983; Leitner et al., 1993). Goal of the evaluation is an analysis of the planning and mental processes as well as mental overload in the context of work tasks. The planning and mental processes are termed requirements for regulation, whereas the work conditions leading to mental overload are termed hindrances of regulations. The measures are checked in a qualitative and quantitative way. They concern:

- task descriptions,
- a description of mental requirements in a layered framework, and
- a description of mental work load according to different factors, such as intensity, periods of involvement, monotony.

The data are acquired through interviews, documents, questionnaires, and observations. The methodology consists of several phases: preparation, first observation and interview, systemisation, second observation and interview, and integration of results. Although users participate throughout the evaluation process, the individual interpretation of tasks and the organization of work are not considered for redesigning individual workplaces. It is assumed that tasks are prestructured according to a global, static organization of work and the predefined processes to accomplish tasks.

The methodology is structured in a similar way to KABA. In extension to KABA RHIA/VERA's theoretical framework, namely the theory of regulating activities, interprets activities as elements of processes. However, its focus is also the set of conditions and constraints for work activities, and not the individual workplace including the individual capabilities and specific possibilities for improvements.

Methodologies stemming from occupational psychology basically address only one aspect that is embedded into user interfaces, namely tasks, including their static and dynamic organization. Due to their focus on the environment of an interactive system, i.e. the work system, they do not focus on the users within the context of an interactive system. They focus on the user within the work system, but not on the user as the social component of the interactive system. As part of this system the user is directly involved in the human-computer interaction at the user interface. Hence, we have also to consider the cognitive dimension of task appropriateness.

THE COGNITIVE DIMENSION

When discussing the cognitive dimension of task appropriateness we will follow two directions: One direction is to find experiments, criteria or methodologies that have been performed or developed, respectively, in relation to task performance. These approaches allow us to identify relationships between organizational and cognitive dimensions almost immediately. Hence, the process of achieving an integrated approach, capturing both dimensions of task appropriateness is facilitated.

The other direction that we will follow concerns the analysis of user-oriented measurements. Their development is not driven by the relation of measures to work tasks. It could focus either on human perception or on the handling of interaction styles, such as direct manipulation. In this case, an integration of these approaches into a multidimensional approach for designing and evaluation task appropriateness of user interfaces requires further elaboration. In the following we will give examples for both directions.

One of the first established approaches for usability testing involving work tasks as well as end users is the Keystroke-Level Model (Card et al., 1980). It captures a way of measuring the cognitive processing time of users. In the model this time is represented through time allowances for mental processes. These allowances are part of key points of interaction. They are identified through work analysis. Task performance times for interaction are estimated under the assumption of error-free and undisturbed performance of expert users in the given context. Although the model supports the measurement of user performance and the development of guidelines and principles for design there is a gap between the tasks at the semantic

level and the measures, i.e. the keystrokes, since the latter address the syntactical layer of interaction and task accomplishment. As a consequence, for the development of guidelines and improvements for the support of work tasks at the user interface, such as writing a letter, additional effort has to be spent for bridging the addressed gap.

The MUSIC Performance Measurement Method (Corbett at al., 1993) has been developed for the evaluation of practicality and suitability of commercial applications. Its output consists of measures and diagnostic data concerning the

- relative user efficiency
- productive period
- snag, search, and help times.

The relative user efficiency addresses the effectiveness related to the costs of performance where effectiveness measures how correctly and completely tasks can be achieved by a particular user. As such, an indicator is developed how learnable software functions are for the accomplishment of tasks for users. The quantitative measurement is achieved through a comparison of specific users to that of expert users in a particular context.

The methodology to be followed in MUSIC for a specific product to be evaluated starts with the Usability Context Analysis to identify goals and tasks. Then, the Performance Measurements Handbook and the DRUM Scheme Manager supports the specification of the targets of the evaluation. With the help of the DRUM Evaluation Manager the evaluation is prepared. After performing the tests involving the users (supported through the DRUM Kit and Recording Logging for video analysis), with the help of the DRUM Metrics Processor and Evaluation Managers the test data are analysed, interpreted, and compiled into an usability report.

The productive period is considered to be the proportion of time spent not having troubles in interaction. This measure may either be directed towards tasks or interaction features. In both cases the mental capabilities to perform activities (either for controlling the interface or accomplishing tasks) are involved. The snag, search, and help times capture those times when users search for help. Again, this measure may either concern the understanding of tasks or the control of interaction, and thus, the handling of interaction media or styles.

The Software Checker (TCO, 1992) is an instrument for evaluation that has been developed to support the measurement of task appropriateness. Unfortunately, it does not define a criterion that could capture any of the organizational aspects we have addressed in the previous section. It rather defines a general criterion that is related to cognition and is oriented towards users: logical design. It is explained as the adaptation of programs to mental processes. The more the behaviour of software is adapted towards human cognitive processes the more human work can be supported through software. However, logical design is defined in such a general way that it cannot be measured directly, and thus, has to be refined to lead to design principles or evaluation criteria.

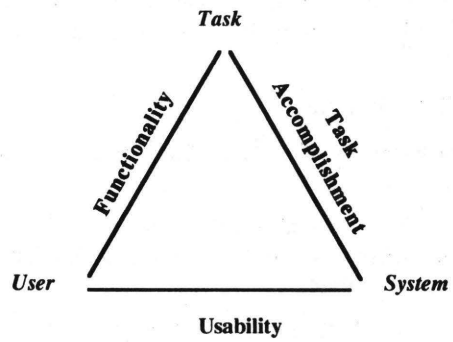
Logical design shows the superiority of task appropriateness towards other design principles. In this case, the adaptation towards user needs is considered to be a subgoal of task appropriateness. Rauterberg (1992) defines the features for individual adaptation as a prerequisite for the individual design and extension of activity spaces. This subgoal relationship between the two principles documents the mutual interdependencies of organizational and cognitive dimensions of task appropriateness. Other principles that are related to task appropriateness are usability principles, such as transparency, consistency, and the provision of immediate feedback, e.g. (Nielsen, 1993). They allow the emergence of user expectations, and lead finally to a principle termed user-conformity (e.g., Stary, 1994).

Then, users are enabled to develop effective behaviour which is manifested in the contribution of mental models (e.g., knowledge of the task and interaction styles) to effective cognition or understanding. User performance, behaviour, and cognition are henceforth referred to as positive indicators of the usability of the product (Sweeney et al. 1993).

A FRAMEWORK FOR INTEGRATION

In order to proceed with an integration of the two basic dimensions of task appropriateness, namely the static and dynamic organization of task as well as the cognitive aspects for task perception and accomplishment, we suggest to utilise a framework. This framework has been developed for describing the mutual context of the elements integrated in any user interface. It has to be extended along the cognitive dimension, in order to meet the requirement of the addressed conceptual integration.

Frese et al. (1989) have developed the addressed framework in order to foster the discussion of interface features within the triad of tasks, users, and the (computer) system (see also the Figure). The connection between the system and the tasks denotes the *functionality* of the system. The relationship between the system and users establishes the *usability* of the system, and the connection between the tasks and users is termed *task accomplishment.*



In contrast to the contingency model proposed in (Stary, 1994) this framework requires the user to be considered in the context of the system and the tasks at the same time. Hence, the role of the user as part of the work system

cannot be handled isolated from the role of the user at the user interface any longer. Organizational and cognitive aspects have to be considered in mutual context.

Functionality in the sense of Frese et al. refers to the set of software functions enabling task accomplishment at all. Usability addresses the ease of use for users. Task appropriateness addresses the use of the software functions for the accomplishment of tasks. In addition, they name fitness, controllability, and error handling as further criteria. The fitness plays an important role for the integration process. It captures the mutual context of the tasks, users, and the system. It denotes the mutual suitability of functionality, usability, and task accomplishment.

However, for the integration of cognitive aspects the framework of Frese et al. has to be extended, since they consider the fitness of the task and the system to be a sufficient condition for task appropriateness. As we have shown, due to the mutual interdependencies of organizational and cognitive aspects the fitness of users, tasks, *and* the system is the prerequisite for task appropriateness. Once this integration has been achieved the development of valid design principles and instruments for evaluation is enabled.

REFERENCES

Apple Computer, Inc. (1992). *Macintosh Human Interface Guidelines*, Reading, Mass.: Addison-Wesley.

Card, S.; Moran, T.P., Newell, A. (1980). The Keystroke Model for User Performance Time with Interactive Systems. *Communications of the ACM*, 23 (7), 396-410.

Corbett, M.; McLeod, M. Kelly, M. Quantitative Usability Evaluation - The ESPRIT MUSiC-Project (1993). Human-Computer Interaction: Applications and Case Studies. *Proc. HCI'93*. Elsevier. Amsterdam, 313-318.

DIN 66 234 Teil 8 (1988). *VDU Workplaces. Fundamental Principles for Dialogue Design* (in German). Beuth, Berlin.

Dunckel, H.; Volpert, W.; Zölch, M.; Kreutner, U.; Pleiss, C.; Hennes, K. (1993): Contrastive Task Analyse for Offices. *The KABA Manuals* (in German). Teubner, Stuttgart.

EC-Directive (1990): Directive 90/270/EWG, Amtsblatt der Europäischen Gemeinschaft. L 156, 29.5.1990.

Eberleh, R.; Oberquelle, H.; Oppermann, R. (1994). *Cognitive Ergonomics* (in German). deGruyter, Berlin.

Frese, M. & Brodbeck, F. (1989). *Computers in Office and Administration. Putting Psychology to Practice* (in German). Springer: Berlin, Heidelberg.

Hacker, W. (1994). Occupational Psychology and Psychology of Organizations. *Fundamentals in Cognitive Ergonomics* (in German). *Eberleh et al.* (1994), 53-93.

IBM Corp. (1991a). SAA/CUA Guide to User Interface Design, IBM Report SC34-4289-00. Cary, North-Carolina.

IBM Corp. (1991b). SAA/CUA Advanced Interface Design Reference, IBM Report SC34-4290-00. Cary, North-Carolina.

ISO 9241, Part 10 (1990). ISO 9241 Part 10: Ergonomic Dialogue Design Criteria, Version 3, Committee Draft, December 1990.

ISO 9241, Part 11 1990). ISO 9241 Part 11 Usability Statements, Version 2.5, Committee Draft, July 1990.

Leitner, K.; Lüders, E.; Greiner, B.; Ducki, A.; Niedermeier, R.; Volpert, W. (1993). Analysis of Mental Requirements and Loads for Office Work. The RHIA/VERA Methodology. Manual (in German). Hogrefe, Göttingen.

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

Open Software Foundation (1993). *OSF/Motif Style Guide Release 1.2*. Engelwood Cliffs, New Jersey: Prentice Hall.

Opperman, R.; Murcher, B.; Reiterer, H. & Koch, M. (1992). *Evaluation in Cognitive Ergonomics. EVADIS II* (in German), 2nd edition. Berlin, New York: de Gruyter.

Rauterberg, M. (1992) Are We Able to Measure Usability? (in German) Software Ergonomie. GI.

Ravden, S.; Johnson, G. (1989). *Evaluating Usability of Human-Computer Interfaces. A Practical Method.* Chichester: Ellis Horwood Limited.

Rosson, M.B.; Carroll, J.M. (1995). Integrating Task and Software Development for Object-Oriented Applications. *Proc. CHI'95*. Denver, CO, ACM Press, 337-384.

Stary, Ch. (1994). Interactive Systems. *Cognitive Ergonomics and System Development* (in German). Vieweg, Wiesbaden.

Sweeney, M.; Maguire, M.; Shackel, B. (1993) Evaluating User-Computer Interaction: A Framework. *Int. Journal on Man-Machine Studies*, 39, 689-711.

VDI 5005 (1990). Verein Deutscher Ingenieure e.V. VDI 5005, *Cognitive Ergonomics in Office Automation* (in German), Düsseldorf.

Volpert, W.; Oesterreich, R.; Gablenz-Kolakovic, S.; Krogoll, T.; Resch, M. (1989). *A Methodology for the Acquisition of Regulation Requirements for Work Activities VERA* (in German). TÜV Rheinland.

Ulich, E. (1994). *Occupational Psychology*. Poeschl, Stuttgart.

Practical Insights From Studies Of Operator Diagnosis

Erik Hollnagel
Asgeir Drøivoldsmo
OECD Halden Reactor Project
P.O. Box 173, N-1751 Halden, Norw
erik.hollnagel@hrp.no
asgeir.droivoldsmo@hrp.no

Barry Kirwan
Industrial Ergonomics Group
University of Birmingham
Birmingham B15 2TT, UKay
KIRWANBI@novell2.bham.ac.uk

ABSTRACT
The OECD Halden Reactor Project is carrying out a Human Error Analysis Project over several years. Whilst the primary purpose of this project is to improve the modelling and understanding of "cognitive errors", a series of pilot experiments has yielded valuable insights with regard to operator diagnosis. The insights, reported in this paper, address issues related to diagnostic behaviour, diagnostic support, and diagnostic reliability. The outcome of the experiments performed so far has helped clarify the role of diagnosis in Nuclear Power Plant emergencies with specific implications for risk assessment and design of operator support systems.

Keywords:
diagnosis; information presentation; training; nuclear power plant; experiments; alarm systems.

INTRODUCTION

Diagnosis is an important part of operator responsibilities in process control environments, but there is presently inadequate ergonomics knowledge on how to support operators in situations requiring diagnosis. Operator support given via alarm handling, VDU information mimics, or training is based mostly on format considerations and generalised notions of what information should be processed. Ideally such operator support systems should be based on a detailed understanding of the operator's knowledge needs and a description of their strategies in controlling dynamic and time dependent events.

The Halden Reactor Project is presently engaged in a Human Error Analysis Project (HEAP), which was started in 1994, and is planned to continue at least until 1999. The overall objectives of HEAP are: (1) to provide a better understanding and explicit modelling of how and why "cognitive errors" occur, and (2) to provide improved design guidance for the development of advanced integrated control rooms. Together they raise a number of issues that are central to cognitive ergonomics, namely: the notion of human erroneous actions (and "cognitive error"), the modelling of human erroneous actions, and the classification and description of the causal structure of human erroneous actions as a basis for qualitative and quantitative predictions.

The objectives of HEAP are addressed by an experimental programme, carried out in the Halden Man-Machine Laboratory (HAMMLAB) which specifically aims at obtaining a clearer understanding of diagnosis in difficult scenarios. This paper summarises the insights gained from the analysis of the first set of pilot studies with regard to the nature of diagnosis and the use of operator support systems. The studies used experienced Nuclear Power Plant (NPP) operators in realistic and challenging scenarios, with trained observers using sophisticated observation methods. The insights reported here include data from the first three pilot experiments of HEAP, as well as data from the Staffing Project (Follesø et al., 1995), which investigated central control room staff level requirements for next generation Advanced Light Water Reactors.

Whereas the various HEAP pilot studies used combinations of concurrent verbal protocols and eye movement tracking, the data obtained from the Staffing Project were based on the communication that naturally took place amongst the operators in a crew. This report therefore combines results and insights from two experimental programmes that differ on internal and external validity. The specific outcomes of the programmes can be compared, thereby improving insights on real world performance.

DIAGNOSIS AND OPERATOR TASKS

Diagnosis is itself a complex topic, and it is useful to define three inter-related areas within which the details of the studies can be considered:

- Diagnostic behaviour - linked to cognitive psychology and modelling of diagnosis
- Diagnostic support - linked to ergonomics, the design of training, and procedural systems, and alarm and interface systems which support diagnosis
- Diagnostic reliability - linked to Human Reliability Assessment, in particular what error types occur and with what likelihood. Diagnostic reliability will, however, not be discussed in this paper.

Diagnostic Behaviour

Diagnosis is performed either by single operators or by teams. A review of the literature (Kaarstad et al., 1994 & 1995; Kirwan, 1994) suggested that the core of diagnosis is problem-solving, and that there were a number of definable strategies available, with evocative names such as topographic, symptomatic, data-driven, and hypothesis-testing. It was further surmised from the literature that such strategies were competitive rather than complementary, i.e., that subjects would tend to exhibit one or another

but would not be flexible in their application, and that strategies would be relatively stable across time. Since the literature about these strategies in realistic NPP environments is meagre it was decided to test these assumptions and also to determine if certain strategies were more likely to result in diagnostic success than others.

Concerning diagnosis in teams, there is little known about how it proceeds and to what extent team diagnosis is preferable to single operator diagnosis. The topic has for some years been studied in the field of aviation under the name of Cockpit Resource Management (CRM; Caro, 1988) but it is uncertain whether these results can be transferred to the NPP domain. Given the scant prior knowledge in this area, the studies focused on describing the characteristics of team diagnosis and the function of communication between operators and other plant personnel.

Diagnostic Support

Diagnosis can be supported in several ways, for instance via procedures, training and experience, information presentation, and alarms systems.

Diagnosis is typically regulated by **procedures**. It is obviously important that these procedures are easy to understand and error resistant, and that the need for operator judgement is minimised. Unfortunately this approach is fraught with problems. Firstly it is difficult to describe with sufficient precision the conditions that may exist when a diagnosis is required. Secondly, the more "rule-bound" operators become, the less they will be used to rely on their judgement, and the skill of diagnosis without reliance on procedures may diminish over time, unless sustained by appropriate training.

Training and experience may have a positive impact on diagnosis; highly trained and experienced operators have better pattern recognition abilities (i.e., more experience of seeing and recognising clusters of symptoms), more mature judgement processes, and probably more confidence in their judgements. Yet experience may also be a major factor leading to misdiagnosis (e.g. the frequency gambling and similarity matching biases described by Reason, 1990). Since operators often will develop their own style or approach, it is of interest to determine what aspects of training or experience lead to success, failure, or recovery from failure. Such insights may lead to recommendations concerning the problem solving tactics and strategies that operators should be taught.

Information presentation and interface design are predicated on general monitoring and supervisory control in normal conditions: the aim is to provide the relevant information in an ergonomic manner, complying with established standards for human factors and cognitive ergonomics. It is possible that this will not be optimal to support diagnosis, which takes place in out-of-normal conditions. It is therefore appropriate to determine whether and how the information presentation and interface should be improved to support this activity.

Although **alarms** are a major means of supporting diagnosis, NPP alarm systems are complex and contain so many alarms that they can easily become a hindrance. Alarms are by definition "attention gaining" devices, and can as such both facilitate and disrupt diagnosis. Since alarm system design, and associated issues such as alarm suppression, are quite complex, the studies mainly tried to observe how aspects of the specific alarm system in HAMMLAB supported or hindered diagnosis. Operators might, for instance, be "alarm-driven", i.e., reacting to each alarm as it occurred. Alternatively, they might be more strategic once the initial flurry of alarms had receded, which would mean that they might occasionally ignore alarms for a short period while they carried out their current diagnosis.

RESULTS: DIAGNOSTIC BEHAVIOUR

For each subject-scenario combination of the HEAP studies, the performance data were analysed to determine which strategies the operators used and how often they succeeded / failed. The most striking observation was that each subject only seemed to use one or two strategies, although five subjects used as many as three strategies within a single scenario. Hypothesis driven, topographical and data driven strategies were the most frequent in the pilot studies, and were used in a combination rather than one by one. Although the Staffing Project did not yield sufficient data to make a similar analysis possible, both topographical and hypothesis driven strategies were observed.

Table 1 reviews the combinations of strategies utilised by the subjects for all the scenarios. In almost all cases a combination of strategies was utilised indicating that operators used them flexibly, hence that there was potentially an inter-dependence between the strategies. This places emphasis on training and interface design to support flexible usage of multiple strategies. The one exception may be symptomatic search, which did not appear to have been utilised to any large extent by these operators with the scenarios (Table 1).

Table 1: Strategy combinations observed during pilot studies

Combination				Frequency
Topographic	Symptomatic	Hypothesis driven	Data driven	(no. of scenarios observed)
•				2
•		•		19
•			•	2
	•		•	1
•		•	•	7

Table 1 shows that only five of the possible strategy combinations were seen in the studies; in particular, symptomatic diagnosis was only seen once. Topographic and hypothesis-testing strategies appeared to dominate, both being supplemented by data-driven search. For instance, a topographic strategy was first used to determine the affected area and its likely "causal neighbourhood", then hypothesis-driven or testing approach took over to identify a specific cause or causal pattern. A data-driven strategy might be used both to integrate new data and to check the adequacy of the current diagnostic trajectory, as well as

when neither topographic search or hypothesis generation could proceed without further data.

Insights On Team Diagnosis

Two important aspects that distinguish diagnosis in a real control room from the performance in an experimental setting are distributed decision making and the sharing of information between team members. Distributed decision making implies that diagnosis is carried out by control room staff in co-operation with field operators and support staff, such as the safety engineer and the technical support centre. Sharing of information means that the team as a whole will generally have more knowledge and experience than the individual. This is assumed to enhance diagnosis if and only if such knowledge and experience is shared at the appropriate time.

The following is a summary of the insights gained on team-based diagnosis, based mainly on the observation of crews in the Staffing Project scenarios.

- Crews did appear to exhibit similar strategies to single operators.
- Symptomatic search appeared to be more evident with teams of operators than with single operators.
- Operators in the real control room relied significantly on outside operators for local information, delegating some of the search aspects of the diagnosis to outside operators (e.g. when compared to the HEAP studies in HAMMLAB). The operators therefore appeared to prefer not to rely solely on control room based information.
- Operators in the real control room also preferred to get confirmation from outside operators before concluding on a diagnosis.
- The timely sharing of information is a critical aspect of team performance, and can be both a positive factor and a negative factor in determining team performance. If so, it may speed up diagnosis and can lead to error recovery; if not, it can contribute to delayed diagnosis and diagnostic failure or misdiagnosis.

Finally, it was found that teams were not immune from the effects of tunnel vision or cognitive lock-up. This effect is thus neither confined to single operators, nor to the experimental control room.

RESULTS: DIAGNOSTIC SUPPORT

Use Of Procedures

There is an ongoing discussion on the need for diagnosis in operations that are heavily supported by procedures. Some empirical studies show that even when procedures are available, operators are still sometimes required to diagnose to handle transients successfully. In the Staffing Project some of the scenarios demonstrated that even if the mitigation of the disturbance is supported by procedures, the quality of the mitigating actions partly depended on the quality of diagnosis of the faults involved in the scenario. The outcome of this diagnostic behaviour can greatly influence the crew's ability to cope successfully with the scenario.

Procedures are usually devised to avoid misdiagnosis in the first place but misdiagnoses will inevitably occur. In these cases procedures are an important corrective or recovery mechanism for the operators. Although this is achieved in many cases, there is room for enhancing this function and procedures should include provisions for misdiagnosis recovery as a specific functional objective. This could add significantly to the safety integrity of NPPs.

Although not formally measured, it was noted in several of the scenarios that procedures were not referred to rapidly, i.e., the operators would begin handling the event for 15 minutes or so before accessing formal procedures. This is consistent with European approaches to event handling (as contrasted with US NPPs where procedures are accessed almost immediately), since operators spend the first phase of the event accessing information and carrying out initial trip-related stabilisation activities or pre-trip stabilisation efforts. This initial period also helps to determine which procedure to access.

Information Presentation

In the HEAP pilot studies the information system was fully computerised and process control was carried out primarily through the use of tracker-balls. The process was represented by around fifty process formats, showing larger or smaller portions of the plant using common principles for process mimics.

Although all operators were less familiar with the interface in HAMMLAB than with their own respective interfaces, few of the problems the operator had in handling the information in the pilot studies could be directly linked to diagnostic faults. However, all the interface problems represented distractions that took operator time and therefore contributed to increased time pressure and a corresponding failure to diagnose within the time windows.

Following Flow Lines Within And Between Formats

It is a generic problem for computerised information presentation to decide whether there shall be any overlap between pages, and where to "break" between pages. The following of flow-lines was often a critical activity and was made difficult in cases where the fault origin was at the edge of one page, bordering on another, and where flow-lines were terminated (abbreviated) and not drawn to meet their true connection point on the format. A potential solution is to use the screen as a "window" that moves seamlessly over the display topography. There also appeared to be a problem in operator usage of different types of format. This problem can probably be resolved by improved training, and by ensuring that search tasks can be completed for all scenarios without having to use more than one format type. Lastly, there was a problem with differing levels of detail in formats that otherwise were at the same level in the display hierarchy. This problem may be resolved by improving consistency of format information representation, or by development of a task analysis tool specifically for display format design.

Interaction With The Process

Interaction with the process took place through addressing plant components via a tracker ball. Actions were specified through function keys (start/stop) or by entering specific values into a dialogue box in the lower section of the process formats. In order to manipulate a process component the operator needed first to put that component into manual control mode. When this was forgotten the operator was not able to do anything to the component.

Several interface aspects facilitated diagnosis. For instance, the display showed key process parameters for parallel sub-systems together, thus making comparison easy. This was of help in situations where the operator had a suspicion that a component was not working as it should. In such situations, it is hard to remember what the actual value should have been for that component, and by comparing it to a parallel component, inferences can be drawn to whether the component is working normally. Another useful feature of formats was the line of key process parameters displayed on the top line of the formats. The eye-tracker revealed that the operators frequently would look at this top line, checking primary pressure, secondary pressure etc., without having to swap to another format to get that information.

Alarms

The alarm system is one of the key information sources for operator diagnosis, and the quality of the alarm system is therefore important. In HAMMLAB the overview alarm display was heavily used. The overview is not only used for the purpose of acquiring alarm information, but also to get a general overview of the process in terms of overall thermodynamic balance. The operators seemed to prefer a style of working where they repeatedly could go back and check the overall balance of the plant and the status of key process components via the alarm overview.

The alarms displayed on the overview display were group alarms. To get more detailed information the operator was referred to a list containing all the alarms in a chronological alarm list, using different colours for different priorities. Some of the operators preferred to use the alarm list as their key alarm information. In some scenarios the key diagnostic alarms only remained briefly on the alarm screen before they scrolled off the list. In these cases the operators could easily fail to detect this alarm and would never recover to make the correct diagnosis. Diagnostic failure was therefore linked directly to deficiencies of the alarm system. Certain key alarms require more salience in the interface (e.g. secondary side radiation), or even their own dedicated displays. Some of these alarms are highly informative in diagnosing scenarios, but may be missed by the operators.

Alarm lists could be displayed in two modes, either automatically scrolling so that the last arrived alarm always was displayed, or in manual mode where the operator decided which pages should be displayed. The function buttons for these alternatives were located very close to each other, and it was possible accidentally to freeze the alarm list by using the wrong button. In addition, the button used for silencing alarm bells was located right next to the buttons that would freeze the alarm list. The silence button was often used during a scenario without looking at it. It was therefore possible accidentally to freeze the alarm list, and it actually happened during one scenario.

On the whole the monitoring and control (management) of scrolling alarm lists led to several problems of inefficient alarm handling with the current system. Alarm information and associated alarm management functions (e.g. the compress function) must also be very fast if they are to be used and relied upon by the operators in dynamic and time-pressured events.

CONCLUSIONS

The investigations into diagnostic behaviour and performance have led to a number of insights related to the fields of cognitive psychology, operator support system design and ergonomics, and Human Reliability Assessment and risk assessment. These insights are preliminary in many cases, but nevertheless have increased our understanding of the nature of diagnosis in NPP emergency events. In this respect, the programme has achieved a major aim.

In terms of understanding diagnostic behaviour, the *a priori* strategies were found to be complementary and flexibly used, rather than in direct competition. Two strategies (topographic and hypothesis-driven) appeared to dominate in single operators; teams also exhibited the symptomatic strategy. For teams it was noted that sharing of information is a critical activity, and that teams were not immune from error mechanisms such as cognitive tunnel vision, and misdiagnosis.

Although diagnostic strategies have shown some utility as a concept during the studies, they may be at too gross a level for more detailed insights. Two alternative avenues for research are to identify operator tactics used during diagnosis, and to consider the depth of knowledge utilised and required for solving individual diagnostic problems. Both of these approaches may be more likely to yield tangible means of improving training for diagnosis.

Four strategies were identified in the studies, both in single operator and team-based scenarios. These strategies appear therefore to be meaningful constructs in diagnosis. However, their relationship to diagnostic success and failure is not clear. It may be possible that these strategies are described at too high a level of abstraction to be useful for performance prediction.

Analysis of operator support systems focused on training, procedures, and the interface and alarm systems. With respect to training, useful tactics were gleaned from real operators' behaviour, and may be provided to other operators via training.

Procedures supported diagnosis but did not preclude the need for it. Their ability to help operators recover from misdiagnosis or diagnostic failure could be also improved. It was therefore recommended that the functionality of procedures be re-considered to include specifically diagnostic recovery as a procedural objective.

The interface appeared to have both positive and negative impacts on diagnostic performance. Some of the negative

impacts were simple "usability" problems. An extensive ergonomics evaluation of the interface should be carried out to remove further usability problems before trying to provide specific diagnostic support systems.

The major problem areas with alarms were the design and use of alarm lists, and the audible alarm under extended alarm conditions. However, the basic advantages of having a chronological list and an audible alerting system are likely to remain. Certain key alarms appear to have insufficient salience. There may be scope for changing or enhancing their status to improve diagnostic performance. There may well be other scenarios and other associated key alarms which are "under-emphasised" in the alarm system interface.

The prospect of multiple faults and misleading / missing indications is probably the most challenging type of scenario for operators and the interface and operator support system designers. The flexibility of an operating team in response to such events, and hence their optimum mode of working together as a team, needs to be better understood before adding in procedure tracking systems.

ACKNOWLEDGEMENTS

The insights reported here are based on the work performed by the HEAP team at the OECD Halden Reactor Project. The dedication and efforts of all the team members, as well as of the operators who served as experimental subjects, are gratefully acknowledged.

REFERENCES

Caro, P. W. (1988). Flight training and simulation. In E. L. Wiener & D. C. Nagel (Eds). *Human factors in aviation*. San Diego: Academic press.

Follesø, K., Drøivoldsmo, A., Kaarstad, M., Collier, S. & Kirwan, B. (1995). *Human error - the third pilot study* (HWR-430). Halden, Norway: OECD Halden Reactor Project

Kirwan, B. (1994). *Human error project experimental programme* (HWR-378) Halden, Norway: OECD Halden Reactor Project.

Kaarstad, M., Follesø, K., Collier, S., Hauland, G. & Kirwan, B. (1995). *Human error - the second pilot study* (HWR-421). Halden, Norway: OECD Halden Reactor Project.

Kaarstad, M., Kirwan, B., Follesø, K., Endestad, T. & Torralba, B. (1994). *Human error - the first pilot study* (HWR-417). Halden, Norway: OECD Halden Reactor Project.

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

Reusing processes and documenting processes: toward an integrated framework

Françoise Détienne
Ergonomic Psychology Group
INRIA-Rocquencourt
Domaine de Voluceau,
BP 105, Rocquencourt
78153 Le Chesnay cedex
France
+33 1 39635522
Francoise.Detienne@inria.fr

Jean-François Rouet
Language and Communication Laboratory
95 Avenue du Recteur Pineau
86022 Poitiers cedex
France
+ 33 49885976
rouet@hermes.univ-poitiers.fr

Jean-Marie Burkhardt
Ergonomic Psychology Group
INRIA-Rocquencourt
Domaine de Voluceau,
BP 105, Rocquencourt
78153 Le Chesnay cedex
France
+33 1 39635196
Jean-Marie.Burkhardt@inria.fr

Catherine Deleuze-Dordron
INRIA-Grenoble
46 avenue Félix Viallet
38031 Grenoble cedex
France
Catherine.Deleuze-Dordron@imag.fr

ABSTRACT
This paper presents a cognitive typology of reuse processes, and a cognitive typology of documenting processes. Empirical studies on design with reuse and on software documenting provide evidence for a generalized cognitive model. First, these studies emphasize the cyclical nature of design: cycles of planning, writing and revising occur. Second, natural language documentation follows the hierarchy of cognitive entities manipulated during design. Similarly software reuse involves exploiting various types of knowledge depending on the phase of design in which reuse is involved. We suggest that these observations can be explained based on cognitive models of text processing: the van Dijk and Kintsch (1983) model of text comprehension, and the Hayes and Flower (1980) model of text production. Based on our generalized cognitive model, we suggest a framework for documenting reusable components.

Keywords
software design, software reuse, software documenting, text processing, situation model, textbase, problem solving, revising

BACKGROUND AND OBJECTIVES
Empirical studies have pointed out some critical aspects of the software design situation. Software design problems are known to be ill-structured (Guindon, 1990; Visser & Hoc, 1990). Specifications are lacking and must be inferred. There is a multiplicity of acceptable solutions for a design problem, and a multiplicity of constraints and criteria of evaluation for the solutions. No solution is optimal according to all criteria. Software design involves at least two knowledge domains: the problem domain and the programming domain.

One strategy to cope with this situation is to lean on previous designs. Empirical studies report that software design is rarely done from scratch. Designers reuse information about previous designs, both from external sources or internal sources (memory). An attempt to systematize and increase the efficiency of this strategy has been implemented in the software engineering approach referred to as "software reuse". In the literature on software reuse, reuse is often considered only as far as some external representation of a source is reused, for example by copying/modifying source code or using a specialisable component.

In this paper we will take a broader view of the reuse activity not limited to the external reuse view. In our view reuse refers to all situations in which information about an old solution is exploited during a new software development. Thus exploiting an old solution may refer to the reuse of an external representation of this solution but also to the use of related information inferred based on a source component even if this component is not actually reused. Furthermore software reuse may occur in any phases of software design. Sofware development is usually described as composed of three main phases: problem understanding, design, and coding. We will argue that exploiting an old solution may occur in any of these phases.

The software reuse issue is intrinsically linked to the documentation issue. Documenting programs properly is essential to support reuse. Our objective is to construct a generalized cognitive model which accounts for both reuse and documentation processes. Based on this model, we will suggest a framework for documenting reusable components.

SOFTWARE REUSE AND DOCUMENTING: STATE-OF-THE-ART
Software Reuse is currently one of the most active and creative research areas in Computer Science. This is mainly because software quality and productivity are assumed to be greatly increased by maximising the (re)use of (part of) prior design products instead of repeatedly designing from

scratch (Gall, Jazayeri & Klösch, 1995; Krueger, 1989). Supporting the reuse activity has become a big challenge in Software Engineering. The "large" definition of reuse which has been recently adopted by the software engineering community is:

"Software reuse is the reapplication of a variety of kind of knowledge about one system to another similar system in order to reduce the effort of development and maintenance of that other system. This reused knowledge includes artifacts such as domain knowledge, development experience, design decisions, architectural structures, requirements, design, code, documentation and so forth" (Biggerstaff & Perlis, 1989, p XV).

Even with this large definition, in practice, methods and techniques are implemented at the programming language level or code level and they aim to make code reuse safer. The main issue is to supply designers with "components" (e.g., code, specialisable components, class reuse by inheritance, specifications) and mechanisms to use these components into the current design. But in any cases, whatever the type of reusable components, reuse relies mostly on the existing code[1].

Currently, the predicted level of reuse has not been reached yet, due to technical, organisational and ergonomic factors (Biffl & Grechenig, 1993; Tracz, 1987). No actual integration of reuse into process models and design methodologies has been proposed yet. Tools and environments are designed without taking reuse into account. Furthermore a common belief is that a "well-designed" component does not need to be provided with any further information. On the contrary, in practice, the need for software documentation is often emphasized as a critical factor in software reuse. One main issue is to properly document software components in order to support the reuse activity.

We believe that the main problem in software reuse is that it relies on the existing code. In this way, Green et al. (1992) insist on the fact that if the code is the essential element of the knowledge of the programmers, it "does not express all the knowledge that the programmer has or may wish to employ". These authors advocated the addition of a separate "description level" decoupled from the code; this decription level would be at the programmer's disposal for recording arbitrary attributes and relationships in a browsable form. This description level would be a kind of knowledge-base; it would record both transient facts and long-term facts about program components and program relationships.

However, as concerns software documenting, there is no agreed-upon methodology for documenting computer programs, and even less a general theory of how documenting relates to software design. Empirical studies (e.g., Rouet et al. 1995a) have shown that professional software designers may have different representations of what documenting is good for, even though they agree on general documentation guidelines (e.g., comments should support difficult algorithms). We have started to investigate the activity of software documentation with the general hypothesis that documenting is a natural component of skilled design. We claim that this approach is supported by evidence from a wide range of situations, which include documenting completed programs and documenting while designing a program.

A GENERALIZED COGNITIVE MODEL

Empirical studies on design with reuse and on software documenting provide evidence for a generalized cognitive model. First, studies emphasize the cyclical nature of design: cycles of planning, writing and revising occur. Second, natural language documentation follows the hierarchy of cognitive entities manipulated during design. Similarly software reuse involves exploiting various kinds of knowledge depending on the phase of design in which reuse is involved.

We suggest that these observations can be explained based on the cognitive models of text processing: the van Dijk and Kintsch (1983) model of text comprehension, and the Hayes and Flower (1980) model of text production. These models account for (a) the hierarchy of entities manipulated during design-as-comprehension (textbase versus situation model) and (b) the cyclical sequence of planning-translation-revision in text production. In the two following sections we present a cognitive approach of documenting and of reuse. Results from empirical studies will be discussed in the text processing theoretical framework.

A cognitive approach to documenting

In our perspective, documenting is an essential component of the design process. We claim that documenting serves two essential purposes: to assist the designer during problem-solving, and to improve the design product by making the program easier to communicate (i.e., easier to understand for another designer). Studies of program documentation have evidenced that designers output several types of comments. Some comments have to do with the surface structure of the program or textbase (e.g., paraphrases and syntactic explanations), others have to do with the problem being solved through the program or situation model (e.g., goal structure, justifications of design decisions). One dimension is of interest when analyzing the relation between documentation and software design: whether the designer is documenting an already completed program, or documenting as part of design. Empirical studies have shown interesting differences between those two situations.

Documenting completed programs

The task of documenting a program produced by someone else has been studied by Rouet et al. (1995b). The analysis of comments elicited several categories of information: paraphrases (comments that paraphrase program statements and do not include any new information), syntactic explanations, semantic explanations (about solutions being implemented), meta-comments (statements about commenting) and inferences from labels. When com-

1. An alternative approach, which seems promising, is based on analogical reasoning in the reuse of software specification (Lung et al. 1995; Maiden, 1991).

menting simple procedures or simple programs, both intermediates and experts issued mostly semantic explanations (in this case, mostly low level functional information), then paraphrases. A general hypothesis was that comments reflect the designer's cognitive representation of the entity being commented. The results suggest that the representation constructed reflects (1) low level functional information close to elementary operations and (2) control flow information (paraphrases). This suggests that programmers have constructed a textbase representation. We refer here to the distinction between the textbase and the situation (or mental model) made in the van Dijk and Kintsch's model of text understanding (1983). The textbase represents what is said in the text and how it is said whereas the situation (or mental) model represents the situation referred to by the text. The structure of the textbase is isomorphic or homomorphic to the text structure whereas the structure of the situation model is not. The situation model is constructed on the basis of the textbase and inferences made by activating generic or episodic knowledge.

Another result from Rouet et al.'s study was that "structural" units, e.g., beginning of loops, were the most frequently commented. This suggests that programmers have constructed a textbase representation since the structure of the representation constructed reflects the structure of the program text (as defined by the control structure). To sum up, in the experimental task of documenting someone else's program, programmers tend to reason at the level of the textbase, i.e., they do not question the rationale for implementing a particular solution.

The same kind of results was found in another study on documentation (Riecken et al. 1991). Expert programmers documenting a program produced by someone else generated more comments detailing given instructions explicitly stated in the code rather than general domain information associated with the task. Subjects generated nearly twice as many detailed comments as abstract comments. These results suggest again that programmers in this kind of documentation task construct a textbase representation rather than a situation model. Furthermore, it was found that subjects located vertical spacing according to the program text structure, e.g., between routines. This last result also supports the textbase construction hypothesis since the constructed representation preserves the text structure.

The study of designers documenting someone else's program indicate that they may focus on a shallow level of representation: the program textbase, i.e. the surface and propositional structure.

Documentation as part of design

Software designers usually output a lot of natural language information as they design programs: design notes, temporary comments, or comments that are meant to remain with the final program (Davies, 1996; Henry et al. 1992). Even though the final goal of the activity is to write a list of code statements, natural language information is part of the designer's solution.

From this point of view, there is an interesting parallel between writing a program and writing a natural language text. One of the most influential cognitive model of text production is the one proposed by Hayes and Flower (1980). Hayes and Flower have defined three major phases in the writing process: Planning of the text structure as a function of domain knowledge (organizing) and communication purposes (goal setting); translating the text plan into a linguistic representation; and reviewing the text as a function of the writer's evaluation. One important feature of their model is that the overall process is cyclical rather than strictly linear. Moreover, the ordering of phases is not strictly predefined. Instead, the organization of writing phases is a function of the writer's strategy (Flower & Hayes, 1981). As a result, the writer generates a lot of intermediate results, e.g., planning notes, temporary text, additions and so forth. The writer also revises his or her drafts in order to improve clarity, coherence and/or to make the points stronger (Bereiter et al. 1988). Even though a first draft may be understandable, the polished product is usually better suited to the writer's purposes.

Software design also includes phases of planning, translation and revision (usually called problem solving or design, coding, revising). Gray and Anderson (1987) showed that such cycles occur in software development. Revising processes lead to changes characterised as stylistic, strategic and tactical. Stylistic changes involve revising *coding* whereas tactical and strategic changes involve revising *planning*.

Planning involves both retrieving problem-relevant knowledge and building up an abstract solution. Translating is equivalent to implementing the solution in a particular language. Finally, revising may include either modifying the implementation, the abstract solution, or even one's understanding of the problem structure. Some of the language generated during design will be later modified, some will be added, and some will be removed. Due to the particular purpose of computer programs -- to be used by both humans and machines -- the final product includes both natural and computer language. Following this approach, the status of documentation is not different from the code being generated. Instead, both documentation and code may be seen as a natural outcome of skilled design.

The empirical study of design with reuse provides good evidence that documenting is part of the design process (Rouet et al. 1995a). When asked to design with reuse of high-level components during the object-oriented design of a complex application, designers produce comments at all levels of abstractions: from the justification of high-level design decisions, to the description of implementation details. Moreover, designers have trouble narrowing the scope of their comments while designing. When asked to write only a particular category of information in a natural language field, they will often depart from the assigned category, and mix up different categories of information. We believe that this behavior reflects the cyclical-interactive nature of the design process. Both kinds of informations are tightly connected in the designers' representation, not as a function of their syntactic or semantic status, but as a function of the sub-

problem being solved at the moment. Some comments serve as an external memory, and they are meant to be deleted later (e.g., consequences of a design decision on another design decision to be made later on). Others -- e.g., justifications for particular solving decisions -- will stay as comprehension aids for the human reader. It is likely that the latter category contain information which connect the program surface structure (or "textbase") to the underlying solution structure (or "situation model"). However, the evidence for what comments are really helpful in program comprehension is still scarce.

A cognitive approach to reuse

Empirical work on software design with reuse has been conducted in the last years. Two contradictory results emerge from the literature. These studies tend to show, either that the effect of the reuse processes is an enrichment of the representations constructed during design (Burkhardt & Détienne, 1995a; Rosson & Carroll, 1993), e.g., by the inference of new constraints, new goals, or that the effect of the reuse processes is the lowering of the level of control of the activity during design, in particular by the use of test/debug strategies and comprehension avoidance strategies (Lange & Moher, 1989; Détienne, 1991). We defend the idea (Burkhardt & Détienne, 1995b; Détienne, 1996) that the former type of results concern reuse processes involved during the analysis and problem solving phases (which we call "reuse in planning") whereas the latter type of results concern reuse processes involved during the implementation phase (which we call "reuse in coding"). Our interpretation is that reuse in planning involves the construction of a situation model of the source whereas reuse in coding involves, but not necessarily, the construction of a textbase representation of the source.

Reuse in planning

The effect of the reuse processes may be an enrichment of the representations constructed during planning. Burkhardt and Détienne (1995a) show that evoking a reusable component may allow the addition of constraints, and new goals. Rosson and Carroll (1993) note that sometimes the borrowed code is not directly reusable itself but rather is used more as a functional specification. In another domain, architectural design, De Vries (1993) found that exploiting examples of old designs may allow the inference of new constraints for the new design and allow the constraints to be envisaged at a more abstract levels.

Several studies show that when a source component is evoked or retrieved during planning (as opposed to coding), information about the source situation from which the component comes from is searched or inferred. Burkhardt and Détienne (1995a) observed that this allowed programmers to infer solution goal structure, constraints, evaluation criteria or design rationales. Rosson and Carroll (1993) note the importance of knowledge about "example application" of a reusable class. In their field study, Rouet et al. (1995a) found that when selecting a reusable component in a library designers were looking for information on the application from which the component was extracted. This contextual information which seems to be highly important is rarely present in the documentation of components because software engineers generally believe that reusable components must be generic and application-independent.

In all these situations, it seems that reusing a component implies more than constructing a textbase representation of the source component itself. It implies constructing a situation model of the source application. In design with reuse, this situation model of the source allows the representation constructed for solving the problem on hand to be enriched and the search space to be enlarged.

Reuse in coding

Reuse results in the lowering of the level of control of the activity during the implementation phase. We refer to the hierarchy of levels of control developed by Rasmussen and Lind (1982). These authors distinguish between automatic activities, activities based on rules, and activities which involve high-level knowledge. The lowering of the level of control of the activity consists in switching from activities which involve high-level knowledge, e.g. problem solving activities, to activities based on rules and automatic activities, e.g. execution of procedures.

The use of the copy/edit style attests this effect. It reflects comprehension avoidance of the copied code and use of surface-level features to construct a representation of it[2] (Lange & Moher, 1989; Rosson & Carroll, 1993). The designers make "probable" modifications and rely heavily on the debugging tools to evaluate the code.

The analysis of what we called "new code reuse" episodes (Détienne, 1991) also shows that the level of control of the activity lowers. In these situations, the designer anticipates the reuse of a component in the same program while developing it. In new code reuse designers construct an operative representation of the source as well as a procedure for modifying the code of the source into targets. Designers then execute this procedure to develop targets. In this case also, there is a lowering of the level of control of the activity, which causes errors by propagation of source errors or by omission of changes. It was shown that the programmers constructed a representation of the source at various levels of abstraction. At the highest level, it represented the function of the source component whereas at the lowest level it represented the surface structure of the component, i.e. elementary operations. This suggests that a textbase representation of the source component is constructed.

IMPLICATIONS

We believe that both our studies on design with reuse and our studies on software documentation provide evidence for a generalized cognitive model, in which software design is seen as a particular instance of text processing. Based on this model we suggest a framework for documenting reusable components.

---

2. These observations have been made in the object-oriented design. It is worth noting that the copy/edit style does not conform to the style of reuse encouraged for OOP, i.e., reuse by inheritance.

Our generalized cognitive model is based on the cognitive models of text processing: the van Dijk and Kintsch (1983) model of text comprehension, and the Hayes and Flower (1980) model of text production. Our model highlights two main aspects. The first aspect is that the nature of the design activity is cyclical. As in text production, the three main cognitive processes of planning, coding and revising occur in a cyclical way. Furthermore, depending on the process involved, the nature of knowledge manipulated varies. As in text processing, the nature of the representation constructed may be either a textbase representation or a situation model.

The second aspect of our model is that there is a parallelism between the nature of information reused depending of the design stage and the nature of documentation produced in various design stages. Studies of design with reuse have shown that reuse may occur during planning or coding. Reuse in planning involves the construction of a situation model of the source whereas reuse in coding involves, but not necessarily, the construction of a textbase representation of the source. Similarly, studies of documentation have indicated that documentation outputs -- e.g., inserted comments -- follow the hierarchy of entities manipulated during design: from abstract solution schemas or design rationales to low-level code statements.

Our contention that (a) design unfolds as a cyclical activity and (b) natural language documentation follows the hierarchy of cognitive entities manipulated during design has a potential implication for the design of design environments.

It has been shown that designers produce comments at all levels of abstractions: from justification of high-level design decisions, to the description of implementation details. Similarly software reuse involves exploiting various types of knowledge depending on the phase of design in which reuse is involved. We argue that a theoretical framework for documenting reusable components should be based both on our typology of reuse processes and on our typology of documenting processes. Clearly, documentation on justification of high level decisions linked to one or several components would be mostly useful when reuse is involved during planning (problem analysis and problem solving) whereas documentation on implementation details of one component would be useful when reuse is involved during coding. But we argue that both types of information are useful depending on the design phase.

The use of free annotations (Green et al. 1992) could be a way to implement this approach. We believe that the tools provided to store and edit documents should not assign a particular status to the information being stored. Chances are that during the design phases the designer will have to store information at various levels at the same location. On the other hand, the tools should include some type of a documentation checker, so that the designer can trim documentation after the design is completed, and remove the pieces of information that are no longer relevant.

Finally, we should point out a limitation of this approach. It is common to think that a design has a rationale or that the design space, i.e., the space of possible designs, can be analyzed by designers. In our approach, this information could be belong to the documentation of a reusable component. However a limitation of this approach on documentation, which also applies more generally to the design rationale approach, has been discussed recently by Karsenty (1996). This author made an empirical evaluation of design rationale documents. He found that designers having to evaluate a previous design for reusing it asked design rationales questions not answered by the design rationales documentation provided with the previous design. These kinds of questions, not handled by the designers of the source, could not be avoided and in fact it is likely that the number of this class of questions will increase as the time goes by and designer's culture changes. Karsenty speculates that long term reuse of design rationales should be more problematic that short-term reuse.

ACKNOWLEDGEMENTS

This research was partially sponsored by the European Esprit III SCALE 6334 project (System Composition and Large Grain Component Reuse Support). Thanks to Denis Alamargot for his insightful comments on the parallel between software design and text production.

REFERENCES

Bereiter, C. Burtis, P.J. & Scardamalia, M. (1988). Cognitive operations in constructing main points in written composition. *Journal of Memory and Language, 27,* 261-278.

Biggerstaff, T. J., & Perlis, A. J. (Ed.). (1989). *Software reusability : concepts and Models.* New York: New York ACM press.

Biffl, S., & Grechenig, T. (1993). Degrees of consciousness for reuse of software in practice: Maintainability, balance, standardization. In *Proceedings of the Seventeenth Annual International Computer Software and Applications Conference,* 1-5 Nov. 1993, Phoenix, AZ, USA: IEEE Comput. Soc. Press. 107-114.

Burkhardt, J-M. & Détienne, F. (1995a) An empirical study of software reuse by experts in object-oriented design. In K. Nordby, P. H. Helmersen, D. J. Gilmore and S. A. Arnesen (Eds): *Proceedings of INTERACT'95.* Chapman & Hall. 133-138.

Burkhardt, J-M. & Détienne, F. (1995b) La réutilisation de solutions en conception de programmes informatiques. *Psychologie Française, 40-1,* 85-98.

Davies, S. (1996) Display-based problem solving strategies in computer programming. In W.D. Gray and D.A. Boehm-Davis (Eds): *Empirical studies of programmers, sixth workshop.* Ablex Publishing Corporation, 59-76.

Détienne, F. (1991) Reasoning from a schema and from an analog in software code reuse. In J. Koenemann-Belliveau, T.G. Moher and S.P. Robertson (Eds): *Empirical studies of programmers, Fourth Workshop.* Ablex, Norwood, NJ. 5-22.

Détienne, F. (1996) *La conception et réutilisation de logiciels: l'approche de l'Ergonomie Cognitive.* Research report 2902, INRIA, Rocquencourt.

De Vries, E. (1993). The role of case-based reasoning in architectural design : Stretching the design problem space. In W. Visser (Ed.): *proceedings of the Workshop of the Thirteenth International Joint Conference on Artificial Intelligence "Reuse of designs : an interdisciplinary cognitive approach".* Chambery August 29, 1993: INRIA Rocquencourt. B1-B13.

Flower, L. & Hayes, J.R. (1981). Plans that guide the composing process. In C. Frederiksen and J. Dominic (Eds.): *Writing: the nature, development and teaching of written communication. Vol. 2: writing: process, development and communication.* Hillsdale, NJ: Erlbaum. 39-58.

Gall, H., Jazayeri, M., & Klösch, R. (1995). Research directions in software reuse: where to go from here? *SIGSOFT Software Engineering Notes* (spec. issue. on SSR'95: Symposium on Software Reusability), 225-8.

Gray, W. D. & Anderson, J. R. (1987) Change-episode in coding: when and how programmers change their code? In G. M. Olson, S. Sheppard and E. Soloway (Eds): *Empirical studies if programmers, second workshop.* Ablex. 185-197.

Green, T. R. G., Gilmore, D. J., Blumenthal, B. B., Davies, S., and Winder, R. (1992) Towards a Cognitive browser for OOPS. I*nternational Journal of Human-Computer Interaction, 4*, 1-34.

Guindon, R. (1990) Knowledge exploited by experts during software system design. *International Journal of Man-Machine Studies, 33(3)*, 279-304.

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

Henry, D., Green, T. R. G., Gilmore, D. & Davies, S. (1992) Improving the communicability of spreadsheet designs: annotating with descriptive tags. In F. Détienne (Ed): *proceedings of the fifth workshop of the psychology of programming interest group.* Paris, December 10-12, 1992. 91-93.

Karsenty, L. (1996) An empirical evaluation of design rationale documents. In M.J. Tauber, V. Belloti, R. Jeffries, J.D. Mackinlay and J. Nielsen (Eds): *CHI'96 conference proceedings*, Addison Wesley. 150-156.

Krueger, C. W. (1989) *Models of reuse in software engineering.* Report CS-89-188. Carnegie Mellon University.

Lange, B.M. and Moher T.G. (1989) Some strategies of reuse in an object-oriented programming environment. In K. Bice and C. Lewis (Eds.): *Proceedings of CHI'89 Conference on Human Factors in Computing Systems.* ACM Press. 69-73.

Lung, C.-H., & Urban, J. E. (1995). An expanded view of domain modeling for software analogy. In *Proceedings. Nineteenth Annual International Computer Software and Applications Conference (COMPSAC'95)*, 9-11 Aug. 1995, Dallas, TX, USA: IEEE Comput. Soc. Press. 77-82.

Maiden, N. (1991) Analogy as a paradigm for specification reuse. *Software Engineering Journal*, 3-15.

Rasmussen, J., & Lind, M. (1982) *A model of human decision making in complex systems and its use for design of sytem control strategies.* Roskilde, Danemark, RISO, M-2349.

Riecken, R. D., Koenemann-Belliveau, J. & Robertson, S. P. (1991) What do expert programmers communicate by means of descriptive commenting? In J. Koenemann-Belliveau, T. G. Moher and S. P. Robtertson (Eds): *Empirical studies of Programmers, fourth workshop.* Ablex. 177-193.

Rosson, M.B. and Carroll, J.M. (1993) Active programming strategies in reuse. *Proceedings of ECOOP'93*, Object-Oriented Programming. Berlin: Springer-Verlag. 4-18 .

Rouet, J-F., Deleuze-Dordron, C., & Bisseret, A. (1995a) Documentation as part of design: exploratory field studies. In K. Nordby, P. H. Helmersen, D. J. Gilmore and S. A. Arnesen (Eds): *Proceedings of INTERACT'95.* Chapman & Hall. 213-216.

Rouet, J-F., Deleuze-Dordron, C. & Bisseret, A. (1995b) Documentation skills in novice and expert programmers: an empirical comparison. *Proceedings of the seventh workshop of the Psychology of Programming Interest Group (PPIG)*, Edinburgh, Scotland, 4-6 January 1995.

Tracz, W. (1987). Software reuse : Motivators and inhibitors. In *Compcon'87, 32nd IEEE Computer Society International Conference*, San Francisco. 358-363.

van Dijk, T.A. and Kintsch, W. (1983) *Strategies of Discourse Comprehension.* New York: Academic Press.

Visser, W. & Hoc, J-M. (1990) Expert software design strategies. In J-M. Hoc, T.R.G. Green, R. Samurçay and D.J. Gilmore (Eds): *Psychology of programming.* Academic Press. 235-250.

The Power of Habit in the Failure to Diagnose Unusual Faults

J. Patrick
School of Psychology
PO Box 901
UWCC
Cardiff. CF1 3YG
+44 1222 874 007
patrickj@cardiff.ac.uk

Operators in process control develop habits or routines that are, by definition, adaptive and normally support the fast and effective diagnosis of most faults. However some challenging faults are difficult because they require that operators break these habits in order for diagnosis to be successful. Evidence that operators' inability to make shifts from "normal" assumptions and interpretations is a major source of difficulty comes from various reports of performance during real incidents and field studies of cognitive activities during simulated events (e.g. Reinartz, 1991). This evidence will be discussed bearing in mind that there is no criterion available that is able to differentiate this source of difficulty from others unambiguously.

A field study, in process control, was designed to examine operators' ability to deviate from habitual assumptions in order to diagnose unusual faults. Faults were designed that challenged two assumptions. One assumption concerned an aspect of the nature of the fault, namely whether it had a single or multiple cause. The second concerned the belief that signals faithfully represent plant state i.e. instrument readings are valid indicators of plant variables. Two nonroutine fault scenarios were developed, one requiring both assumptions to be broken in order for diagnosis to be successful. One routine fault was included for comparative purposes. In total five senior operations personnel participated in the study and each tackled two faults on a high fidelity simulator.

Process tracing methodology (e.g. Woods, 1992) was devised and applied successfully to analyse operators' reasoning during both routine and nonroutine scenarios using video recordings and concurrent verbalisations. This provided a fine-grained analysis of the interaction between operators' strategies, including hypotheses, the developing fault scenario and their reactions to it. Modified problem behaviour graphs were developed after Newell and Simon (1972) and Bereiter and Miller (1989).

Faults that violated routine assumptions were indeed problematic as reflected in both quantitative and qualitative measures of performance. Nonroutine faults were not diagnosed successfully. Qualitative analyses suggested that operators were unable to break routine assumptions in the generation of hypotheses and in the interpretation of instrument readings. There was a strong tendency for operators to believe their instruments even when they had a correct hypothesis based on corroborated evidence that was highly inconsistent with an invalid instrument reading. Unexpectedly multiple faults that did not include instrument failures were not easier to diagnose even though operators only had to shift with respect to one routine assumption. These findings are discussed in the light of other sources of difficulty when attempting to find faults in complex industrial processes.

References

Bereiter, S.R. and Miller, S.M. (1989). A field based study of troubleshooting in computer-controlled manufacturing systems. *IEEE Transactions on Systems, Man and Cybernetics*, 19, 205-219.

Newell, A. and Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Reinartz, S.J. (1991). A multi-faceted examination of how teams cope with complex nuclear power plant incidents. In Duncan, K.D. and Patrick, J. (Eds), *Proceedings of the Third European Conference on Cognitive Science Approaches to Process Control*. Cardiff, UK: School of Psychology, UWCC.

Woods, D.D. (1992). Process tracing methods for the study of cognition outside of the experimental psychology laboratory. In Klein, G., Calderwood, R. and Orasanu, J. (Eds), *Decision making in action: Models and methods*. Ablex.

Adapting models to cases : a study of collective problem solving

Catherine Sauvagnac & Pierre Falzon
Laboratoire d'Ergonomie
CNAM, 41 rue Gay-Lussac, 75005 Paris
falzon@cnam.fr

Medical decision is not always the result of an individual process of analysis. In complex, unusual and risky situations, decisions are taken by groups of highly trained specialists who meet in order to reach a common understanding of the case and to decide on a therapeutic procedure. We are studying such a situation of collective decision making with two objectives. The first one is to reach a better comprehension of the way in which these groups work in order to optimize their functioning. The second objective deals with knowledge elaboration. Beyond case-specific therapeutic decisions, do practitioners build new general knowledge while processing successive cases ? How could this knowledge elaboration process be assisted ?

Most of the time, therapeutic decision in oncology makes use of pre-established protocols : they are rules set up by domain experts and validated by statistical studies. Protocols specify their inclusion criteria, related to tumour features and patient features. Decision is then made by examining the inclusion criteria in one or another protocol. In reality, protocols do not specify all of the therapeutic aspects : physicians have to tailor the protocol to the case.

The cases which are debated in the decision committee are those which cannot be solved using these protocols, either because of the cancer disease itself or because of specific characteristics of the patient. The committee meets once a week. It includes senior specialists in each necessary discipline (radiotherapy, oncology, surgery, hormonotherapy). The committee size varies from 7 to 15 participants. The number of cases considered in each meeting varies, from 1 to 7 cases.

The elaboration of a common knowledge space

During the meeting, each case dealt with is brought in by a long statement concerning patient features, disease story and features that make this case a particular problem. This statement is made by the physician who treats the patient. All along the statement, the other participants ask for more explanations.

Not everything is explicitly said. First, the presentation of the case provides a filtered view of what is known about the case : only the information believed to be unknown and necessary for decision making is provided. Second, much medical knowledge is assumed to be shared by the participants. This general knowledge is implicit and compose the mutual knowledge space. However, it may appear explicitly, especially when justifying a previous assertion.

Decision-making

Classification criteria are the object of the exposure of the case and also the object of the requests of explanation. A first aspect of decision-making thus concerns the elaboration of a description of the case on which everyone agrees. This description is realized through an assessment of the differences between the case and the protocols. A difficult issue is that the boundary of a protocol may appear fuzzy when applied to a specific case.

As far as therapeutic actions are concerned, protocols may leave room for discussion. First, recommendations may be conditional or at the discretion of the team. Advices of the participants help to build a more comprehensive list of the possible side-effects of the therapeutic means. Second, the treatment recommended by the protocol may not be applicable to the case for some particular reasons (another ongoing treatment, psychological factors, etc.).

The striking observation is that, although the cases that are discussed are supposed to be outside the domain of application of the protocols, the decision process is structured by the protocols. Cases are "solved" by differentiation from -or aggregation to- a protocol.

The traces of the decision committee activity

After each meeting, a member of the decision committee (always the same one) writes a report (10 to 30 lines per case), which is then stored in a "decision committee" folder.

The report appears as a selective and biased trace of what was actually discussed. Some elements disappear from the report (justifications of decisions, rejected therapeutic solutions, pieces of knowledge that were provided in the course of the discussion of a case). Inversely, some elements of the patient's file that were not mentioned in the committee's discussion are mentioned in the report.

Thus, the report is not an accurate trace of the work of the committee : the lack of justifications and rejected solutions makes it impossible to trace the reasons of the decision.

From a cognitive standpoint, our goal is two-fold. On one hand, we wish to facilitate the elaboration of a relevant collective memory of successive cases. Traces should be kept of justifications, alternatives, rejections of solutions, etc.

and of the general knowledge that is necessary to understand the decision. Case-based reasoning and model-based reasoning should be associated. On another hand, the objective cannot be limited to the construction of a data base of cases : physicians should be encouraged to engage in an abstraction task. The processing of a succession of similar (but not identical) cases makes it possible to elaborate semantic knowledge from episodic representations. The conjunction of repetetivity and variability is a powerful means for abstraction. We wish to assist the physicians in this abstraction task, for instance by providing them a methodology for the collective analysis and pairing of cases.

The adaptation of the protocols, necessary to process uunusual cases, has two results : the therapeutic decision and the elaboration of new knowledge about protocols, pententially usable in future cases. Functional and meta-functional activities thus appear as tightly intermingled.

The issue of ground/cockpit integration: results from field studies

A.-L. Amat and A. Bellorini
European Commission, Joint Research Centre
Institute for Systems, Informatics and Safety
21 020 Ispra (VA), Italy
+39 332 78 51 37
anne-laure.amat@jrc.it, anna.bellorini@jrc.it

ABSTRACT
A pilot activity analysis has been performed to study the relation between the cognitive behaviour of an expert aircraft pilot and the worksystem. The cognitive demand of the pilot's task is related to the level of automation (or mode) of the supporting tools. The mode choice stems both from a personal "flying style" and worksystem constraints. One of the most relevant constraints is the air traffic control operator's requests. A study on air traffic control operator stress shows that this demand actually comes from flow management and is propagated from the ground to the cockpit. Thus, our research, based on natural setting observations, confirms that operators, and designers, from both ground and cockpit sides, should better cooperate and consider that there is one and only one worksystem.

Keywords
pilot task, air traffic control, activity analysis, worksystem, cognitive demand, stress, use of automation.

INTRODUCTION
An increasing number of accidents caused by human factors has reinforced the demand for research in the field of cognitive behaviour in complex systems. In the aviation domain, a kind of human factors-related accident which is particularly disastrous—68% of 1995 fatalities (Learmont, 1996)—is controlled flight into terrain, where a perfectly operating aircraft, under the control of the crew, is flown into terrain (or water) with no prior awareness on the part of the crew of the impending disaster (Wiener, 1977). In these cases, the crew has lost the awareness of what is going on and realises it too late or not at all. In other words, the problem stems from the fact that a situation, judged as safe or not judged at all, should actually not have been tolerated. Thus, human error and accident issues have to be tackled not only by studying human behaviour in emergency but also by analysing normal activity. Our contribution can be placed within this last perspective: it describes pilot cognitive activity in a normal flight context, in cases both of good and poor performance.

The flight context can be seen as the worksystem status. The worksystem gathers the components that help the pilots in their task but also the elements that might raise problems. The mission of the commercial pilot is to fly from point A to point B respecting safety, economy, schedule, and comfort criteria. During their task performance, pilots are supported by paper documentation—maps, procedures, check lists. They listen to radio bulletins, giving weather and airport information, and communicate with other human agents of the worksystem: crew members, cabin crew, passengers, ground people and air traffic control operators (ATCOs). They can also use highly automated instruments such as the flight management system (FMS) that provides them elaborated displays describing past, present and future situations and can perform "for them" navigation, performance calculations and systems control subtasks. Thus, pilots can choose between worksystem sources of information and control modes in order to counteract difficulties stemming from variations in meteorological conditions, route or profile modifications or even system failure. In modern "glass" cockpits, most of the worksystem-pilot relations can be apprehended by looking at the interaction with the FMS[1].

In the first part of this paper, we will present the main results of the pilot activity analysis, describing the variation in the cognitive activity in relation with the worksystem status. We will show how air traffic control (ATC) requests are demanding for the pilot since they increase cognitive demand and require great flexibility in the use of flight instruments. In the second part of this paper, we will present the ATC point of view that actually enlarges the perception of the same problem: we will show, using results from a field study at an Italian airport, how certain types of ATC clearances that increase the pilot's workload, actually stem from strategies aiming at reducing the ATCOs workload.

PILOT'S COGNITIVE BEHAVIOUR
As previously mentioned, cognitive activity is closely related to worksystem status. Moreover, the description of behaviour combines both a representation of the competence used by the pilot and a specification of the general cognitive mechanisms supporting task performance. Thus, we considered as a prerequisite for our study that a complete description of pilot cognitive activity must fulfil, at least, three requirements:

1. it must be related to a *cognitive architecture* supporting the general mechanisms governing human cognition that are relevant in the case of an expert pilot. So it

1. We focused on the case of the Boeing 767.

must cope with situation awareness and decision making success and failure.

2. it must provide a list of the worksystem events (or *scenarios*) that affect the pilot's cognitive activity, i.e. that modify the cognitive demand of the situation.

3. it must comprehend the *competences* used by the pilot during his activity (knowledge on instrument use, rules of thumb, procedures...) and express them in a format meaningful for the cognitive architecture.

A knowledge engineering methodology (Guida and Tasso, 1994) has been followed in order to comply with this prerequisite on cognitive activity description. It consists of two main steps: sampling and extensive knowledge elicitation.

The first step, sampling knowledge engineering, led to a general view of pilot activity, matching the first requirement. It has been performed by synthesising the background supplied by existing models such as COSIMO (Cacciabue et al., 1992) and COCOM (Hollnagel, 1993) and the work carried out to develop of a Crew Resource Management Course (Amat et al., 1995; Kjaer-Hansen, 1995). This view or theory about pilot activity has led to the first version of a cognitive architecture (1st requirement) and to the definition of an initial format for describing the competences used by the pilot.

The second step, extensive knowledge elicitation, is being performed through pilot interviews and observations in the cockpit. We could thus identify the elements modifying cognitive demand (2nd requirement) and grasp the competence used for performing the task (3rd requirement). A flight can be seen as a sequence of flight phases. During this second step, we focused on the descent phase, where the cognitive activity is very relevant in terms of both planning and monitoring and where the complexity of the situation is still manageable for an external observant. Starting by following the hierarchical and goal/means task analysis philosophy (Sébillote, 1991), the activity has been described as a "task units" graph, each task unit associating the strategies performed and the goals aimed at by the pilot.

Our study on pilot cognitive activity provides three kinds of information, complying with the three requirements previously developed. It also encompasses the relationship between those aspects of the activity. We have focused, here, on the relation between pilot activity (requirements 1 and 3) and the worksystem context. Since the loss of situation awareness often seems to be connected to an inadequate cognitive demand or workload (Tenney et al., 1992; Sarter, 1991), we have chosen to examine the relation in light of cognitive demand variations. This approach has some links with the one described in Woods and Hollnagel (1987); however, our objective is not to provide a domain independent cognitive description but to stick to the case of pilots in normal conditions.

According to the theory of our cognitive architecture, pilot task performance can be analysed in terms of four main cognitive activities: (1) strategic planning, (2) tactical planning, (3) control and (4) monitoring. We have presented hereafter some results of our pilot study, giving examples of these four activities and specifying the related cognitive demand.

(1) Strategic planning

Strategic planning copes with goal management. Indeed, it regards both the elaboration of a plan—i.e. building or, more often, retrieving a goal breakdown— and the capability of adapting this plan in case of disturbance, i.e. of questioning the current goal, dealing with competing goals.

As regards the descent phase, most of the strategic planning is made during descent preparation. This "sub-phase" aims at preparing aircraft systems and pilot for the "ideal"[2] descent. The ideal profile and its related performance calculations are done by the flight management computer, on the basis of data given by the pilot and stemming from the choice of approach procedure. Thus, the choice of approach procedure and the identification of control points—points indicating a change in the aircraft system or target and summarising the constraints of descent—are fundamental for the descent execution. They also provide elements for performing the briefing, a kind of mental simulation of the approach phase. The approach procedure is chosen by considering aircraft and airport facilities, meteorological conditions and by following some personal safety trade-offs. So the cognitive demand of this task comes from the variety of parameters and their interdependency. In reality, experience and synchronisation constraints among subtasks give structure to this task: the choice, in routine cases, can be made by following a sequence of tests. Furthermore, the choice is also facilitated by the information available that allows the reduction of the number of alternatives.

Thus, the cognitive demand basically depends on the availability and reliability of information and on the familiarity with airport approaches procedures.

The descent plan might have to be modified during the execution of the descent: a descent path or profile could be modified to avoid turbulence for instance, or to comply with ATC requirements. The case of ATC interventions might simply correspond to the introduction of a new goal in the previous plan. But sometimes, the requests can be such that the current plan has to be replaced by another: see the case presented in the ATCO section, where the pilot of the A2 aircraft heard the ATCO telling him "A2 climb to Flight Level 110 (FL110)" while he was descending and approaching the airport.

The descent planned during the preparatory phase corresponds to the theoretical one (ideal descent). Sometimes, the experienced pilot who knows the habits of the ATCOs, and can predict the modifications that could be made later on in the descent path, can elaborate another descent plan,

---

2. the ideal descent plan optimises the tradeoff between fuel consumption, respect of schedule, passenger comfort and safety principles.

a back-up plan, not directly supported by cockpit tools. This is done to avoid the cognitive demand related to the unexpected change of plans and might allow choosing a tactic facilitating a switch from one plan to another: for instance, when arriving at the northern Italian airport we focused on, a pilot will more easily accept being late with regard to the planned path if he knows that the chance of cutting a planned curve is great (cutting a curve might allow him to get back on schedule). This double planning has thus some advantages in the sense that it allows better anticipation of changes; however, the drawback is related to the overconfidence bias: a very experienced captain with a moderate attraction for highly automated system could neglect the preparatory phase of the descent, saying "anyway, they never let us follow the standard procedure, so we just have to wait to be taken in charge and directed by radar vectoring".

Since we focused on cases of normal activity where information is available, modifications of the plans are seen as the main source of cognitive demand. The cognitive demand stemming from strategic planning could be evaluated by considering the number of interrupted goals and the frequency of the goals triggered.

(2) Tactical planning
Tactical planning deals with strategy management. It consists of choosing a strategy for current goal achievement.

During descent preparation, the cognitive demand related to this level of planning is quite low since only a few strategies could reach the same goal, or since there is a perfect match between the current context and conditions related to the first strategy that comes into mind. During descent execution, the effort related to tactical planning should also be quite low, if it is be performed by following the plan. But, very often in European countries, ATC intervenes during the flight, and modifies the descent profile. We saw that this could induce strategic replanning, but it also provokes a reconsideration of the current strategy. Indeed, typically, it might question the current navigation mode. Vertical navigation can be performed using three main modes, corresponding to different levels or style of automation, with different time responses, and allowing regulating different parameters. In case of a request such as "descend to FL140 at a rate of 1500ft/min", the pilot would disengage the highest automated function "vertical navigation" and engage the vertical speed mode, since it allows the vertical speed to be maintained by regulating the thrust. These transitions from one mode to another, or mode "usage" (Degani et al., 1995) are the manifestation of the pilot's flying style. In our study, we considered one style and, based on it, we studied the cognitive demand. We did not consider that tactical planning was, in itself, cognitively demanding since the effort made for choosing the mode is very context driven. However, the consequences of this switching of modes (or strategies) might be demanding since control of the system and monitoring activities depend on the mode. In our example, after having switched to vertical speed mode, the pilot has to monitor that the aircraft is not close to overspeed and perform a control action if that is the case.

Thus, we did not consider that strategic planning was, in itself, cognitively demanding; however frequent strategic changes might be significant since they induce changes in controlling and monitoring activities.

(3) Control
We call "control of the system" actions to be performed when the need of intervening on the system process has been identified, mostly during monitoring activity. Since we deal with a highly automated cockpit, we have not considered manual control; we have coped with actions such as entering a value in the FMS, setting an instrument, modifying a target value on the autopilot and the cognitive activity related to this.

The cognitive demand related to these control actions varies according to the effort needed for determining the value of the parameter to set. During the execution of the descent, most of the target values are directly given by the ATC ("descend to FL180", then the target flight level becomes 18000ft) or established by simple calculations ("descend to FL100 in 4min": when at FL180, this means loosing 8000ft in 4 min or flying at a vertical speed of 2000ft/min for 4 min—FL100 =10000ft and FL180=18000ft with a standard pressure setting on the altimeter.) The control actions that allow sticking to the programmed profile or staying in the current mode (in other words, remaining in the same strategy) are not very demanding either, since, most of the time, control is made by a trial and error process based on the experience and knowledge of the dynamics of aircraft systems: for instance, in vertical navigation mode—the highest level of automation— a control action could be to add "a little" speedbrakes or spoilers when the message "DRAG REQUIRED" appears on the screen or, in flight level change mode, a medium level of automation, it could be to increase the speed "a bit" because, during the monitoring activity, the need has been identified. It is important to note that these subjective values are closely related to the result of the monitoring activity that gives an idea of how much the speed, for instance, has to be increased. Furthermore, monitoring is also important in the trial and error process to check that the modification of the parameter was right. During planning of the descent, a control action might require more demanding calculations: for instance, the evaluation of the reference speed is made by considering current speed, time before landing, and wind.

Thus the cognitive demand stemming from the control activity is related to the number of parameters taken into account in the rule determining the value to set, and to the number of actions to perform.

(4) Monitoring
Monitoring activity aims at maintaining situation awareness. This ranges from perception, through identification, to the interpretation of a parameter status. Monitoring activity is directed by the current goal, or, more precisely, by the strategy that specifies relevant parameters, provides some expectations with regard to system evolution and defines the current safety envelope.

The effort for monitoring the situation depends on the flight context (approach to control point, unexpected event...) and on the mode used. At the highest level of automation (selected for instance during the planning of the descent), the demand is mostly at the level of attention since FMS interfaces allow a quick check of the profile. At medium level of automation, such as flight level change, the anticipation of future status is not obtained directly by looking at instruments. It needs a more elaborated treatment of the information that might require the consideration of parameters not relevant in other modes. For instance, when checking if the actual altitude is acceptable while descending, pilots might monitor the difference between the current altitude and the one obtained by using a rule-of-thumb which transforms the distance to the airport into a prescribed altitude.

Thus, indicators of monitoring demand could be the "distribution of the relevance" among the parameters (our idea of relevance is close to the notion of cognitive and physical salience presented in Cacciabue et al., (1992)) and the "complexity" of the calculations to do to realise the implications of the values of the parameters.

We have presented a view of pilot cognitive behaviour, focusing on four types of complementary activities. To complete this description, the dynamic dimension of the task has to be stressed. Indeed, it plays a determining role in the evaluation of the cognitive demand, even if it is difficult to underpin with indicators. Furthermore, we noted that the temporal horizon is divided into segments that are actually defined by a backward chaining scheduling of the operation which is also at the basis of goal subdivision. The temporal pressure is thus relative within these segments and can be evaluated by looking at the execution conditions inherent to the strategy used.

The description of the nature of the cognitive demands encountered by the pilot provides a qualitative view of cognitive demand. For switching from this qualitative view to a more quantitative one, the notion of cognitive profile has to be introduced. Indeed, the evaluation of cognitive demand is subjective: each individual has his/her own rating scale and gives a different weight to each of these indicators. By writing some functions of these indicators, some rules indicating the personal priorities, one can trace the variation of the cognitive demand. However, we have chosen here to use common sense assumptions regarding the respective weight and measurement of the indicators in order to be able to generalise our findings and affirm that some scenarios are more demanding than others.

In routine cases, the highest cognitive demand is encountered during the execution of the descent when the pilot has to change his plan several times. Indeed, common policy is to use the highest level of automation, checking that the programmed profile is followed. When the profile has to be changed, switches to lower levels have to be made and a strategy set up to return to the most automated, or to intercept the planned trajectory, hence an increase in cognitive demand. Furthermore, medium levels of automation are more flexible to use, have a quicker time response but are also more demanding cognitively, mainly because of the monitoring and control activities that have to be continuously done. A variation in the meteorological conditions might induce such a transition towards a lower level of automation. More frequently, changes are caused by ATC intervention. Moreover, in case of step-let-down descent, these changes are not only unexpected but also repeated: the ATCO first gives a clearance for a flight level, the pilot responds to this "order" by adapting his original plan, then other clearances are given, while the pilot does not know what the next clearance will be. Thus, in these cases, often found in Europe, the pilot has at his disposal a set of highly automated functions, programs them and prepares himself/herself for their use, but is constrained to use a medium level of automation because of unforeseeable profile changes provoked by the ATCOs.

So, from this point of view, it seems that the improvements made to pilot worksystem instruments are, in operational context, limited by the ATCO problem to cope with the situation. In order to analyse more precisely the causes of pilot's cognitive demand variation, we will now give a description of the ATCO worksystem and activities.

FEATURES OF ATCO ACTIVITY

One important category of perturbation that has considerable impact on the reliability of the predicted flight profile is related to the human actions of the ATCO. At any instant, the ATCO may change the route, cruise level and airspeed as a function of the perceived overall traffic situation.

In this part of the paper we will focus on the discrepancy between the structure of the pilot's task, guided by the automatic tools provided in the cockpit, and the constraints stemming from ATC interventions. These "annoying changes" from the pilot's point of view are actually only the visible part of the larger problem of flow management.

The objective of ATC is flow management according to three criteria: (1) optimisation of flight safety in order to avoid possible conflicts between aircraft and transgression of minimal separation rules; (2) flight economy to minimise fuel consumption; (3) fluidity of traffic flow in order to maintain scheduled flights.

The air traffic control task is a continuous task since the controller regularly verifies that each aircraft follows the right trajectory. It is composed of sub-tasks: radar screen monitoring; management of communication between pilots, tower controllers, controllers of adjacent sectors through different means—radio, telephone, intercom; strip management; monitoring of weather conditions at controlled airports.

When ATCOs detect a possible conflict, they solve it by modifying level, heading and speed with respect to safety, economy and fluidity of traffic flow. They have to constantly verify the evolution of the density of the traffic they are in charge of, and to plan the distribution of the traffic over time.

Different conditions can challenge the task performance of ATCOs and increase their level of workload and stress.

Stress for the ATCO seems to be a joint effect of high attention required and high level of responsibility as measured by the cost of an error (Warr, 1987).

As a result of a field study in an Italian airport, where the activities of the ATCOs were examined at the approach control working position, the following conditions affecting ATCO cognitive demand have been identified (Bellorini, 1996): geographical allocation of airways; traffic conditions in terms of complexity, volume (number of aircraft in charge) and number of holding aircraft; airport capacity and support; procedures and rules; communication density and possible overlapping; appearance of unexpected situations or conflicts.

In order to cope with the increase in cognitive demand related to these conditions, ATCOs apply safety-centred strategies based on task sharing and workload distribution.

They decompose the general problem into localised subproblems, giving priority to safety rather than flight economy and fluidity of traffic flow and treating the problems sequentially, one by one. The number of parameters taken into account decreases and strategies become economical in terms of the cognitive effort of the ATCOs.

Another observed economical solution is the application of individual strategies and actions, successfully tested, to avoid the main conflict. These actions are generally directed to the aircraft which the ATCO knows the flight history (previous heading, speed ...).

The following example shows how ATCOs have managed a conflict between four aircraft (Bellorini and Vanderhaegen, 1995). This case involves two ATCOs of the approach control working position (C1 and C2), controlling two specific geographical sectors, and the pilots of four aircraft (A1 to A4), landing at the same airport. C2 is in charge of A1 and A4; C1 is in charge of A2 and A3. C1 receives A2 at only 10 NM before the "gate" from an ATCO of an adjacent sector. A2 is at the same level as A3, that is, in holding pattern on the gate. In order to maintain minimal separation[3], C1 decides, first of all, to increase the level of A2, in order to avoid a conflict of level with A3, and, secondly, to give an heading to A2 so that it avoids the gate.

We can see, in this case, how the solution applied by the ATCO, on one hand, reduces workload and stress by solving the salient problem: safety is ensured by minimal separation, even if the increase in level is not economical and does not help traffic fluidity. But, on the other hand, as explained in the section on pilot activity, this solution increases pilot cognitive demand. Thus, the increase of the pilot's mental load can be seen as an effect of the ATCOs workload transfer.

CONCLUSION

In normal conditions, the workload of the pilot and the ATCO varies and can reach high levels. Moreover, there is often a link between these increases of workload. Mostly during the descent and approach phases, the ATCO workload can be transmitted to the pilot. At this stage of the flight, an increase in cognitive demand, is not the most welcome, since it can damage the pilot's situation awareness and might contribute to a controlled flight into terrain accident.

The current aviation worksystem allows this workload propagation. Nor does it guarantee an optimum co-operation. What happens, in practice, is that there is no integration between the pilot and ATCO worksystems, which have been considered as two separate entities. The existing mismatch between cockpit automation and ATCOs support tools shows this clearly (Gras et al., 1994).

Furthermore, an important way to cope with workload propagation in distributed systems is to share knowledge about the respective tasks (Hutchins and Klausen, 1992) and to efficiently support the agents of the system. A contribution towards better integration of the two worksystems could be made through training programs, which includes a presentation of the task, and the means available to the other agents within the system. But fundamental changes will come by modifying the approach taken while developing support tools and considering the air traffic management (ATM) problem as the common denominator.

To maintain safety as well as fluidity of traffic flow, integration between the two worksystems–ATC and cockpit–has to be supported. Thus, further research towards the development of automatic systems supporting the ATCOs and towards a better transfer of information to the cockpit has to be undertaken. The Communications, Navigation and Surveillance (CNS) and Air Traffic Management (ATM) concepts developed by the International Civil Aviation Administration (ICAO, 1994) point in that direction. NASA and FAA, along the same lines, have been field-testing a new generation of decision support software for ATCOs (Descent Advisor, (Erzberger, 1992)), that also reduces delays and increases fuel efficiency, thus complying with pilot criteria.

3. minimum distance allowed between two aircraft

REFERENCES

Amat, A.-L., Bacchi, M., Cacciabue, P.C., Kjaer-Hansen, J., Larsen, N.M., Lie, A. M., Ralli, M. (1995), The Ispra Crew Resource Management Course - Facilitator Preparation Manual, *ISEI Technical note* I.95.48. ISEI/IE/2926/95. May 1995.

Bellorini, A (1996), Task-related factors of stress: the analysis of conflict resolution in air traffic control, *Proc. 1st International Conference on Applied Ergonomics (ICAE)*, May 21st-24th 1996, Istanbul, Turkey.

Bellorini, A., Vanderhaegen, F. (1995), Communication and cooperation analysis in Air Traffic Control, *Proc. 8th International Symposium on Aviation Psychology*, April 24-27, Columbus, Ohio, USA.

Cacciabue, P.C., Decortis, F., Drozdowicz, B., Masson, M., Nordvik, J.P. (1992), COSIMO: A cognitive simulation model of human decision making and behaviour in accident management of complex plants, *IEEE*

Transactions on Systems, Man and Cybernetics, 22 (5), 1058-1071.

Degani, A., Shafto, M., Kirlik, A. (1995), Mode usage in automated cockpits: some initial observations, *in* T.B Sheridan, *Proc. International Federation of Automatic Control; Man Machine Systems (IFAC-MMS) Conference*, Boston, MA. IFAC.

Erzberger, H. (1992), *CTAS: Computer intelligence for air traffic control in the terminal area.* NASA-TM-103959.

Gras, A., Moricot, C., Poirot-Delpech, S.L.,Scardigli, V. (1994), *Faced with automation: the pilot, the controller and the engineer*, Publications de la Sorbonne, Paris.

Guida, G., Tasso, T. (1994), *Design and development of knowledge based systems*, John Wiley and Sons, Chichester.

Hollnagel, E. (1993), *Human reliability analysis- Context and control-*, Academic Press, London, UK.

Hutchins, E., & Klausen, T. (1992), Distributed cognition in an airline cockpit, *in* Middleton, D. & Engestrom, Y. (eds), *Communication and cognition at work*, Cambridge University Press.

ICAO, (1994), Human Factors Digest N. 11, *Human Factors in CNS/ATM Systems*, Circular 249.

Kjaer-Hansen, J. (1995), Strategy and method for adapted human factors training: the JRC-CRM experience, *Proc. 8th International Symposium on Aviation Psychology*, April 24-27, Columbus, Ohio, USA.

Learmont, D. (1996), Off target, *Flight international*, 17-23 January.

Sarter, N.B. (1991), Situation Awareness: a critical but ill defined phenomenon, *International Journal of Aviation Psychology*, 1(1), 45-47.

Sébillotte, S. (1991), Décrire les tâches selon les objectifs de l'opérateur: de l'interview à la formalisation, *Le Travail Humain*, 54, (3).

Tenney, Y.J., Adams, M. J., Pwe, R.W., Huggins, A.W.F., Rogers, W.H. (1992), A principled approach to the measurement of situation awareness in commercial aviation, *NASA contractor report 4451*.

Warr, P.B. (1987), *Work, Unemployment and Mental Health*, Oxford: Oxford University Press.

Woods, D., Hollnagel, E., (1987), Mapping cognitive demands in complex problem-solving worlds, *International Journal of Man-Machine Studies*, 26, 257-275.

Wiener, E.L. (1997), Controlled flight into terrain accidents: system-induced errors, *Human factors*, 19 (2), 171-181.

Multi-level Human-Machine Cooperation in Air Traffic Control : an experimental evaluation

Marie-Pierre Lemoine & Jean-Michel Hoc
Université de Valenciennes et du Hainaut-Cambrésis
Laboratoire d'Automatique et de Mécanique Industrielles et Humaines
Phone : +33-27141393 - Fax : +33-27141294
e-mail : {lemoine, hoc}@univ-valenciennes.fr

ABSTRACT
Results of an experiment done to evaluate dynamic task allocation between two air traffic controllers — a radar controller and a planning controller — and assistance tools are reported. Three experimental conditions are compared with a method of identification of controllers' cognitive activities from thinking aloud protocols: (a) a control condition without assistance, (b) an 'explicit' condition where the controllers allocates the tasks, and (c) an 'assisted explicit' condition where task allocation is done by the computer system with a possibility of counter-order by the planning controller. The assistance tools, especially in the assisted explicit condition, appeared to reinforce the strategic aspect of the activity and to improve human-human cooperation. The need for elaborating an overall representation of the traffic seemed to be stronger in the explicit condition than in the assisted explicit condition.

Keywords
air traffic control, human-machine cooperation, dynamic task allocation, cognition

PROBLEM
Dynamic task allocation between human operator and assistance tools is one of the more important problems of automation and computerisation in industry and transport, but now it also concerns a vast range of dynamic situations, for example, anaesthesia and intensive care. The reliability of human-machine systems, which allows human operators to remain in control of situations, includes multidisciplinary research themes, and the examination of real human-machine cooperation. The dynamic task allocation theme is not new and probably gets to the heart of the human-machine cooperation problem in dynamic situations, where the control of the situation is partial and requires a continual adaptation to changing and unexpected conditions (Debernard, Vanderhaegen, & Millot, 1992; Millot, 1988; Millot, Asse, Marrakchi, & Willaeys, 1985; Rieger & Greenstein, 1982).

In most studies, the evaluation of different dynamic task allocation principles is based on final indicators of performance (for example, fuel consumption). This type of objective indicator is important to underline differences, but is absolutely insufficient to understand human-machine cooperation mechanisms which explain and allow the generalisation of results to other situations. Subjective evaluation of operators can also be too global and introduce contradictions with performances (Debernard *et al.*, 1992). A clear evaluation, which tries to understand the cognitive mechanisms of human-machine cooperation, has to identify these mechanisms.

If the cognitive activities of a computer program are directly accessible by the examination of the explicit instructions introduced by the designer, it is not the same with the human operator. The observer must consider that human operator cognitive activity has two parts: an explicit part, directly observable, the behaviour, and an implicit part, only attainable by interpretation and determining the behaviour structure (Piaget, 1947). The change from behaviourist psychology to cognitive psychology brings many theoretical and methodological reflections to understand implicit activities, especially by the means of different methods of verbalisation (Bainbridge, 1979; Caverni, Bastien, Mendelsohn, & Tiberghien, 1988; Ericsson & Simon, 1984; Leplat & Hoc, 1981).

Cognitive activity coding has still never been systematised in a general manner because it is necessarily based on inferences from verbal and non-verbal protocols, under the control of a theoretical model of these activities. Thus such a frame-model, at once sufficiently general and precise to perform such a task has never been formulated. So we have to use models that are specific to some categories of situation. The importance of the frame-model, even if it is specific, must be underlined, especially to analyse verbal protocols. Linguistics and psycholinguistics have developed many methods of verbal protocol coding. But they have a limited relevance when we do not want to code the verbal protocol but what it allows one to infer about implicit activities. As such, verbal protocol is a behaviour among others and cannot be, anymore than the others, more than an indicator of an object, unattainable without inference.

In this short talk, we emphasise the presentation of the main results, so we cannot present in detail the frame-model and the method. The reference used to code the diagnostic and decision-making activities is the conceptual model of dynamic situation management proposed by Hoc and Amalberti (1995). This model comes from research on diagnosis, both in experimental psychology and in cognitive engineering, and it is mainly new because of its deviation from traditional serial models (e.g., Rasmussen, 1986). It also pushes forward the strong link between diagnosis —

considered as the continual updating of the current representation of the situation — and action goals. Our approach to human-human and human-machine cooperation comes from a new way of thinking, developed by cognitive psychology, but also social and engineering sciences (Hoc, 1996). According to this conception, two agents of a system are in a situation of cooperation if: (a) each of them strives towards goals although interfering with the goals of the other, and (b) they try to manage such interference to make easier the activities of the other.

In the next section we present the design of our experiment. Next, we describe shortly the protocol coding scheme, and we report the main results obtained with this method and other tools.

EXPERIMENTAL DESIGN

For this research we have built an experimental platform — SPECTRA V2 — which simulates a real control position (Lemoine, Debernard, & Millot, 1995). It integrates two work places, the former for the radar controller (RC), and the latter for the planning controller (PC). Each controller works with a radar and an electronic strip screen, and in addition with assistance tools (cf. Fig. 1). A strip (a cell of the array) describes the flight plan of each plane entering in the controlled sector.

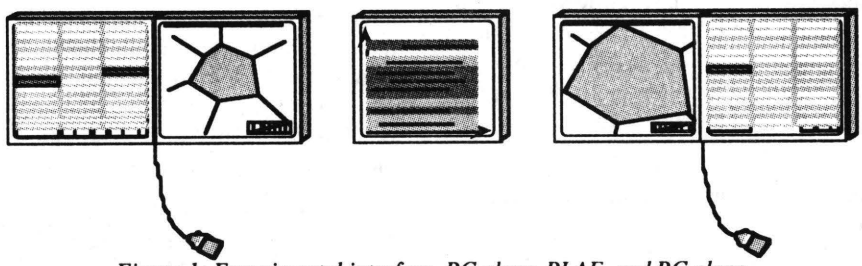


Figure 1: Experimental interface. PC place, PLAF, and RC place.

The controllers' tasks are different and complementary. PC has a strategic role because s/he ensures the aircrafts' entering and exiting while regulating RC's workload. RC fills a tactical role because he has to fulfil aircraft' flight plans and traffic safety by conflict resolution.

There are two types of assistance tools. The former — SAINTEX — assists aircraft guidance by conflict detection and resolution. It can detect all conflicts, and solve those between two stable planes (planes with no change of flight level) by deflecting one of them, and those between a stable plane and another one which has to climb or descend. The latter — PLAF (French acronym for Planning of Allocation) — is dedicated to PC and helps him/her to anticipate the influence of entering aircraft on the existing traffic, to measure the influence of the implementation of solutions from SAINTEX, on the RC's conflict resolution.

Three experimental conditions have been implemented to allow the evaluation of the assistance tools (cf. Fig. 2). The first condition — without aid (NA), is a control condition, and is similar to a real work place. The second condition — the 'explicit' condition (EX) — integrates the tactical assistance SAINTEX. The shareable conflicts (solvable by SAINTEX) are allocated by default to RC, and both controllers can make the allocation. The third condition — the 'assisted explicit' situation (AE) — has an additional assistance tool, in the explicit mode, that ensures the allocation of the task to SAINTEX if RC is overloaded. However, this decision can be changed by PC.

The experiment took place during three weeks in the Eastern En-Route Control Centre of France, with three pairs of experienced controllers. Three similar scenarios were built with samples of real traffic of the eastern sector of France airspace, and order effects were counterbalanced. The experiment began with a training stage using three other scenarios with light traffic and was used to describe all the features of the platform and the situations. The test scenarios were built with about sixty planes with different behaviours (landing, takeoff, flying over), and many shareable (6) and non-shareable conflicts (30). The duration of a scenario lasted about 1h30.

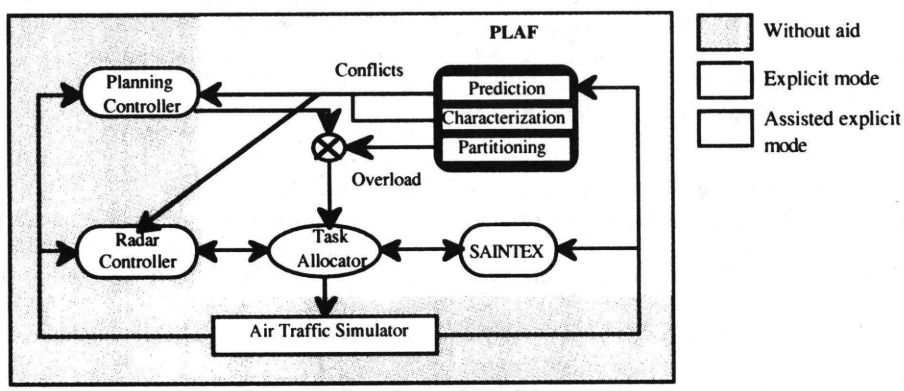


Figure 2: Experimental conditions.

The analysis is based on the study of task execution, subsequent self-observation obtained from each controller after each task execution recorded by a computer (traffic and actions) and a laser disk recorder (verbal reports).

METHOD

General principles of coding

For the coding and some analysis of protocols, we used the MacShapa software which brings editing, searching, and statistical functions (Sanderson *et al.*, 1994). This software is very useful for applying our methodology, without imposing a particular frame-model, and for creating a coding scheme. It has been developed to help to make exploratory analysis of sequential, verbal as well as non-verbal data.

Figure 3 presents the organisation of a MacShapa spreadsheet. It is composed of six columns, synchronised by the onset and offset time of the units. The units are defined on the basis of the verbal contents. The third and sixth column present protocols recorded during task execution: the spontaneous verbalisation of both controllers and the actions performed by controllers or by the system, with the principal events of the traffic. The subsequent self-observation reports of each controller fill the two first columns; and, the fourth and fifth columns present the manual encoding, with the predicate-attribute format, of each controller's activity.

The coding scheme used is common to the three last columns. The last column (actions on the traffic) and the third one (spontaneous reports) allow one to determine the activities to code, the predicates. These columns and the two first (subsequent reports) allow one to fix the attributes of the predicate, according to the context and the frame-model. The reproducibility of the coding is ensured by the convergence obtained by two coders. The two first columns (subsequent reports) are not coded, but they are used to control the inferences while coding the task execution protocol. While this method allows one to make precise some aspects of the activity, it does not ensure the exhaustiveness of the coding of this activity, because the basis is only the verbal reports. But we believe that operators verbalise the crucial part of their activity (under attentional control).

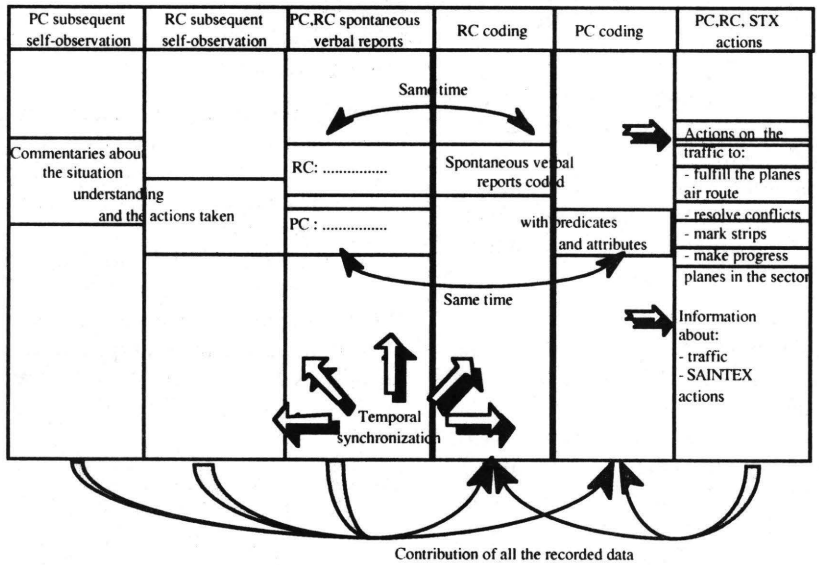


Figure 3 : Extract of a protocol coding with MacShapa

Apart from the events in the traffic, the coding scheme is decomposed into three levels (see Hoc & Lemoine, 1996) for a more extended description) : (a) activity class, (b) activity sub-class for a predicate, (c) attributes associated to a predicate (cf. Fig. 4). Among these attributes we find cross-references to others units of the protocol.They are only used to verify coding and they are not discussed in this paper.

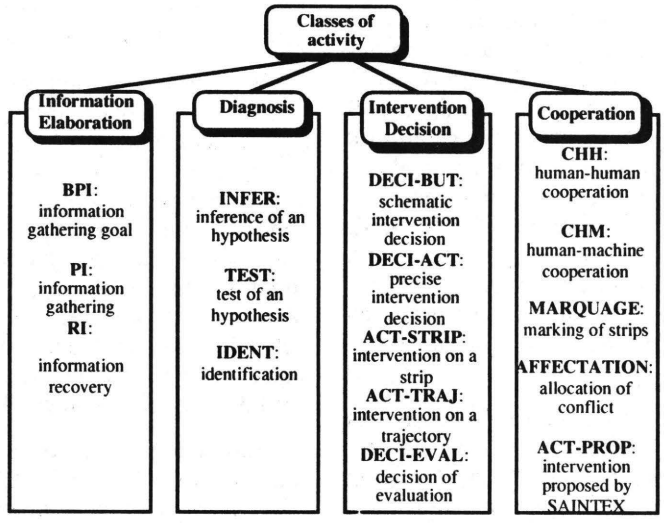


Figure 4: Predicate organisation for the activity coding

The diagnostic activities allow operators to build and to manage relevant interpretations for action. The difference between identification (IDENT) and inference (INFER) is the degree of certitude. Identification represents the transformation of raw data into interpreted data, categorised with certitude by the operator. Inference elaborates hypotheses with limited certitude; and test updates this degree of certitude.

Part of intervention decision activities is automatically recorded. These activities come from actions through interfaces to give orders to

pilots (ACT-TRAJ) or to change the plane state in the sector (entry, radar contact, exit of planes by ACT-STRIP actions). DECI-BUT represents schematic intervention decision which will be specified when the decision is effective (DECI-ACT). These activities underline the schematic character of planning in intervention decision-making.

Although cooperation activities are partly redundant with the other activities, their coding allow one to have a better idea of the way both controllers work together, and how they cooperate with the machine. And, especially, one can stress the particular position of PC who must translate information that comes from the machine, for the RC.

RESULTS

The spontaneous verbal report coding, the computer recording of all the air traffic control actions, the human-human and human-machine cooperation allowed us to analyse many aspects of controllers' activities and we present some essential results. These results come from the study of the three pairs of controllers in the three experimental conditions.

Performance

Some criteria allow us to evaluate the performance of the controllers. There are the number of air-misses (when the distance between two planes is under 8 nautical miles), the number of actions by plane, and the type of action used. The number of air-misses emphasises that the more the degree of automation increases, the better are the results (on the average, NA: 4.3; EX: 2.7; AE: 1.7). The other result comes from the study of the types of action used (heading for compass degrees, heading for a beacon, flight level) and the goals of action. We determine the goal by analysing the verbalisation about each action or the context in which the action takes place. The objective can be accurate route keeping, or conflict resolution. In comparison with NA and AE, EX shows proportionally more compass orders (NA: 23%; EX: 32%; AE: 22%), and so more resolution actions (NA: 47%; EX: 61%; AE: 54%). This is not a good result because a plane is penalised, according to fuel consumption or time of transit, if it often has to change its flight level or its trajectory. Moreover, we note proportionally more actions to respect the route in NA (NA: 53%; EX: 27%; AE: 36%), and more modification of flight level (NA: 41%; EX: 28%; AE: 34%) than in the conditions with assistance tool. An explanation could be that the controllers made less planning in the situation without assistance tool. They prefer to let the traffic evolve by itself, and to give intermediate flight levels, not always to avoid detected conflicts, but to be sure that a conflict will not appear. This principle allows the controller to supervise current conflicts and to wait before giving final decisions.

Subjective evaluation of workload

Every five minutes, we asked both controllers to give an estimation of the RC's workload, by choosing a value between one and seven. The highest value (seven) is chosen when RC is overloaded. The answers to this subjective scale show that PC has a good estimation of the RC's workload. S/he overestimates and anticipates it, but the trend between the response of both types of controller is similar. So, our hypothesis that the strategic level is able to control the task allocation is confirmed. The slope of the workload estimation graphs increases with the number of planes in the sector.

Cognitive and cooperative activities

All these results seem to point out the interest of the assisted explicit situation. But we have to go further in the analyse, particularly by using the verbal reports and their coding, to understand how the controllers work, and the effects of the assistance tools on their work.

A previous paragraph gave a short presentation of the coding principle, but to understand the results we obtain with this method, we must give some more details about some of the predicate arguments.

To study the information gathering we analyse:

- the type of object concerned by information gathering (flight level, plane, conflicting planes),
- the type of information: static, dynamic, alarm or action referred to,
- the type of condition: what activity requires information gathering.

When studying an intervention decision, we analyse the condition and the action type. Human-human cooperation as well as human-machine cooperation are analysed according to the type of cooperation and the means of cooperation.

The types of the cooperation are:

- goal: cooperation deals with the agents' goals: identifying the other agent's goal, facilitating the attainment of the other agent's goal, detecting and managing interferences between the goals of both agents;
- meta-cooperation: it deals with the definition of a communication code;
- common system of reference: cooperation deals with the representation of the context to share, to make the cooperation about the goals possible,
- role: the cooperation deals with role allocation between (human or artificial) agents.

The means of cooperation are:

- information supply or request,
- agreement or disagreement, within mutual control,
- confirmation or invalidation, without deep analysis,
- precision, when a message is reformulated more precisely.

The distributions (over predicates and attributes) of the results which come from the coding of the controllers' verbal reports points out two types of difference. We distinguish the difference between the condition with assistance tool (A: AE & EX) and that without assistance tool (NA), and the difference between the assisted explicit condition (AE) and the explicit condition (EX).

The condition with assistance tool and the condition without assistance tool (A vs. NA)

The main differences between conditions deal with individual cognitive activities and human-human cooperation.

Firstly, we note that the operators' behaviour is more anticipative than reactive in the situation with assistance tool. Activity is more anticipative (planned) and less reactive to events. Many indicators lead to this interpretation. In comparison with NA, with assistance tool (A), we find proportionally:

- less information gathering triggered by alarms, and more concern with conflict supervision (PI>condition>salient: A: 22%; NA: 29%);
- more intervention decisions about the organisation of the strips board (planning) and the condition is rather monitoring than alarms (DECI-ACT>action>classifying: A: 18%; NA: 11%).

The assistance tools thus seem to play a positive role by producing a planned activity. The reason would be the decrease of the workload, thanks to a part of the traffic given to SAINTEX, and an organisation around an earlier detection of conflicts.

Secondly, the assistance tools seem to make the human-human cooperation (HHC) easier:

- the cooperation frame is more operational (action-oriented) because it deals with goals and not with the building of a common system of reference (HHC>goal : A: 78%; NA: 68%. HHC>common system of reference: A: 13%; NA: 25%);
- when cooperation deals with goals, it is to facilitate the other controller's task, and not to detect and manage interferences between tasks (HHC>goal>facilitation : A: 84%; NA: 69%. HHC>goal>task interferences: A: 15%; NA: 19%);
- the radar controller appears more active, because we find more supplies and requests, and less confirmations (simple acknowledgements).

It seems therefore that the interfaces associated with the assistance tools create a common system of shared information which facilitates human-human cooperation.

The assisted explicit and the explicit condition (AE vs. EX).

The difference between AE and EX deals with the control of the overall traffic, planning, and human-human and human cooperation.

AE seems to present less difficulties in managing the entire traffic than EX. The activities of representation elaboration are less dominant: there is proportionally less diagnosis (DIAG: AE: 17%; EX: 22%); PC makes less decision with inferences as conditions (DECI-ACT>condition>inference AE: 21%; EX : 39%).

In AE, the objects of information gathering are proportionally more groups of planes in conflict (PI>object>planes: AE: 30%, EX: 24%), than individual planes (PI>object>plane: AE: 68%; EX: 71%). The assistance tools seems to focus the attention of the control on the conflicts. This is a positive result if we consider that this structure of representation of the traffic focused on conflicts is more operational.

AE reinforces the controllers' anticipative behaviour; planning appears through the actions on the strips and the decisions made on the basis of monitoring. Moreover, AE improves the human-human cooperation :

- it produces less explanation from PC to RC about the situation state,
- it is more operational and deals with the goals, decision-making and its evaluation,
- PC is more involved in the decision-making,
- information gathering is more closely linked to cooperation activities,
- but, there is less cooperation about inference activities.

It seems that the common space of shared information is more rich in AE than in EX, and the implication of PC in the task allocation leads him/her to make a better analysis of the situation.

Because RC cannot directly allocate or take back a conflict from SAINTEX, we note in AE that the human-machine cooperation (HMC) is reinforced. The main subjects of this cooperation are roles and allocations. We find proportionally more agreements, and less disagreements. So, we can think that in AE, there is a better acceptance of the assistance tools by the controllers than in EX.

The way in which controllers are involved in the task allocation is different according the experimental condition. When we analyse the goals and the means of the HMC, we note that PC is more involved in task allocation in AE than RC, whereas in EX, PC can control the task allocation as well as RC. We note that the inclusion of the controller (PC or RC) in task allocation changes human-machine cooperation by modifying :

- the goals, which become proportionally more the detection and management of interferences between goals, and less facilitation,
- the subjects of the cooperation, with more supplies of information and less confirmation of information.

CONCLUSION

The possibility to code the activities underlying the verbal reports and the non-verbal behaviours using precise predicates is already an important result. We can cover all the verbalised activity without neglecting any part of it. The synchronisation of the coding of the verbal reports and of the actions upon the traffic ensures the coherence of the coding and the evaluation of the importance of the verbal activity as regards to the global activity. The expertise in the air traffic control domain and the knowledge of the human cognitive behaviour allow us to evaluate the data which come from the frequencies given by MacShapa. The performance criteria of air traffic control, like the air-misses, the number of actions upon a plane, allow us to stress the relevance of the tactical and strategic assistance integration into the control position. But the analyses of the coding show more precisely the characteristics of the radar and planning controllers' activity which lead to better cooperation and to a better structuring of the global control activity. Thanks to the analysis of the information gathering activities, inferences, decision making and evalua-

tions, we see that the activity has a better structure in the assisted explicit situation.

The analysis of verbal reports points out three main results:

- the assistance tools reinforce the strategic aspect of the activity and improve the human-human cooperation,
- between the assisted explicit and the explicit situation, there is an inversion in the role of the radar and planning controller in the human-machine cooperation, and the assisted explicit situation induces a mental representation more centred on conflicts,
- the explicit situation underlines the necessity for the controllers to have a better control of the overall situation.

According to these results, therefore we can say that the assisted explicit mode, which really tests the new organisation, including a real sharing of the tasks between the radar controller and SAINTEX, seems to provide the best results.

ACKNOWLEDGEMENTS

This study was supported by the French research centre on air traffic control (CENA). We thank P. Millot for his scientific advice and S. Debernard, I. Crévits, and M.O. Bès for their collaboration.

REFERENCES

Bainbridge, L. (1979). Verbal reports from the process operator. *International Journal of Man-Machine Studies, 11,* 411-436.

Caverni, J.P., Bastien, C., Mendelsohn, P., & Tiberghien, G. (1988). *Psychologie cognitive: modèles et méthodes.* Grenoble, Presses Universitaires de Grenoble.

Debernard, S., Vanderhaegen, F. & Millot, P. (1992). *An experimental investigation of dynamic allocation of tasks between air traffic controller and A.I. system.* Paper presented at the *5th IFAC/IFIP/IFORS/IEA MMS,* Amsterdam.

Ericsson, K.A., & Simon, H.A. (1984). *Protocol analysis: verbal reports as data.* Cambridge, MA: MIT Press.

Hoc, J.M. (1996). *Supervision et contrôle de processus: la cognition en situation dynamique.* Grenoble, F: Presses Universitaires de Grenoble.

Hoc, J.M., & Amalberti, R. (1995). Diagnosis: some theoretical questions raised by applied research. *Current Psychology of Cognition, 14,* 73-100.

Hoc, J.M., & Lemoine M.P. (1996). *Identification des activités de diagnostic, de prise de décision et de coopération en situation dynamique : le cas du contrôle aérien.* Communication présentée à *ERGO'IA,* Biarritz, France, Oct.

Lemoine, M.P., Debernard, S., & Millot, P. (1995). Men-machines cooperation : toward a more cooperative assistance in air traffic control. *Proceedings of the IEEE 1995 International Conference on System, Man and Cybernetics* (pp. 1503-1508). New York: IEEE.

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

Millot, P. (1988). *Supervision des procédés automatisés et ergonomie.* Paris: Hermès.

Millot, P., Asse, A., Marrakchi, M., & Willaeys, D. (1985). *Approaches to man-machine cooperation in control posts of automated processes.* Paper presented at the *5th European Conference on Human Decision-Making and Manual Control.* Berlin, June.

Piaget, J. (1947). *La psychologie de l'intelligence.* Paris, Armand Colin.

Rasmussen, J. (1986). *Information processing and human-machine interaction.* Amsterdam: North-Holland.

Rieger, C.A., & Greenstein, J.S. (1982). The allocation of tasks between the human and computer in automated systems. *Proceedings of the IEEE 1982 International Conference on "Cybernetics and Society"* (pp. 204-208). New York: IEEE.

Sanderson, P., Scott, J., Johnson, T., Mainzer, J., Watanabe, L., & James, J. (1994). MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA). *International Journal of Human-Computer Studies, 41,* 633-681.

Cognitive Aspects of ATC: Experience from the CAER and PHARE Simulations

Christopher J. Kelly and Peter J. Goillau
ATC Systems Group
Defence Research Agency (DRA)
St. Andrews Road
Malvern WR14 3PS, UK
+44 1684 894684 +44 1684 894250
kelly@atc.dra.hmg.gb goillau@atc.dra.hmg.gb

ABSTRACT
This paper describes some cognitive aspects of air traffic control (ATC) in the context of two research projects carried out by the DRA over a two-year period, 1994-95: 'CAER' (Computer Assistance for En Route ATC) and 'PD1' (PHARE Demonstrator 1). Both projects were concerned with the effects of providing air traffic controllers with certain "computer assistance tools" in future ATC systems, and both involved sophisticated real-time simulation experiments. The paper discusses three cognitive issues: worksystem automation, the controller's 'picture', and mental workload.

Keywords
air traffic control, CAER, PHARE, PD1, controller's picture, automation, mental workload.

EXTENDED ABSTRACT
Over the past two years (1994-95) a number of real-time, air traffic control (ATC) simulation trials, investigating human-computer interaction (HCI) aspects of future ATC systems, have been carried out at the DRA in Malvern. The trials are the products of two particular projects: CAER (Computer Assistance for En Route ATC) and PHARE Demonstrator 1 (PD1), the latter a collaborative European project involving Eurocontrol and other key ATC research establishments. The main aim of the trials, for both projects, has been to assess the effects of the introduction of certain prototype "computer-assistance tools" which, as implied by the term, are intended to assist the air traffic controller at his/her work. In the case of CAER, the tools consist of a number of window-based alerting aids, in conjunction with an electronic ("paperless") flight strip data display, for possible implementation around the year 2000; in the case of PD1, the tools also consist of various conflict-alerting displays, but in combination with a highly-sophisticated "problem-solver", advanced planning functions, and a "stripless" radar data display, for possible introduction in the period 2010-2015 (Maignan, 1994).

Projects such as these, involving the design and evaluation of complex worksystems, provide a wealth of information on a gamut of system, HCI and cognitive ergonomics issues. This paper will provide firstly an overview of the two projects, and discuss their main aims and objectives. Second, the paper will focus on a number of cognitive issues. Three issues will be considered: the air traffic controller's 'picture', worksystem automation, and lastly techniques for the measurement of mental workload. Of course, these issues are not new, but the CAER and PD1 projects provide an opportunity to examine them afresh.

The air traffic controller's picture (i.e. his/her mental model of the current and predicted air traffic situation) has long been a topic of cognitive research (Whitfield and Jaclson, 1982). Much attention has been given to understanding the importance of the controller's (paper) flight progress strip in maintaining that picture, and particularly the effects of replacing and/or removing it with electronic representations. The future ATC system developed under CAER, featured fully electronic strips; the PD1 system had no strips (in the conventional sense) but instead had interactive aircraft labels. Some initial findings on the effect of these changes upon the controllers' mental picture, are presented.

Following on from this, the paper will consider the issue of automation. Both the CAER and PD1 systems are based upon providing the controller with various forms of computer assistance. Although it is generally well accepted that automation is a "good thing", it is also recognised that there is often too much "technology push" looking for problems to solve (a situation common to many other fields and not just ATC). From a cognitive standpoint, the problem remains that the provision of extra tools just gives the controller another job to do, including checking the tools (Goillau and Kelly, 1996). The paper will show that automation works best where it is properly integrated into the controller's task.

The main driving force behind today's developments in ATM systems is the need for further increases in capacity (i.e. handling more aircraft). Broadly speaking, the limiting factor is seen to be the service capacity provided by the controller, as opposed to physical airspace limits. This implies the need to reduce the controller's workload for each aircraft worked and use the "spare capacity" to increase the aircraft movements per controller (Stoner, 1995). This further implies the need to be able to measure the controller's workload. In the CAER and PD1 trials a number of techniques were employed: Instantaneous Self Assessment "ISA" (e.g. Jordan, 1992), NASA Task Load Index "TLX", and Malvern ATC Capacity Estimate "MACE" (Goillau and Kelly, 1996). The paper will present some initial results and observations on the use of these techniques.

REFERENCES

Cox, M. (1992). Report on the cognitive aspects of the air traffic control task. (Unpublished.)

Goillau, P.J. and Kelly, C.J. (1996). MAlvern Capacity Estimate (MACE): a proposed cognitive measure for complex systems. *Proc. First Int. Conf. on Engineering Psychology and Cognitive Ergonomics*. Avebury. (To be published.)

Jordan, C.S. (1992). Experimental study of the effect of an instantaneous self assessment workload recorded on task performance. (Unpublished.)

Maignan, G. (1994). What is PHARE?. *Air Traffic Control Quarterly* 2, 2, 79-83.

Stoner, C. (1995). Controllers as air traffic managers. In *Proc. of Global NAVCOM'95*, Montreal, 23-25 May.

Whitfield, D. and Jackson, A. (1982). The air traffic controller's picture as an example of a mental model; in G. Johannsen and J.E. Rijnsdorp, *Analysis, Design and Evaluation of Man-Machine Systems*, Proc. of IFAC Conf., Baden-Baden, Germany.

Supporting Continuous Learning

Reinhard Oppermann
HCI Research Division (FIT.MMK)
GMD - German National Research Center for Information Technology
D-53754 Sankt Augustin, Germany
Tel.: +49 2241 14 2703, Fax: +49 2241 14 2065
Reinhard.Oppermann@gmd.de

ABSTRACT
This paper deals with problems of learning in the working environment and introduces ideas of supporting the learning on demand process. With the system Lear (Learner's Living Repository), we propose a solution that supports users in exploiting and exchanging learning and consultation episodes: Users can describe problems or solutions when accomplishing tasks by memos or by registered screen actions, comment on them, and store them in a personal "demotheque". Users can send memos and registered screen actions that describe questions or problems as a request for off-line help to a consultant. Memos and registered screen actions that describe representative solutions can be made available to a group of users in a "purse for demos".

Keywords
learning and re-learning support, exploiting and exchanging learning and consultation episodes

INTRODUCTION
Today's working life and its widespread use of technology require more than ever to acquire permanently new domain and tool knowledge. As the user's task competence can dynamically be increased by a flexible work organisation and task support, the user's tool competence should dynamically be increased by systems suitable for learning, exploration facilities tolerant for correction, and support environments reinforcing recapitulation and re-evaluation of problems and solutions. We focus our view on the latter: to increase the tool competence of the user by strengthen the learning process and the reuse of already acquired knowledge in further working situations. Learning is ubiquitous, it has to be supported in every working situation, not only in particular learning phases or environments. Learning is a combination of exploration and instruction: people learn by trying things out and by asking other people for advice. Learning is a iterative phenomenon, it evolves step by step using early knowledge for later understanding.

THE SITUATION
Users' motivation to learn. We assume the learning process as being integrated into the task accomplishment (Dutke and Schönpflug, 1987, pp. 295f.; Paul, 1995, p. 168). A substantial part of learning does not happen during the training but during task performance. Users explore the system in use and try out functions to reach their goals. A "guided exploration" facility was proposed to support this kind of learning (Carroll et al., 1987-88; Carroll, 1990). Guided exploration owes its origins in the concept of "discovery learning" out of the late '60s and early '70s (Williams and Farkas, 1992, p. 41). Accomplishing tasks users often do not consider problem solving as a learning event. As Carroll and Rosson cite: "adults resist explicitly addressing themselves to new learning" (Carroll et al, 1987-88, p. 101); see also (Knowles, 1973, p. 73; Kidd, 1977, p. 77). In particular, if the critical situation is supposed to occur only once the user is not motivated to *learn* a solution. It is sufficient if he or she is enabled to create the solution, for instance by the help of step by step instructions not meant to induce a knowledge acquisition with the user. Williams and Farkas give an example where a user who has exceptionally to produce a footnote instead of known endnotes for a particular journal will not accept being "compelled (...) to 'learn' or 'remember' the procedures that he or she explicitly needs now in order to create the footnotes" (Williams and Farkas, 1992, p. 44). Only for recurrent problems and tasks will new knowledge be acquired.

Support from on-line help. When problems arise addressing the on-line-help is often insufficient for the user. The support users get from on-line help systems is restricted to the information that experts have brought into the system. Help from the system is restricted to information about system functionality and to well-known notorious problem situations (Fox et al., 1994), pp. 186f.). We only know one example that provides growing support based on questions of users and answers of consultants: "Answer Garden", see Ackerman and Malone (1990)[1]. On-line help support should be extendible to the user's individual results of exploration-based learning and to learning with consultants so as to integrate the learning results into technical support facilities (individualised help system). This individual help environment can be perceived as a user's own created guiding solution in contrast to the "guided exploration" manuals proposed by (Carroll, 1990).

1. "TeamInfo" was developed as a shared repository for informal group-relevant information by (Berlin et al., 1993). For producers of software, a "Living Design Memory" was proposed by [Terveen et al, 1993).

Consulting local or central experts. The learning process may occur individually where the user helps him- or herself by exploration (trying things out) but often the user asks for help consulting a competent colleague ("power-user") in face-to-face interaction or consulting an expert by telephone or remote diagnose. Empirical studies show that users have problems with consultants and consultants have problems with users (Brezizinski, 1988; Liechti, 1988; Moning and Winkelmann, 1993): Consultants are overloaded; their increasing number is over-compensated by a yet increasing number of clients; members of the user service units show limited availability; they are often not interested in the needs of users; they "forget" promises of problem solving that can't be executed immediately. Consultants have to solve (in their eyes) trivial problems and are therefore not motivated. User support is often organised on several levels (Brancheau et al., 1985) where the communication requires an exchange of problem and solution representations and where verbal or written descriptions are expensive and misunderstandable.

PROPOSED SOLUTION: SUPPORTING LEARNING AS AN ITERATIVE PROCESS

Learning does not occur as a single step. Learning is composed of detecting new knowledge, forgetting details and applying acquired knowledge. The learner proceeds in his or her competence by several trials of acquisition and application of qualifications. The first trial to acquire knowledge may be (a) exploratory, (b) supported by technical or human consultants, (c) error prone, (d) with indirect solutions, and (e) with dead ends. The first step of learning provides the user with rudimentary knowledge about errors, risks, and solutions. Making only one experience is not sufficient for full understanding and it is not robust to forgetting. It has to be reinforced and extended by re-use in later similar situations. In other words, the learning process is iterative.

With Lear (Learner's Living Repository), we introduce a conceptual framework that places special emphasis on integrating working and learning and on supporting self-directed and group learning. Prototypes of a support environment for learning and consultation in and after face-to-face or remote interactions will be developed and evaluated in a realistic work setting. The components of Lear can be described as follows:

Individual Memo: There are situations the user learns a particular feature of an application that only implies "declarative knowledge" (Anderson, 1976). A simple description of handling details, parameter settings or other *static* features is helpful if this description is at hand on demand. For this purpose we propose an explanation facility that we call an "individual memo" where the user enters his or her text or voice comments, sticks it to the relevant interface element, and consults it on demand.

Multimedia Demonstrations with Annotation Facilities: Scenic films can support the (re-)understanding of "procedural knowledge" (Anderson, 1976). To exploit *dynamic* features of tasks or system interaction, we propose a recording facility to produce a film of the interaction sequence. Such a film as a living interaction record is easier to grasp than a textual description. The demonstration supports the understanding in showing the process of actions and its effects. The film can be annotated to comment the rationale, the reasons for a solution, warnings to misleading assumptions, hints to unexpected side effects, etc., to support the transfer of the former solution to the current problem. Annotation facilities can support the user with respect to these goals.

Social learning: Learning in the working environment takes place in a social context. People have similar tasks using similar applications. They consult and support each other and exchange experiences and solutions. Technical support should enable the users to send and receive typical solutions (demo clips) found by an individual provided to the group.

Users can identify portions of an animated interaction sequence describing problems they encountered or solutions they found when using the tool, comment on them, and store them as episodes in a database called "demotheque". Users can send episodes that describe questions, problems with the tool, or breakdowns when using the tool as a request for off-line help to a consultant. Episodes that describe learned tool knowledge can be stored in the "demotheque" for later use. Episodes that describe users' personal experiences of solutions can be made available to a group of users in a "purse for demos". The elements of the proposal comprise the following features:

- **Individual memo:**
  The user can stick an electronic description to system elements in order to explain system features for individual needs.

- **Recording interactions**:
  The user can replay and explore duplicate the own interaction history (e.g. an error situation).

- **Defining relevant demos**:
  The user can select a relevant episode and keep it for similar future situations in a personal "demotheque".

- **Annotating demos:**
  The user can add comments and warnings referring to what he or she has done. Different modes of annotations should help to avoid information overload of a single sense organ.

- **Retrieving demos:**
  The user can select different access methods to retrieve the relevant episode from the demotheque.

- **Selecting Views:**
  The user can select different kinds of views to exploit a relevant episode from the "demotheque" for supporting the re-learning of a solution and for supporting the transfer of the solution to the current task.

- **Exchanging questions and answers:**
  A remote consultation can be supported to enable the

user and the consultant to exchange questions and answers independent of their time and space constraints.

- **Exchanging solutions:**
  The users of a co-operative work environment can exchange task and tool competence by providing and requesting solutions typical of the workspace ("purse for demos").

CONCLUSION

The problem of exploiting own and peer experience is of great practical relevance. People learn continously but they have difficulties in finding and using prior established competence. Lear is a self - learning support system to bring the user in touch with his or her learning history and to open the access to the competence of peers. The learners might exchange their problems and solutions by this mean. The learners can exchange their problems and solutions through a network so that cooperative work is supported more effectively than by a mail system. The idea of Lear is designed for the working environment but it might also stimulate in distant learning.

Lear is limited in the form of the acquired knowledge: using memos or registered screen actions and exchanging them within a group of domain workers for the moment. The initiative remaines fully in the head/hand of the user. Multi media techniques are only used scarcely not to distract the user from his or her main task.

First results of the implementation of the proposed facilities will be prepared for the conference. Evaluations have to show if and to which extent the proposed functionality of Lear will be accepted and used in a real working environment. Variables like time pressure, frequency of similar tasks, but also personality traits may influence the resonance of the facilities.

ACKNOWLEDGEMENTS

I would like to thank Christoph Thomas and Helmut Simm (GMD), Udo Bleimann (FH Darmstadt), Gerhard Fischer (University of Colorado at Boulder), and Marcin Sikorski (University of Gdansk) for their contribution to the ideas desribed in this paper.

REFERENCES

Ackerman, M.S., Th.W. Malone (1990): Answer Garden: a tool for growing organizational memory. In: *Proceedings of the Conference on Office Information Systems.*

Anderson, J.R. (1976): *Language, Memory, and Thought. Hillsdale, N.J.: Lawrence Erlbaum Associates.*

Berlin, L. M., R. Jeffries, V. L. O'Day, A. Paepke, C. Wharton (1993): Where did you put it? Issues in the design and use of a group memory. In: *Proceedings INTERCHI '93*, pp. 23-30.

Brancheau, J.C., D.R. Vogel, J.C. Wetherbee (1985): An investigation of the information center from the user's perspective. *Data Base* 17 (1985), 1, 4-17.

Brezizinski, R. (1988): When it's time to tear down the info center. *Datamation* 33 (1988), 21, 73-82.

Carroll, J.M., P.L. Smith-Kerker, J.R. Ford, S.A. Mazur-Rimetz (1987-1988): The minimal manual. *Human-Computer Interaction* 3, (1987-1988), 123-153.

Carroll, J.M. (1990): *The Nurnberg Funnel: Designing Minimalist Instruction for Practical Computer Skill.* Cambridge, MA.: MIT Press.

Dutke, S., W. Schönpflug (1987): When the introductionary period is over: Learning while doing one's job. In: M. Frese, E. Ulich, W. Dzida (Eds.): *Psychological Issues of Human-Computer Interaction in the Work Place.* Amsterdam: Elsevier Science Publishers B.V. (North-Holland), 295-310.

Fox, Th., G. Grunst, K.-J. Quast (1994): HyPLAN: A context-sensitive hypermedia help system. In: R. Oppermann (ed.): E. Hillsdale: Lawrence Erlbaum Associates, pp. 126-193.

Kidd, J.R. (1977): *How Adults Learn.* New York: Associated Press.

Knowles, M. S. (1973): *The Adult Learner: A Neglected Species.* Houston: Gulf Publishing Company, American Society for Training and Development.

Liechti, M. (1988): Das "Information Center" – wichtige Supportstelle im Betrieb, 1. Teil, io *Management-Zeitschrift* 57 (1988), 574-575.

Moning, U., B. Winkelmann (1993): Entwicklungsphasen von Information Center, Ergebnisse einer empirischen Untersuchung über betriebliche Endbenutzerunterstützung in der Schweiz. In: *Wirtschaftsinformatik*, 6 (1993), 532-541.

Paul, H. (1995): *Exploratives Agieren. Ein Beitrag zur ergonomischen* Gestaltung interaktiver Systeme. Frankfurt am Main: Peter Lang Verlag.

Terveen, L. G., P. G. Selfrigde, M. D. Long (1993): From "Folklore" To "Living Design Memory". In: *Proceedings INTERCHI '93*, pp. 15-22.

Williams, T.R., D. K. Farkas (1992): Minimalism reconsidered: should we design documentation for exploratory learning? *SIGCHI Bulletin* 24 (1992), 2, 41-50.

What's the Story? Narrative and the Comprehension of Educational Interactive Media

Lydia Plowman
School of Cognitive & Computing Sciences
University of Sussex
Brighton
BN1 9QH, UK

+44 1273 678647
lydiap@cogs.susx.ac.uk

ABSTRACT
The paper summarises findings from previous studies of interactive media in the classroom which suggest that the design of most software leads to a fragmented experience which impedes comprehension. The role of narrative in providing structure and coherence is discussed in the context of educational interactive media and initial findings from current research into the role of narrative are reported.

Keywords
interactive media, education, multimedia, narrative, interface design

INTRODUCTION

A recent large-scale evaluation study of interactive media in the classroom, based on over 200 schools in England and Wales (Laurillard *et al*, 1994), showed that although teachers are generally satisfied with the hardware of these technologies they are very disappointed with the design of the software. And although it is clear that children enjoy using interactive multimedia and find it motivating, this is in spite of its design rather than because of it. Children were frequently unable to make sense of the materials without considerable teacher preparation or supervision, even though much of the software is marketed as suitable for unsupervised use.

Interactive learning long pre-dates the introduction of computer-based learning and interactive multimedia. I will take two commonplace examples - one from a formal educational situation and one informal - as a starting point for considering some teaching and learning processes and how they might relate to the design of interactive media for school use. One example is of a competent teacher in the classroom, the other is a parent, sibling or friend reading a picture book with a novice or non-reader.

Interactivity and learning

The assertion that the best teachers are the best storytellers may be a cliché, but this demonstrates interactivity as it was understood before it was applied to the relationship between computers and people. Thus teachers are able to assimilate digressions into their stories, repeat points and tie up threads in a highly interactive way which takes account of the individuality of learners, the social context, relevant artefacts and the environment. They are able to elicit relevant knowledge from students and respond to what they say, to initiate, confirm, evaluate, reformulate, and give feedback (Mercer, 1995).

Jointly reading a picture book requires a certain degree of intersubjectivity between partners (i.e. language, the knowledge evoked by language, and sharing objects of attention, such as the book) which can be negotiated according to circumstances (Devescovi & Baumgartner, 1993). Reading a picture book is an opportunity to talk about both the story within the book and how it relates to aspects of the child's and more experienced reader's lives. Narrative has a vital role in the development of children's language and social skills whether as a result of exposure to narrative texts, such as television programmes and picture books (Meek, 1991), or from constructing their own narratives to make sense of their experience (Engel, 1995; McCabe & Peterson, 1991; Nelson, 1989).

NARRATIVE AND UNDERSTANDING

Clearly, computer programs which can simulate the complexities of the types of interactivity demonstrated by a classroom teacher or a more experienced reader have not yet been developed, but these are models to which the design of educational interactive media could aspire. The narrativisation of knowledge, whether directly by a person, or indirectly through a text, is a way of capturing and holding attention and without students' attention, learning cannot take place. The question 'What's the story?' refers not only to fictionalised accounts but, more often, is a request for a series of more or less related events to be filtered and wrapped up in a coherent package. Narrative aids reconstruction because of its dependence on cause and effect and, therefore, linearity. Narrative has a vital role in the processes of teaching and learning because, apart from any motivational aspects, it is an aid to memory and, coupled with this, a way of structuring information. For the purposes of this discussion I am therefore interested in the role of narrative as a macro-structure, although the perception of structure also alters our understanding of content.

Narrative in its various forms underlies these examples of positive interactivity because it is a structure which is used to generate and convey information and knowledge between people. Narrative depends on interactivity because it is 'a solution to the problem of how to translate knowing into telling' (Scholes, 1980) and therefore has its own dynamic—the transition from one state to the other and, sometimes, from one person to another. In an educa-

tional context both types of process are necessary because, whether it is a learner-centred situation or one based on a transmission model, information or knowledge needs at some stage to be conveyed from teacher to pupil. This information needs to be explicitly structured before it is internalised.

Narrative is not therefore associated only with fiction in the Aristotlean sense of texts with a beginning, a middle and an end, it *shapes* our knowledge and experience and we have what Bruner (1990) describes as a predisposition to finding and creating narrative. He claims that the principal property of narrative is its inherent sequentiality and that it is composed of a unique sequence of events, mental states, and happenings involving human beings as characters or actors which are its constituents. The meaning of these constituents is given by their place in the overall configuration of the sequence as a whole rather than as separate units.

The generation of narrative is an active process of construction through which we make sense of the chaos of our daily experiences, whether for others or for ourselves. It creates a structure for interpreting and understanding connections and links between people and events. But although we are skilled at imposing order in this way, this is not necessarily an efficient use of cognitive resources when engaging with difficult texts.

There is a significant body of research which suggests that texts which are unfamiliarly structured because they do not conform to mental models of narrative make excessive demands on the learner's cognitive processes, and that memory and comprehension can be used most productively when the text is clearly structured and navigable. Learners are constantly adjusting their understanding in accordance with their exposure to conventional narratives, making the construction of 'story' a central cognitive goal (Bower & Morrow, 1990; Cain, 1996; Thorndyke, 1977). This research has focussed on written texts; so far there is a dearth of such research for interactive media, although Thüring *et al.* (1995) discuss this in relation to hyperdocuments. They describe increasing the readability of hyperdocuments by reducing cognitive overhead and increasing coherence, calling for an 'interpretative context' for such texts. Using the example of SPI (Sepia Presentation Interface), a hypermedia authoring environment, they make a number of suggestions for designing for comprehension. These include context preservation, visual orientation cues and graphical maps, but one way in which an interpretative context can be provided for educational materials is by narrative.

THE STRUCTURE OF INTERACTIVE MULTIMEDIA TEXTS

The two main forms of educational interactive media software are those which are presented in an encyclopedia format, designed to support both free exploration and specific information enquiries, and those which use a more guided and focussed tutorial format. Both formats have structures, the question is whether these are perceptible to users. Although encyclopedia-style software is more commonly supposed to be lacking in structure, at least the structure at the highest level is recognisably that of an encyclopedia, even if the specifics of navigation are not clear. Tutorial-style multimedia software does not have a comparable provenance, so although it may seem to lend itself more obviously to a narrative structure, the lack of precedents makes design decisions difficult, especially given the lack of empirical work in this area. Either way, comprehension problems are exacerbated for users when compared to books or live tutors. Although multimedia programmes superficially appear to combine media with which we are already familiar, such as books, film, and television, the interpreting skills we have acquired from exposure to these traditional media are not directly transferable, partly because the interactivity is often an intrusion rather than integrated into the text.

It makes sense, therefore, that educational texts - in any medium - should have a recognisable structure so that learners are able to commit more of their cognitive resources to understanding the content than labouring to understand its form. Textbooks and educational television programmes use familiar structures and their use is generally supervised by the teacher within the classroom. They operate, therefore, within a cultural and social context. Many educational interactive multimedia programmes, however, are designed to substitute to some extent for the teacher and children are frequently sent to a different room containing the appropriate hardware—typically the library—to use them. In such a situation, they are confronted not only by a new medium[1] but they are left to cope with it on their own.

As part of another study I developed media interaction charts (MICs) for graphically representing interaction with the system, interaction within the group using it, and the transitions between media (Plowman, 1992, 1996). These charts were based on observations of use of interactive video on laserdiscs. However, the cognitive and design implications are equally relevant to CD-ROMs and materials presented over the World Wide Web as the platform is not as significant as the interaction of media and cognition (Salomon, 1994). If anything, the new media are more confusing as the introduction of multiple windows means the screen is likely to be busy with different symbol systems.

This detailed observation of children using interactive media for tutorial-style lesson sequences demonstrated how using interactive multimedia can be a fractured and disorientating experience. A series of discrete screen events, often lasting little more than a few seconds, is interspersed with group decision-making or interaction with the machine for entering text or selecting from a menu. The MICs showed that using interactive media was highly interactive in the sense of frequent and numerous interactions with the machine, but many of these interac-

---

1. Even though all schools in the U.K. are equipped with CD-ROM drives as a result of various government initiatives, a surprising number of children involved in the current research are using a CD-ROM for the first time.

tions were empty or gratuitous and did not contribute to the learning experience—other than by developing social skills through their problem-solving discussions.

Although I earlier pointed to the interdependence of narrative and interactivity in certain learning situations, these are where the interactivity is flexible and sensitive enough to adapt to the learners' needs, i.e. in a live situation. This flexibility and sensitivity is still not apparent in interactive media and so the narrative becomes disturbed or fragmented.

Narrative structure is one of the most important ways in which the educational message comes to be understood but the structure of interactive multimedia differs from those of other media with which we have more familiarity because

- the media formats switch frequently between video, text, animation, graphics, sound, and silence;
- there are combinations of different media on the screen at the same time;
- the learner can be given control over aspects of pace, sequence, choice of activity, and input to the system;
- typical usage is in groups, so there are frequent pauses to discuss decisions relating to learner control or who is to use the input device and there are likely to be multiple interpretations of the different representations.

The effects of this fragmentation of the learning experience are under-explored for both cognitive and affective aspects of learning (but see Blissett & Atkins, 1993; Laurillard & Taylor, 1994; Stratfold, 1994). Further research is now under way to investigate in more depth some of the findings from the previous studies and to consider the role of narrative as a structure for increasing comprehension of interactive media. At this early stage of MENO (Multimedia, Education and Narrative Organisation) we can only report initial findings. Fieldwork is currently taking place in two secondary schools to investigate use of CD-ROMs in as naturalistic a context as possible. Findings will be used to develop and test software which presents material in different narrative and non-narrative structures.

A NEW MEDIUM, A NEW INTERFACE

The fact that interactive multimedia is relatively new means that conventions of representation have not yet fully emerged (Plowman, 1994). In some respects, it is at a similar stage of development as film was at the turn of the century. This is described by Burch (1986) as the 'primitive mode of representation' because it had not yet developed into the 'institutional mode', with which we are now familiar. Most films in this classical Hollywood tradition exhibit conventions of seamless editing, clear transitions from one time or place to another, and a coherent narrative. Modern films which do not conform to these conventions are often defined as avant-garde texts which are more concerned with explorations of form than coherence of narrative and so are perceived as difficult to understand. Indeed, this difficulty is integral to their purpose, which is often to disturb the cognitive comforts of the audience.

Interactive multimedia does not conform to many of the Hollywood conventions. As it is a new medium with different forms of representation and relationship to the audience this is not necessarily problematic. Nevertheless, I argue that it is still in its 'primitive mode' and has not yet evolved into a more stable mode with conventions that have developed over a period of time. Even where conventions have emerged, designers cannot be sure that novice users are familiar with them (see below). This causes problems for the users just as the 'primitive mode' of film caused problems for its audiences.

The early years of cinema

During the early years of cinema audiences had problems adjusting to a nascent medium and various devices were used to help them understand what was happening. Key among these was the use of a film explainer or *explicador* whose job was to interpret what was going on for the audience (Hofmann, 1995). According to Carrière (1995), these survived in Spain until 1920 and in Africa until the 1950s: 'the cinema had brought with it a new language, which few spectators could absorb without effort, without help'.

The film explainer stood between the cinema screen and the audience, anticipating and responding to problems by using their prior knowledge of the films being screened to explain both structure and content. They 'talked the early audience through the film, directing their eyes to significant details which their untrained perception might otherwise miss' and 'anchored the ambiguous images on the screen' (Chanan, 1996).

Audiences were unavoidably confronted by the impact of the technology on the experience of engaging with a film. They had to be quite patient as they waited for the next, short piece of film footage (each segment was typically 50 feet and ran for 45 seconds in the mid-1890s) and so narrative could not be seamlessly developed. The film explainer did not simply *explain* what was happening in films and deal, in current jargon, with interface problems, he was also largely responsible for the narrative. As segments of film increased (20 minutes by 1910) there was increasing opportunity for narrative development to be carried by the film rather than by the explainer.

Thus the explainer corresponds to the experienced reader in the earlier example of interactive learning who mediates the as yet unrealised text of the picture book by making connections between the images and the writing, asking questions, relating the text to their own experiences, and guiding the learner through both structure and content. In that example, guidance includes showing the novice how to hold a book the right way up, how to interpret the symbol systems associated with pictures and writing, and explaining a number of other conventions associated with reading a book. For the film explainer, it included commentary on the characters and plots as well as explaining some of the conventions of this new form of representation, such as close-ups.

The early years of interactive media

The MENO research suggests that designers cannot take for granted even the most basic conventions of user interface design. Several groups of children aged 13 and 14, using *World War II: Sources and Analysis* (the first time they had used a CD-ROM), were not clicking on highlighted text. When asked why they thought the words were in different coloured text they speculated that they were keywords or 'important', but did not know they could access a definition or further information by selecting the word. Other children did not know that clicking in the corner of a window would close it; some were unfamiliar with the convention of scrolling through text. These problems come into Thüring *et al.*'s (1995) category of 'user-interface adjustment' as an example of cognitive overhead. (Other examples are orientation and navigation which I do not deal with directly here, although they are clearly related to narrative structure.)

The children were able to make an informed choice if their interaction was guided. At the end of one segment of information about the Blitz they were told, in text, that if they wanted to watch some related video they should click on the icon. This decision was therefore easy to make although another example of user-interface adjustment was not knowing how to interpret the icons and having to look them up in the manual. Once they had the video window on-screen they were easily able to deal with the controls because they were the same as for any standard video recorder. These have been in existence long enough for controls to be standardised and for most people to have had exposure to them.

Other children looking at some video about the atom bomb from the same CD-ROM asked the teacher what the video was about even though an immediately adjacent window contained written information about the footage. They had failed to notice it because they were focussed on the video and had not yet fully adjusted either to understanding that information from multiple sources is presented simultaneously on the same screen or being able to assimilate it.

This is not a new problem: writing in 1938 about talking films, Arnheim (1958) writes that 'uneasiness is due to the spectator's attention being torn in two directions. In their attempts to attract the audience, two media are fighting each other instead of capturing it by united effort. Since the two media are striving to express the same matter in a two-fold way, a disconcerting coincidence of two voices results, each of which is prevented by the other from telling more than half of what it would like to tell'. Audiences have now adjusted to dealing with a medium such as film in which both sound and vision are combined, but Arnheim's views highlight how different are our responses when we first engage with new media. To return to the examples of interactive learning given earlier, we can see that situations such as these would not happen if a person rather than software was providing the interactivity with the learners, because they would be able to anticipate problems and shape their guidance according to need. Children who had more exposure to interactive media would not need as much help as novices.

As children gain increasing familiarity with a range of multimedia software they become quite alert to similarities and differences in interface design and will transfer knowledge of one programme to another, particularly if it is produced by the same publisher and there is intended similarity of design. The software therefore needs to be capable of adapting to differing degrees of expertise, allowing introductory tours or the presence of an explainer to be overridden. This process of adaption could mean that children learn to adjust to the lack of structure and find ways of compensating for it. Nevertheless, this still represents a cognitive overhead which is a drain on the resources they are able to bring to the learning situation.

NARRATIVE AND THE STRUCTURE OF INTERACTIVE MEDIA

Although it is fairly clear how narrative might help as a structure for tutorial-type programmes, it is more difficult to see how it might contribute to the comprehension of encyclopedia-type formats or help with the examples of low-level user interface problems outlined above. It is not just the narrative structure which holds everything together at a conceptual level which is violated but our basic means of engaging with the text is threatened if we have to learn a new set of conventions before we can begin to learn.

If learners have to overcome basic technical constraints of the medium before they can engage with the narrative (as early film audiences had to) then, as we have seen, this gets in the way of engaging with a learning experience. Part of the problem is that there are still limits to the amount of video which can be included on a CD-ROM — and it is the video segments which carry most of the narrative (Plowman, 1996). The conventions of the filmic 'institutional mode of representation' emerged in an attempt to diminish our recognition of the artifice of the medium of film and sustain narrative momentum. As compression techniques improve and there are other technological advances it is likely that interactive media will also undergo a similar transition.

One of the problems in the design of educational materials is the tension between providing learning tasks as part of the software and providing support for externally provided tasks with information and resources. This manifests itself in design terms as a series of units with varying degrees of relationship to each other and to the overall structure. We have characterised this in terms of the macro-narrative of the global structure and the micro-narratives of the lower-level task units. The following categories are an initial attempt to classify different narrative structures and how they might be applied; we expect to develop this as the research progresses.

Explicit narratives

Some programmes are designed with an explicit narrative which operates at a macro-level to give coherence to the content. An example of this is *Clémentine*, a CD-ROM designed to teach French to less motivated teenagers. The macro-narrative involves Clémentine having to raise a million francs to save the Bengali tigers and all the sub-tasks, such as designing posters, arranging coverage on the local radio station, and Clémentine's appearance on a TV charity

quiz show, relate to this. The structure is relatively linear because this was seen as more appropriate for the guided progression needed for language skills.

But an explicit macro-narrative does not necessarily lead to a coherent learning experience. *School Disco* was designed with a strong narrative of organising a school disco but the tasks involved (such as calculating cumulative frequency) did not always bear a clear relationship to the narrative, or were so complex that it became forgotten. In such an example there is a tension between the macro-level of the narrative and the lower-level task units.

Sub-narratives

Stowaway, a programme designed to teach children about life on an 18th century fighting ship, does not have an explicit macro-level narrative in the sense of the two examples given above, but uses a cutaway image of the ship to provide continuity. Navigation around the ship is achieved by clicking on this icon, which is always accessible. This provides coherence for a number of sub-narratives, one of which is to find the stowaway. Others are the stories of various characters on the boat, such as the captain or surgeon, told in an elctronic book format.

Externally imposed narrative

Initial findings suggest that inventing a suitable meta-level narrative can help children to be more focussed and directed in their use of CD-ROMs designed in an encyclopedia format. One group is using *Ancient Lands* to find information to help them solve a Roman murder mystery story which has been presented by the teacher, based on skeletons and bones found at Hadrian's Wall in the 1930s.

Children using *World War 2: Sources and Analysis* who had a clear task provided by their teacher (to imagine themselves either as a Japanese American living in the United States, or as a Japanese person, or as an American, at the time of the dropping of the atomic bombs in 1945 and present their point of view) were able to enjoy a richer learning experience than those who did not have this framework and were confused by its design at both an interface and structural level.[2]

Negotiated narrative

Simulation games such as *SimCity 2000* allow the user to have a more directive approach to the narrative development although, as this has to be within the confines of the programming, it is negotiated (Stratfold, 1994). The game format means that the teacher needs to make the learning objectives clear.

CONCLUSIONS

Given that group use is the norm in schools, the multimedia programme becomes an artefact in a cognitive system which encompasses the students, the teacher, and the wider social and cultural context. A strong narrative should be a good way of linking individual and group cognition as it is also a link between the individual and the cultural.

Students must be able to relate the constituent units of interactive media programmes to the macro-structure but also perceive the relationship between sub-units. Integrating content and structure is also paramount for comprehension.

Narrative is a means of achieving this. I have outlined different narrative structures and suggested that narrative can also help with low-level user interface problems. If we accept that interactive media is still in its 'primitive mode' then the device of the film explainer is worth considering as a way of dealing with both interface problems and giving coherence to the experience of using educational multimedia.

Although it is more difficult to imagine how narrative can be applied to materials in the encyclopedia-style format, *Stowaway* is an example of how a coherent macro-structure with an integrated task can be designed. Looking for the stowaway is an enticement for children to look through the information as they seek him. It does not, of course, guarantee that they will assimilate the information and if a game becomes the main element there is a risk that children's interests are diverted into playing the game. This is one area which will be the subject of further research.

The example of a teacher providing a narrative context suggests that if the software itself provided these narratives it may be a good approach to helping children get the most out of interactive media which present archive material or use an encyclopedia format. It would also save teachers a considerable amount of preparation. Perhaps a choice of narratives could be built in to the design and provide a means of introducing the children to the different sources and types of material available to them. This could be overridden once they had enough familiarity with the disc.

ACKNOWLEDGEMENTS

MENO (Multimedia, Education and Narrative Organisation) is funded by the ESRC Cognitive Engineering Programme, grant no. L127251018. I am indebted to other members of the team - Diana Laurillard, Alistair Scott, Matthew Stratfold, and Josie Taylor - and to Jim Fanning at Tideway School, Newhaven and Gill Charles at Warden Park School, Cuckfield and their pupils.

REFERENCES

* An asterisk indicates publications which are available from the following web site: http://www-iet.open.ac.uk/iet/iet/MENO/meno-home.html

Arnheim, R. (1958). A new Laocoön: artistic composites and the talking film (orig. pub. 1938) in *Film as Art*. Faber & Faber, London.

Blissett, G. & M. Atkins (1993). Are they thinking? Are they learning? A study of the use of interactive video. *Computers in Education* 21 (1/2) 31-39.

2. At this stage we are collecting qualitative data and not measuring learning gain. At a later stage of the study there will be an experimental dimension which will provide different data.

Bower, G.H. & D.G. Morrow (1990). Mental models in narrative comprehension. *Science* 247 (4938) 44-48.

Bruner, J. (1990). *Acts of Meaning*. Harvard University Press, Cambridge, MA.

Burch, N. (1986). Primitivism and the avant-gardes: a dialectical approach. In *Narrative, Apparatus, Ideology*, ed. P. Rosen, pp. 483-505. Columbia University Press, N.Y.

Cain, K. (1996). Story knowledge and comprehension skill. To appear in *Children with Reading Comprehension Difficulties: Processes and Interventions*, eds. C. Cornoldi & J. Oakhill. Erlbaum, Hillsdale, N.J.

Carrière, J-C. (1995). *The Secret Language of Film*. Faber & Faber, London.

Chanan, M. (1996). *The Dream That Kicks: the Prehistory and Early Years of Cinema in Britain* (2nd ed., first pub. 1980). Routledge, London.

Devescovi, A. & E. Baumgartner (1993). Joint-reading a picture book: verbal interaction and narrative skills. *Cognition and Instruction* 11 (3/4) 299-323.

Engel, S. (1995). *The Stories Children Tell: Making Sense of the Narratives of Childhood*. WH Freeman, N.Y.

Hofmann, G. (1995). *The Film Explainer*, trans. M. Hofmann. Secker & Warburg, London.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, Ma., MIT Press.

Laurillard, D., L. Baric, P. Chambers, G. Easting, A. Kirkwood, L. Plowman, P. Russell & J. Taylor (1994). *Interactive Media in the Classroom: Report of the Evaluation Study*. National Council for Educational Technology, Coventry.*

Laurillard, D. & J. Taylor (1994). Designing the Stepping Stones: an evaluation of interactive media in the classroom. *Journal of Educational Television* 20 (3) 169-184.*

McCabe, A. & C. Peterson, eds. (1991). *Developing Narrative Structure*. Lawrence Erlbaum, Hillsdale, N.J.

Meek, M. (1991). *On Being Literate*. Bodley Head, London.

Mercer, N. (1995). *The Guided Construction of Knowledge - Talk Amongst Teachers and Learners*. Clevedon, Multilingual Matters.

Nelson, K., ed. (1989). *Narratives from the Crib*. Harvard University Press, Cambridge, MA.

Plowman, L. (1996). Narrative, linearity and interactivity: making sense of interactive multimedia. *British Journal of Educational Technology* 27 (2) 92-105.*

Plowman, L. (1994). The 'Primitive Mode of Representation' and the evolution of interactive multimedia. *Journal of Educational Multimedia and Hypermedia* 3 (3/4) 275-293.*

Plowman, L. (1992). An ethnographic approach to analysing navigation and task structure in interactive multimedia: some design issues for group use. In *People and Computers VII*, eds. A. Monk, D. Diaper & M. D. Harrison, pp. 271-287. Cambridge University Press, Cambridge.

Salomon, G. (1994). *Interaction of Media, Cognition, and Learning* (orig. pub. 1979). Lawrence Erlbaum, Hillsdale, N.J.

Scholes, (1980). Language, Narrative and Anti-Narrative. In *On Narrative*, ed. W.J.T. Mitchell. University of Chicago Press, Chicago.

Stratfold, Matthew (1994). *Investigation into the Design of Educational Multimedia: Video, Interactivity and Narrative*. Unpublished PhD thesis, Open University.*

Thorndyke, P. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology* 9 77-110.

Thüring, M., J. Hannemann & J. Haake (1995). Hypermedia and cognition: designing for comprehension. *Communications of the ACM* 38 (8) 57-66.

Multimedia software referred to in the text:

Ancient Lands (Microsoft Home, 1994); *Clémentine's Cause* (BBC Education, forthcoming, 1996); *School Disco* (laserdisc, produced as part of the Interactive Video in Schools project by Centre for Innovation in Mathematics Teaching, Exeter University); *SimCity 2000* (Maxis, 1995); Stowaway (Dorling Kindersley, 1994); *World War II: Sources and Analysis* (Interactive Learning Productions & Yorkshire TV, 1994).

Helping to acquire the perceptuomotor skills underlying the learning of prosody in a foreign language

B. Denoeud, M. Raquillet and A. Foti
Laboratoire d'Ergonomie du CNAM
41, rue Gay-Lussac
75005 Paris
France
tel (33) 1 43 54 18 34; fax (33) 1 43 25 36 14
denoeud@cnam.fr

ABSTRACT
In the perspective of designing help systems for learning an oral foreign language we have developed a cognitive analysis of the acquisition of perceptuomotor skills underlying the learning and practice of stress and intonation. Starting with this cognitive approach, we have defined two main needs to satisfy in the framework of a computer based training system: hearing what is to be perceived and to evaluate more objectively the discrepancies between productions. We have designed an "intermediate representation" of stressing and intonation phenomena, based on the integration of acoustic parameters. We have also implemented a simulation of auditory segmentation of speech based on purely acoustic signs. A first technical evaluation is evoked, but we highlight principally the new experimental situation created by the introduction of prosodic representations, in a domain where cognitive representation are unusual outside the works of phoneticians.

Keywords
ergonomics, artificial intelligence, help system, stress, intonation, phonetics.

INTRODUCTION
"It's perhaps a discouraging thing to say, but learners of English who are not able to listen regularly to colloquial English, are not likely to learn English Intonation, though they may learn very good pronunciation of the segments and use stress correctly."

This quotation from Roach (1983) in his manual "English Phonetics and Phonology" is a dramatic illustration of the difficulty which French speakers have in learning English intonation even though they have studied it for some considerable time and can write English properly. And yet phonetic sciences have accumulated phonetic and phonological knowledge about stress and intonation phenomena. The acoustic substance of intonation is known and is based on the intensity, the fundamental frequency (F_0) and the time. But studies of perception show the gap between physical variables of frequency and intensity and subjective variables which are respectively the pitch and the loudness. The auditory system has an overall perception of variations in acoustic variables, to the point where subjects are unable to recognize the part of each of them (Rossi, 1981). One of the great difficulties in producing perceptive models of stress and intonation is the fact that these phenomena are based on the same acoustic substance and it is difficult to untangle the codes linked to each of them. For these are real codes in which the signs are multiple and can be substituted for each other. The variations of F_0 and intensity should be separated between intrinsic variations linked to the specific features of phonemes and their co-articulation, and linguistic variations which must be interpreted in terms of the codes indicated previously, the majority of which are still to be discovered. Finally, the phonatory system generates random variations of the same variables which explain the usual variability of the oral productions.

From the viewpoint of developing help systems for computer based training, we studied the problematics of helping a Learner to acquire the underlying skills of the learning of stress and intonation in a foreign language. In order to develop the problematics, we shall outline a cognitive approach to this learning activity in order to define the Learner's help requirements. Then we shall develop the help systems which we have implemented in the context of a self-training system.

Cognitive Approach to the Prosody Learning Activity
Let's consider the situation of the adult Learner who learns the stress and the intonation of a foreign language. This Learner, of an intermediate level, could be an professional adult who studied English at secondary school and university. He can write English but has great difficulty in understanding the native English speakers he contacts for professional reasons. Words that are known in writing are not recognized when spoken. As for the intonation, it can result in misinterpretation when the speaker tries to express himself.

In regard to our development targets, we are going to consider the activity of the Learner who studies these aspects of the spoken English language in contact with the recordings of native English speakers and who tries to repeat them through a computer based training system.

We shall not pretend to develop a comprehensive cognitive analysis of the learning activity which still has to be done, but shall simply underline three major points in our problematics:

1. the inadequacies of auditory feedback,
2. the role of auditory feedback in voice control,
3. the role of auditory perception in learning prosody.

The Inadequacies of Auditory Feedback
Teachers of English often notice that their pupils "do not hear" what is to be perceived and that they "do not hear" when they speak. The efforts in language laboratories are

rather ineffective because the naive speaker is also a naive listener (Boulakia, 1987). We could illustrate this fact through an example which is not intended to be representative. In an test of an oral English tutorial, a Learner practised the repetition of oral sentences until he reached what he considered to be a satisfactory level.

The example chosen concerns the sentence: "Good morning Mrs. Anderson. How are you?". The second part of the sentence had to be pronounced with a sentence stress on the "How" and a descending intonation opposite to the rising intonation of the interrogative in French. After having heard a total of some twenty repetitions of the model sentence, one subject stopped studying this sentence, admitting that he was not happy with the end result. Since he had not found the descending intonational schema, he was never satisfied with his intonation of "you" pronounced in a rising tone as in French. This problem was coupled with an ignorance of weak vowels which made him pronounce "are" while the sound segment only contained the /ə/ phoneme, a weak vowel (schwa).

This example illustrates the inadequacies of the auditory feedback to identify the correct intonational schema on a three words expression and the ineffective listening which cannot ensure a correct repetition of the "ə" phoneme of "are".

For the Learner, this observation underlines three phenomena which determine a difficulty in the learning of prosody:

i) the importance of writing, ii) the phonemic restoration and iii) the impact of prosodic schemas of the mother tongue

i) The learner who masters the written language but does not practice speech can only invent the oral product through the image of the written product if he has not been warned of the gap between the written performance and the spoken performance.

ii) Our perception of speech does not need to hear all these phonemes making up the isolated word in order to identify it in the context. The phonemic restoration gives us the feeling of perceiving all the phonemes, even if they are not pronounced. As such, the "ə" phoneme is perceived as "are" in our example and is even repeated with stress.

iii) The significance of the prosodic schemas of the mother tongue explain why a Learner, hearing an interrogative sentence, pronounced with a descending intonation, reproduces it "totally unconsciously" with the rising schema of the French interrogative. As such, auditory feedback does not appear sufficient to learn the stress and intonational schemas of a foreign language since our auditory perception activates the alone available schemas, those of our mother tongue.

The Role of Auditory feedback In Voice Control

When learning the prosody of a foreign language, the Learner has to learn to control his voice in order to produce new stressing and intonation schemas. Auditory feedback is used by speakers to control their voice. Subjects who become totally deaf after learning a language, rapidly lose control of their intonation. The main function of auditory feedback appears to be that of ensuring control of pitch variations which are an essential parameter of intonation. But there are considerable differences in auditory feedback between real-time feedback at the time of speaking and the hearing of one's own voice after a recording. In the production of speech, feedback takes place through both aerial conduction and bone conduction. The latter sounds as if the higher frequencies were removed. This low-pass filter explains a part of the differences in perception between the voice at the time of speaking and the voice reproduced by the recording.

But there is an other input for voice control and sensation, the sensibility and proprioceptive memory that are underscored by studies about the singing voice. Proprioceptive sensitivity is long and difficult to learn. It consists of associating an auditory image with a proprioceptive sensation. In the case of singing, auditory feedback controls the pitch while proprioceptive feedback enables the voice to be placed and to determine its tone colour. It is through the combination of internal sensitivities that the singer controls his phonatory effort and judges the quality of his production (Scotto di Carlo, 1990).

For the mother tongue, it could be that proprioceptive signs are used to control the production of the usual prosodic schemas of the language and that memory of intonation is there. This means also that auditory perception does not control all the parameters involved in the prosody. In order to learn a foreign language, the subject does not have the schemas derived from a proprioceptive memory of the language and he places an excessive burden on the auditory perception compared to its real capacity to control his own voice.

The Role of Auditory Perception In the Learning of Prosody

Any speaker who masters his mother tongue has the competencies for perceiving and producing the stressing and intonation schemas relative to his language. Although he may have memorized these schemas, he does not necessarily know the laws or even the forms. As such, the French speaker is often unaware of the position of the word stress in French, while he applies it unconsciously. He has memorized it because he is capable of detecting any violation of the usual schema. Theories of the motor control of speech do not need to have a conscious representation of stressing and intonation schemas in order to produce them and detect any violation. These theories also explain how the perception and production of phonemes are linked, with recognition resulting from processing through reproduction or synthesis.

In addition, an adult who learns the prosody of a foreign language must adapt his perceptive and articulatory mechanisms in order to produce new stressing and intonation schemas. This adaptation can be very heavy as in the case of English Language II for native French, where the perceptive mechanisms of French, based on the syllable, are unsuitable for the perception of English words, based on stress. (Cutler, 1983). In the learning phase, only auditory perception can guide the Learner if his sources are purely acoustic. In the interaction with a person, there are also the resources of visual perception of the articulatory gestures.

For the adult, it is natural to try to develop learning through differentiation which requires an explanation of the schemas of the mother tongue and language II. Therefore, it is necessary to introduce the subjective parameters of pitch, intensity and time through which a teacher who knows phonetics can express these phenomena. But one should not lose sight of the fact that the target is to develop sensitivity and a proprioceptive memory of the prosodic production in language II.

The analysis of methods of American English Teachers, concerned with these subjects, shows which operative representations they used. They explain the stress as a variation of the force of voice, of intensity, and the intonation as a variation of the pitch. Their teaching methods suggest ideas to design a computer system that might free Learners from the presence of a Tutor. But we must be aware that the guidance of a Tutor operate on a high level perception, significant only between humans. To design an automatic system, we need to 'simulate' parts of perceptive processes or design other help system that could show what to perceive and to produce.

The Learner's Requirements For Self-Training

In terms of the cognitive analysis developed previously, and in view of the self-training which we described heretofore, we can define the requirements to be satisfied in order to help Learner acquire prosody.

1. Provide a *visual representation* of the prosodic phenomena which enable the Learner to focus his attention on these phenomena and to evaluate his products more objectively than through his auditory perception.

2. Present *isolated sound segments* in order to make the Learner aware of what he is producing « objectively » and to show him the gap between oral and written languages. It is easy to understand the utility of listening to isolated sound segments corresponding to sentence stress, prominent points of a prosodic schema or again a single phoneme as in the example mentioned previously. A specific, detailed comparison of two oral productions between a model sentence and its reproduction, or between two repetitions of the Learner requires the rapid, serial presentation of the corresponding segments and segmentation of speech signal into linguistic units which are significant for the speaker.

3. *Mobilize all the perceptive and cognitive resources* around the construction of representations of prosody for learning purposes. Different types of presentation enable to draw upon different resources.

Now we are going to develop the help systems which we developed to answer the requirements in three main points:

1. *an intermediate representation of prosody,*
2. *a simulation of auditory segmentation of speech,*
3. *an interface for direct manipulation of presentations.*

An 'Intermediate Representation' of Prosody

This involves the development of a visualization of prosodic phenomena which represents stress and intonation and is operative for the Learner since it enables him to evaluate his productions.

Our perception of prosody is synthetic. As such, the representation has to take this situation into account. Yet all feedback is analytical. Therefore, we had the idea of designing a visual representation of the perceptive integration of the different acoustic variables of stress and intonation. Since the juxtaposition of the intensity and F_0 curves is difficult to understand and is not pertinent for learners, we shall integrate them in a single image in order to transpose into visual modality that which is difficult to render in auditory modality. An image of the intensity had to be given for the stress which teachers represent through the parameter of intensity. For intonation, which is usually represented by a melodic curve, an image showing high variations had to be produced. These two images were combined in a synthetic image of prosody, the scale of which is that of the sentence.

We have no pretension of simulating a perceptive model of intonation with a system which does not recognise the phonemes pronounced. Such a model, which should eliminate the microprosodic variations before weighing up the linguistic code relative to stress then intonation (Rossi, 1981) cannot be implemented. But we are trying to develop an "intermediate representation" between the separate representation of acoustic variables, which is not pertinent for the learner, and a perceptive representation like the one which could be produced by experts in intonology. This intermediate representation is produced to give the Learner a sign and to stimulate his auditory capacities without giving him the model of what has to be perceived in detail.

The framework of this representation is provided by the F_0 curve which gives an image of the melodic line of the spoken sentence. On this F_0 curve, the intensity is shown in + and in - relative to the F_0 curve.

The superposed intensity curve has two advantages: it indicates the variation of intensity and adds a visual segmentation of the F_0 curve. The F_0 curve on its own is difficult to interpret since it presents a continuity of all the voiced parts without showing significant units for the speaker (words and syllables). As such the synthetic representation highlights the significant units which most often correspond to syllables.

The image produced by the voiced parts of the signal is that of a red connecting line, surrounded by a green sheath whose thickness varies. The visual green masses integrate the dimension of intensity on the vertical line and the dimension of time on the horizontal line. The visual masses and the pitch help underline the prominent points of the intonation (sentence stress, expressive accents) while the F_0 curve and the centre of gravity of the masses show the melodic curve. At a more precise level of analysis, the co-variation of the intensity of F_0 helps to highlight the stress inside words.

The representation of non-voiced consonants raises a problem in as much as F_0 has no sense and the intensity of a non-voiced consonant may be very strong as in the fricatives ("ch", "ss") in which there is no relation with the intensity perceived. As such, we chose to present the corresponding speech signal with a scale of intensity lower

than that of the voice signal which, alone, bears the prosodic information.

Figure 1 shows the representation of isolated words. The highlighting of the stress schema is noticeable.

Figure 2 shows the prosodic representation of a sentence of four seconds and the full system interface. The upper window is reserved for the model sentence while the successive productions of the Learner are displayed and stacked on the lower window. The textual representation of the model sentence, closely connected to the prosodic representation, is noticeable.

In an initial technical evaluation of this representation, we noted that Learners who were simply told that "the red line indicates the melodic line and the green masses represent intensity" did not ask any questions about this representation and, in particular, used the "melodic line" to define correction targets.

Simulation of the Auditory Segmentation of Speech

Through analyzing existing help systems which help with this learning, we noted that use of the tape recorder, as in language laboratories, was not suitable for presenting isolated sound segments due to the difficulty of referencing and handling and the time required for this technology. First of all the defects of the tape recorder have to be corrected. In addition, automatic segmentation of sound productions has to be performed in order to relieve the Learner and enable him to concentrate on his main task. In perceptive learning, the importance of associated tasks and waiting times should be noted since they erase the highly evanescent auditory image.

For all these reasons, we developed an automatic speech signal segmentation system which segments oral productions into 3 to 5 segments for a performance of two seconds. These segments also have a visual image on the

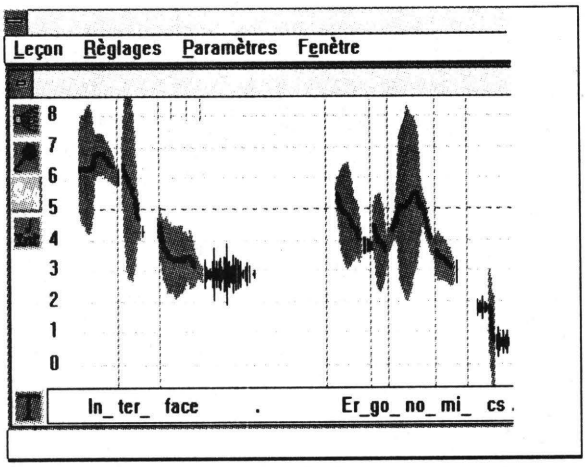


Fig. 1. Prosodic representation of isolated words. The black line is the melodic contour, the grey masses show intensity.

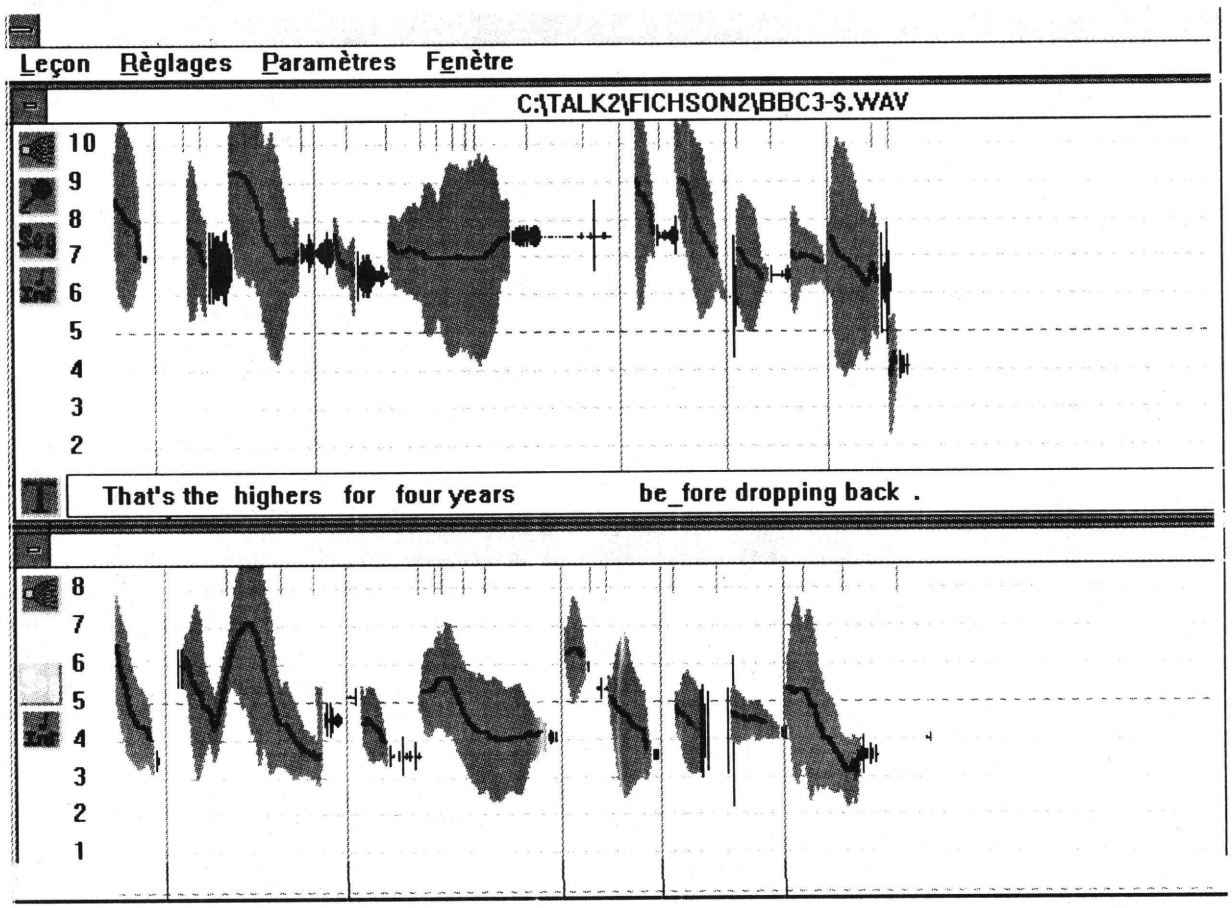


Fig. 2. Prosodic representation of a model sentence of four seconds and of the corresponding production of the learner.

prosodic representation. For reasons of research and development, we also developed a fully manual segmentation system which can cut and present any segment defined by the Learner by moving the cursor to the target area in the figure. With the prosodic figuration acting as a permanent image of sound phenomena, we use this as an access key to oral productions. The advantages of referencing and the easy access to oral productions provided by this mode of access compared to the tape recorder should be noted: the Learner can situate the target segment in regard to the whole sentence and get access simply through pointing. This mode of command is concrete and, in addition, can be used for the tracing and anchoring of auditory perceptions which, by nature, are evanescent.

But this fully manual segmentation system is not suited to the usual presentation of significant sound segments to the Learner. It requires interpretation of the prosodic figuration which acts as an access key and leaves the Learner with the burden and difficulty of pertinent and significant segmentation. For, it should be understood that an oral result cannot be cut just anywhere on the speech signal; auditory segmentation is the result of complex perceptive processes involving recognition of words, syntactical analysis and sometimes even interpretation. Therefore we have to try to simulate auditory segmentation, i.e. segmentation into syllables, words and even phonemes which the speaker represents naturally. Yet auditory segmentation is not simply "given" by the speech signal.

Phonological studies show that linguistic units are not easy to identify from classic acoustic parameters. The inter-word links introduce continuities which do not correspond to the lexical cutting. Inversely, the constitutive silences of certain consonants (occlusive consonants: p, t, k) are not interpreted as natural limits in the speech signal. A random cutting of the speech signal could produce incomprehensible sounds or unexpected phonemes (the end "sse" produces "te").

To simulate auditory segmentation, we defined and developed two levels:

- a level of pre-segmentation corresponding to any instant when a segmentation could have a phonetic sense (approximately the syllabic level). These instants are referenced by a short line on the figurations of the oral productions.

- a level of segmentation itself corresponding to larger phonetic groups (1, 2 or 3 syllables) inside which the cutting is less obvious. These instants are marked by a vertical bar. The learner can still perform relative segmentation as long as he remains in the context of pre-segmentation at his disposal.

For segmentation, as we have conceived it from the auditory viewpoint, to be properly represented, the instants which define the start of a segment must:
i) correspond to important breaks
ii) not be situated in an occlusive consonant or at the start of a voiced consonant when this is preceded by an occlusive consonant.
The pre-segmentation instants correspond to minimum levels of intensity. Among these, we have selected, for segmentation, those for which the variation speed of the following maximum intensity is higher than a threshold.

Since automatic segmentation is based on signs that are purely acoustic, it is not as systematic as one might wish and systematically cuts between the syllables or between the words, for example. We also developed a "Segmentation" command at the level of the interface which provides manual correction of segmentation by switching between the segment terminal and the pre-segment terminal. A segment limit is deleted by clicking in front of it; a new segment limit is validated by clicking in front of a pre-segment. These commands returns an auditory feedback, that provide a control of the induced effect.

Without giving a detailed evaluation of automatic segmentation, we note on figures 1 and 2 that it provides syllabic segmentation. Pre-segmentation is rather overabundant in order to detect all the inter-syllable transitions. As for segmentation, it has a small number of segments (3 to 6 for a sentence of two seconds) which most often corresponds to the words or syllables. The segmentation in figure 2 is the result of automatic segmentation in a rough state, while that of the model sentence and its repetition by the Learner, in figure 2, have each been modified five times (five segment deletions). Segmentation is stored with the model files so that the author of tutorial can define the segmentation he wishes to show his student and free him from any concern as regards segmentation on models.

An Interface For Direct Manipulation of the Different Presentations

To enable the Learner to mobilize all his resources for perception and production of the stressing and intonation schemas of the language II, we developed and multiplied all the presentations which provide useful information. To prosodic representation, we added a representation of the alone intensity which could be useful for underscoring stress. The first evaluation of our system showed that assistance could be necessary to develop more elementary motor skills and reach the targets of this learning more efficiently. While our assistance was targeted on the sentence (lasting a maximum of 5 seconds) and required a processing time of 3 to 10 seconds, it could be thought that a visual feedback that was simpler - but in real time - would have led to better learning of the control of voice parameters. The presentation in auditory modality of our visual feedback of prosody could also be envisaged.

But if we want to mobilize all the resources of the Learner, it is essential for all the available presentations to be closely interconnected with each other. As such, the textual presentation is taped in and adjusted manually to help the understanding of prosodic representation. Since this is visually segmented, it is also the key to access to the sound segments. While the selected sound segment is being played, the corresponding visual segment is displayed in reverse video. The Human Computer Interaction literature has coined the term « directness » for interfaces that give the user the feeling to manipulate the underlying objects. We have not only ensure a feeling of directness about oral productions but we have assembled all the visual presentations around the segmented auditory presentation.

We also tried to encourage the differential listening of two oral productions and evaluation of the Learner through reference to the model. The successive presentation of the corresponding segments of the two oral productions dis-

played on the screen is obtained by pointing in the corresponding visual segments of the upper and lower windows. This differential comparison, based on prosodic representations, leads to a more objective self-evaluation than if the Learner was left to his perception which is clearly disabled from the outset.

This close correspondence no doubt explains why the subjects easily accept our complex prosodic representation because they receive its immediate and concrete illustration in the form of the sound segment.

It would be interesting to know if the prosodic representations are essential as an image of intonation and if they favour the effective acquisition of prosodic schemas. But to evaluate these representations and their efficiency, it would be necessary to include this tool in a course with defined aims. Although we have already established a technical evaluation of our help systems (Denoeud, 1996), we still have to evaluate this tool inside a pedagogical approach and see how a teacher would use it in interaction with a Learner.

CONCLUSION

In order to explore the problematics of the acquisition of underlying perceptuomotor skills and the production of stressing and intonation schemas in a foreign language, we outlined a cognitive analysis of the learning activity which showed that the specific features of perceptive learning were not usually based on cognitive representations. We have come to the conclusion that the representations can be useful, at least on a temporary basis, to support learning activities. But the target of developing perceptuomotor skills requires a connection to be established between perceptions and representations and between representations and motoricity. In order to support perception activities, we proposed a training environment in which these phenomena are given a material representation which expresses subjective sensations and objectifies the productions. The prosodic representation which we proposed is original in that it endeavours to represent, in visual modality, the perceptive integration of the parameters of the acoustic substance. On the other side of motor acquisition, our help system provides a means of evaluation which enables the Learner to perform self-evaluation that is more objective than his auditory perception, which is often disabled at the start.

A representation of prosody formed by signs extracted from the acoustic pressure signal certainly has imperfections. But we observed that the learners make an "intelligent use" of the help system; they do not expect a computer to provide perfect performance under all circumstances. They appear to respect a « principle of reality » which gives auditory reality priority over the representation. This fact, which still has to be confirmed, is an encouragement to design "intermediate representations" to support perception activities; we shall nor have to wait for some considerable time before having models of perception that can be fully implemented.

BIBLIOGRAPHY

Boulakia G., Gautheron B. (1987) Visualisation de l'intonation: application à l'enseignement, in Cling M. et Humbley J., Eds, *Actes du 3 ° colloque d'Avril sur l'anglais oral*, 1987, 12-22.

Cutler A., Mehler J., Norris D., Segui J. (1983) Language specific comprehension strategy. *Nature*, 304, 159-160.

Denoeud B., Raquillet M., Foti A. (1996) Aider à l'apprentissage des phénomènes prosodiques en situation d'autoformation, *Actes du congrès ERGO - IA*, Biarritz, oct 96.

Roach (1983). *English Phonetics and Phonology*. Cambridge University Press.

Rossi M., Di Cristo A., Hirst D., Martin P., Nishinuma Y. (1981). *L'intonation, de l'acoustique à la sémantique*. Klincksieck, Paris.

Scotto di Carlo (1989) Les sensibilités internes phonatoires. *T.I.P.A.*, 12,71.

The application of a combined strategy for the interaction between a memory impaired person and a wordprocessor

Javier A Bajer
Cognitive Ergonomics Research Group
Loughborough University
Loughborough
Leicestershire LE11 3TU
Tel: (01509) 234 362 *Fax:* (01509) 234 362
EMail: J.Bajer@lboro.ac.uk

ABSTRACT

The purpose of this paper is to address performance problems encountered by a memory impaired subject while interacting with a wordprocessor. An assessment of the subject and her ability to perform tasks was conducted. An iterative model was created applying solutions that include both the adaptation of the tool and development of the subject's skills. This paper concludes that as word-processing tasks require a diversity of skills, the strategy for enabling interaction should include a combination of theories and approaches and should be sensitive to the particular characteristics of the memory impaired user and the evolution of his or her skills.

Keywords

human-computer interaction, disabled, memory impairment, word-processing

INTRODUCTION AND OBJECTIVES

The main objective of this project was to study the interaction problems between an amnesic patient (user) and a word processor (tool) and develop a strategy to overcome those problems.

As seen in the model in Figure 1, there is a "gap" represented between the user and the tool. This "gap" was defined in terms of user's skills, motivation, tool characteristics and, of course, performance objectives. Briefly, the user's objective was to operate a word-processing tool at a level comparable to that prior to the accident.

In order to bridge this "gap" different mechanisms were put in place, some of them with the intention of building the user's skills in relation to the tool, while others adapted the tool towards the special needs of the user. At this point it was very important to be able to define how much effort was required in each direction. Questions regarding level of effort included:

- How many new skills can the user assimilate?
- How personalised will the tool be after adaptation?
- Will we be isolating the user to the use of just this one tool?
- If we don't do that, will the user ever be able to use a word processor?

The decisions for the appropriate solutions were influenced by current models of rehabilitation from literature research, interviews with field experts and subject's professional carers, task and memory assessments and an ongoing Usability evaluation of the interaction.

Different models were applied, adapting the tool in some cases and training the user in others. The tasks were performed iteratively until appropriate solutions were reached.

CONCLUSIONS

There are several theories and models that have proved to be effective when addressing specific problems in the memory rehabilitation field. The intention of this study was to put into practice these strategies in order to enable the interaction between a memory impaired person and an off-the-shelf wordprocessor. Although different, the memory rehabilitation and learning theories are not necessarily contradictory. Real life skills, like using a wordprocessor, require a variety of skills and the best approach to achieve those skills is through a **combination of strategies.**

Amnesia is not a binary pathology that either exists or does not, but rather a continuum. The combination of different strategies certainly could be used on a much wider population.

The application of multiple models or theories to the solution of interaction problems should be approached in a very flexible way, as conditions change as a result of each interaction, requiring the replacement, addition or withdrawal of solutions.

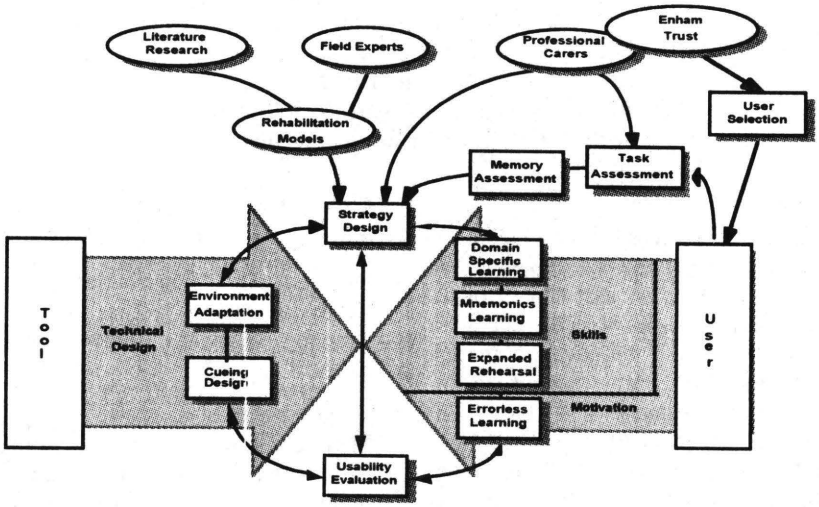
The creation of prosthesis or the development of skills should not be mutually exclusive. On the contrary, this research shows how a prosthetic solution could be used to create an appropriate environment for the development of skills. Furthermore, these new skills allowed a later withdrawal of the initial prosthetic solutions.

The findings in this study represent particular needs of a specific subject. These findings should not be generalised, as it is not possible to "prescribe memory aid X for memory problem Y" (Harris, 1992).

Finally, Word for Windows™, though a very powerful tool, proved to have considerable requirements in both *computer and human memory*, although requirements for this latter could not be measured easily.

REFERENCE

Harris, J.E., 1992, Ways to help memory. In B.A.Wilson and N.Moffat (eds.), *Clinical Management of Memory Problems.* London, Chapman & Hall, P. 59-85.



Interaction model - Fig. 1

Evaluation Of Iconic Representations Of Patient Records In A Medical Workstation

L. Pellegrin [12], V. Méjean[1], N. Bonnardel[1], H. Chaudet[23]

[1.] CREPCO - URA 182, U.F.R. de Psychologie et Sciences de l'Education,
Université de Provence, 29 av. R. Schuman
13621 Aix-en-Provence Cedex 1, FRance

[2.] Service d'Information Medicale, Hopital Secteur Nord
Chemin des Bourelly
13 915 Marseille Cedex 20, France

[3.] Laboratoire de Biomathematiques, Statistiques et Informatique Medicale
Faculté de Médecine de Marseille, 27 boulevard J. Moulin,
13385 Marseille Cedex 5, France

ABSTRACT

The main goal of this study was to compare the users identification of medical records objects described by mixed (with pictures and text) or pictorial icons. The results indicate that the identification score for mixed icons is significantly better than for pictorial one. There is also an interaction effect between groups and icons for this score. There were no significant effects on the group response times from either kind of icons. This experiment confirms the global effect of users' general preferences for mixed icons but shows also an inverse effect on icon identification and time due to a specific expertise in existing medical databases.

Keywords

icon design, medical records, user performance, expertise, evaluation methods.

INTRODUCTION

This work is concerned with the design and the choice of icons in human-machine interfaces applied to medical informatics. Our first goal was to evaluate icons representing different objects in medical records created for the interface of a medical workstation project. A record contains a set of medical information concerning the patient during his stay in an hospital.

This informationis quite varied: clinical observations, biological results, letters, diagnosis reports. They are used during the patient consultations by the physicians and the nurses and updated by the secretaries.

At present there are no specific recommendations for creating icons for medical software, especially concerning patient records, except the general ISO design norms (De Carolis et al, 1995). But, there are some basic foundations for creating icons (Byrne, 1993; Familant & Detweiler, 1993): they must have a strong metaphor (Carroll et al, 1988; Sébillote, 1991), be discriminable (Tversky and Gati, 1978), and only use graphics, without text (Apple Computer, 1992).

On the other hand, empirical studies indicate that mixed icons (with pictorial and text) are most meaningful than pictorial icons for the users (Guastello and Traut, 1989).

Our experimental goals have been to compare the subjects recognition of the object described by mixed or pictorial icons and also, the users identification preferences concerning the different iconic representations of each medical object.

MATERIAL AND METHODS

An experiment was conducted in a Medical Information Service. Forty-two different icons were created which represented 9 objects defined in an earlier project as specific of the patient record by medical experts as, for example, patient records, patient records set, letters, questionnaires.

Each object was represented by a mixed icon, that is to say with a pictorial and textual design, and by only pictorial icons. 32 subjects participated in this experiment. They belong to 4 different professional groups: senior or junior physicians, nurses and medical secretaries. 21 icons (pictorial or mixed) were randomly presented to each subject in a SuperCard® application.. Participants were asked to identify the object represented by each icon. Written responses and reactions time in seconds were recorded.

For interpreting verbal responses, a score (from 0 to 1) was constructed which represented the quality of icons identification. Reaction times would measure the subject cognitive working for this identification task.

RESULTS AND CONCLUSIONS

Global results of this experiment confirm the Guastello and Traut results.

Identification score for mixed icons is significantly better than pictorial ones [$f(1,664) = 11.25152$, $p = 0.008$]. There was no significant difference between subject groups [$f(1,664) = 0.46575$, $p = 0.706$] but an interaction effect between groups and icons [$f(1,664) = 4.16867$, $p = 0.006$] was observed. The medical secretaries scored better for pictorial icons than mixed ones, in contrast to the three other groups which had best recognition of medical objects for mixed icons.

This effect should be a consequence of computer expertise of subjects because they are the most frequent users of medical databases and classical office softwares. In this case, mixed icons has been no help for the identification of the objects.

There are no statistical effects on the response times between groups [f(1,664) = 1.511684, p = 0.210] or icons [f(1,664) = 0.124455, p = 0.724] even if average time for mixed icons are shorter than for pictorial ones. There are also no statistical interaction effects between groups and kind of icons [f(1,664) = 2.402, p = 0.067]. These results indicate that subjects take a longer time for the most meaningful icons, which was opposite to our expectations.

We have compared the total of points responses given for the set of icons of each object separately. Results show that the adequate recognition of object illustrated by icons may be modulated by the experimental factors as icons or groups. For example, the letters mixed icons were better [f(1,72) = 18.000. p = 0.0000648] and more rapidly recognized than graphics ones [f(1,72) = 3.867160. p = 0.053]. Also, letter icons are identified better by senior physicians, secretaries and junior physicians than by nurses [f(1,72) = 3333, p = 0.0241]. Three different icons of letters were proposed to the subjects, the second one receiving the best identification score.

We found statistical differences in scores and time measures between icons only for a few objects such as patient records set or questionnaire.

All these results gives us some user preferences which will have to be studied. Especially, we need to improve the pictorial quality of the icons. Such a study shows also that the design of icons needs to build specific methods of interface evaluation which use both such appreciation score and more refined time reaction measures.

BIBLIOGRAPHY

Apple Computer Inc (1992) *Macintosh Human Interfaces Guidelines*, Addison Wesley Publishing Compagny, Reading, MA

Byrne D. (1993) Using icons to find documents: simplicity is critical, In *Proceedings of INTERCHI' 93*, 446-453.

Caroll J. M., Mack R.L., Kellogg W.A. (1988) Interface Metaphors and User Interface Design, *Handbook of Human Computer Interaction*, Elsevier Science Publisher B.V.

De Carolis B., De Rosis F., Errore S. (1995) A user-adapted iconic language for the medical domain, *Int. J. Human-Computer Studies*, 43, 561-577

Familant M., Detweiler C. (1993) Iconic reference : evolving perspectives and an organizing framework, *Int. J. Man-Machine Studies*, 39, 705-728.

Guastello J. S. , Traut M. (1989) Verbal versus pictorial représentations of objects in a human-computer interface, *Int Man-Machine Studies* , 31, 99-120.

Sébillote, S. (1991) Théories et méthodologies : décrire des tâches selon les objectifs des opérateurs. De l'interview à la formalisation, *Le travail humain*, Vol 54 (3).

Tversky A. and Gati I. (1978) Studies of similarity, in Rosch E., Lloyd B. Eds, *Cognition and Categorisation*, Hillsdale, NJ : Erlbaum, 79-98.

Applying a User-Centered Design approach to an Information Kiosk for Spanish Social Security

Jesús Merino
Tesorería General de la Seguridad Social
CENDAR
28850 Torrejón
Spain
+34 1 503 86 05

Miguel Quesada
Andersen Consulting
Torre Picasso
28020 Madrid
Spain
+34 1 596 60 00
jose.miguel.quesada@ac.com

ABSTRACT
The Spanish Social Security Card Project includes several computer applications. One of them, the so called TAIG (which stands for Terminal Automático de Información y Gestión, meaning Automatic Management and Information System) is a touch-screen based kiosk intended to provide access to social security, employment and healthcare information.

TAIG has been designed for direct use by Spanish citizens. Therefore usability issues of this application are critical. This paper explains the design process followed.

Keywords
smart card, user-centered design, conceptual model, user interaction model, usability testing.

INTRODUCTION

Two years ago, the Spanish Social Security System started a project, known as the Card Project, focused on changing the membership document from a paper-based document to a smart card. There are several institutions involved in the project including: General Treasury of Social Security, National Institute of Social Security, Maritime Social Institute, National Institute of Employment and Andalusian Healthcare Institution.

The project is a very innovative one as the Smart Card International Association (SCIA) recognized. The SCIA conceded the "Outstanding Smart Card Project for 1996" award to Spanish Social Security Card Project during the Card Tech/Secur Tech 96 Congress held last May in Atlanta. This congress is recognized as the most important smart card congress in the world. Former winners of the award are US Department of Defense and German Ministry of Healthcare, amongst others.

This card increases membership document image and durability, serves as a personal identification document for the Social Security System, and facilitates citizen access to information. This information belongs to three different areas: Social Security, Employment and Healthcare.

There are several computer applications associated with the Card including a Healthcare and the TAIG application. TAIG is a touch-screen based kiosk intended to provide access to information in the areas mentioned above.

TAIG is an application designed to be used directly by all Spanish citizens providing general information, personal information, certificates, employment management, change of physician, etc.

TAIG success is closely related to citizens' acceptance and understanding of the application information and functionality.

There is an access control system based on biometrics data (information of the personal fingerprint is stored in the card) to guarantee that nobody can access someone else personal data. This has been the security option chosen in the project because it is safer than the personal identification number (PIN).

Therefore, TAIG includes a fingerprint recognition device requested when the citizen wants to access personal data or to carry out personal operations (obtaining certificates, changing of the physician, etc.).

PILOT EXPERIENCE

The pilot experience of the project was held in the province of Cordoba (Andalusia) in 1995. During the pilot phase, an amount of 500.000 cards were issued, and 19 TAIGs were available in different locations of the institutions involved in the project.

The experience was very useful and allowed the project team to establish the feasibility of the project and to fully understand the process. Regarding TAIG the following results were obtained:

- Citizens got excited about TAIG; they wanted to use it.
- The application "was hard to use" for novice users.
- Information was "hard to read on-line" and the preferred way for obtaining information was print-out.
- The application performance was very slow.
- Citizens didn't understand very well the way the information was arranged.

After the pilot, the project team decided to completely redesign the kiosk using a quite different approach including usability concepts based on a user-centered design philosophy.

METHODOLOGICAL APPROACH

As mentioned before a user-centered design approach was followed. According to Rubin (1994), the main characteristics of this approach are:

- Early involvement of the user.
- Empirical measurement of application's usability.
- Iterative design process including designing, prototyping and testing.

There are two main tasks packages in the usability methodology to be considered:

- Gather and Analyze Usability Requirements.
- Design and Evaluate User Interaction Model.

We are going to explain them in more detail.

Gather and Analyze Usability Requirements
This is the first task package in the methodology and is focused on understanding the users and their tasks, identifying existing users' mental models and defining usability objectives of the application.

The kiosk is intended to be used by 40 millions people (Spain's population). The Spanish "average" citizen is a person with no computer experience, small functional knowledge of social security, employment and healthcare, and not familiarized with computer peripherals like mice or keyboards. The best way to approach this audience was through a multimedia kiosk with a touch-screen as the only input device.

As mentioned before Spanish citizens on average are not familiarized with this kind of applications , so no previous Mental Models were expected. However, during the Pilot it was realized that many citizens tried to apply a 'cash dispenser' mental model to the kiosk.

There are two main Usability Objectives to consider, when designing the application:

- Subjective Satisfaction (through a pleasant aesthetic appearance). The kiosk should be appealing enough to capture citizen attention. This is specially critical when he/she may decide whether or not to use it.
- Learnability. Once the citizen is placed in front of the kiosk Learnability becomes essential. First time usage was the main concern, as there is no possibility of train the users.

Other Usability Objectives like Speed/Throughput, Errors Prevention and Memorability were less critical in the design of this application.

Design and Evaluate User Interaction Model
When defining the Conceptual Interaction Model of the application it was considered that Navigation should be both very explicit and simple. Navigation was an essential issue, because it was necessary to clarify the way to navigate forward and backward into the application. Special visual aids (Navigators) were created to recall the novice user where he/she was in the sequence and to provide a faster way of navigation for experienced users. Navigation was also aided by appropriate use of contextual helps, including spoken helps.

Displaying detailed information was also very important because it was demonstrated during the pilot phase, that sequential reading was not appropriate. According to Horton (1994), on-line documentation is a new media, as different from paper, as television is from books.

It was necessary to change reading style (to a more conceptual one) and the detailed information metaphor by using a "book metaphor" instead of the "page-turning metaphor" used during the Pilot.

The User Interaction Model we chose was based on "Natural Interfaces" (based on natural objects like buildings, rooms, books, etc.) and multimedia as the best choice to deal with novice users without previous computer experience.

Prototyping was started in a very early stage of the design, according to user-centered design philosophy. The first prototypes were actually low-fidelity prototypes (including paper-based), more focused on the User Interaction Model than on the accuracy of the design details. It was critical to know if users were able to understand the interface new approach.

Usability tests proved that that was in the right direction, although some improvements were introduced.

We continued this iterative process using more and more sophisticated prototypes and making several tests with many average users. As a result of these tests many design considerations which led to the final design of the TAIG were obtained.

CONCLUSIONS

User-centered design is the only valid approach when designing multimedia highly interactive applications intended to be used by a high population of users. It is also very important to have a Pilot experience in order to anticipate potential usability problems of the application.

REFERENCES

Horton, W. (1994). *Designing and Writing On-line Documentation.*

Rubin, J. (1994). *Handbook of Usability Testing - How To Plan, Design, and Conduct Effective Tests.* John Wiley & Sons, Inc.

Informal Communication is about Sharing Objects and Media

Judith Ramsay, Alessandro Barabesi and Jenny Preece
The Center for People and Systems Interaction
South Bank University
103 Borough Rd., London, England
Tel. + (0) 171 815 7414
E-mail: {ramsayja, barabea, preecej}@sbu.ac.uk

ABSTRACT
This position paper presents late-breaking findings about informal communication over desktop video-conferencing software. Evidence is presented in support of users coupling media, and in support of video providing some form of shared object perspective (previously known as the 'video as data' hypothesis (Whittaker, 1995)).

Keywords
computer-mediated communication (CMC), desktop video conferencing, informal, shared object, media-coupling.

INTRODUCTION
Computer-mediated communication (CMC) can be textual or an audio visual communication infrastructure involving cameras, monitors, microphones and speakers in users' homes or offices. This allows communication with others who are geographically distributed. CMC tends to be less formal in nature than CSCW and can be used in work or non-work situations.

Of particular interest to us is the extent to which the media are coupled with one another, in interactional units, during computer-mediated communication. An example of this is the following:

User One initiates: "Can you see me?" (Typed)

User Two responds: "Yes" (Spoken)

(In fact, our desktop video conferencing system (CU-SeeMe) couples audio and video (audio/visual), but in this study we focus primarily on audio.)

In addition we believe that focusing on shared objects and their contextual role might provide a fruitful way forward. This reflects Whittaker's (1995) view that work in desktop video conferencing has neglected the importance of shared objects and their role in a shared context.

The aim of this research is to investigate: (i) the way that different media are used, and (ii) the requirement for shared objects and operational spaces in informal communication via desktop video conferencing.

Rationale and set-up
The interactional cues that video transmits are likely to play a more important role among people who know each other, and who are motivated to carry out the communication task (Isaacs and Tang, 1994). We thus investigated informal communication, which is representative of individuals who know one another, and to increase motivation and ecological validity, individuals selected their own issue for discussion.

CU-SeeMe
CU-SeeMe is a desktop video conferencing system designed by Cornell University for use on the Internet or other IP networks. It runs on Macintosh and Windows platforms. Video transmission requires a camera and digitizer. It displays 4-bit grayscale video windows.

METHOD
Ten pairs of subjects (five male-male, one female-female, and four male-female pairings) were video-taped interacting with one another using CU-SeeMe. The pairs knew one another, but all had little experience of using the software. The subject sample comprised students and professors. Each pair discussed a small problem of personal significance. One individual helped the other in reaching a solution to the problem, or in defining a plan of action. Examples of problems discussed included:

- My neighbor's television keeps me awake
- One of my colleagues is slowing my work rate
- I want to create an on-line, distributed book.

Apparatus and procedure
The subjects were seated in two separate rooms furnished with a PowerPC with a 14" monitor. A camera (Connectix QuickCam) and a microphone were mounted on each monitor. The internal speakers were used. The audio channel was a 64 Kb/s full duplex. Video was transmitted through the ethernet at a maximum speed of 500 Kb/s. Each subject was provided with a self view window, a remote view window and a text window (known as the talk window) to send text to the remote partner. The monitor output was connected to a video-recorder through a computer-to-video scanner in order to videotape the interactions.

RESULTS
The taped interactions were examined for instances of: (i) choice of media during interaction, and (ii) the need, if any, for shared objects and operational spaces.

1. Mixed media exchanges
Seven out of the ten pairs occasionally engaged in some form of media mixing or coupling. As the video link is

omnipresent in the interaction, and usually accompanies audio channel usage, the media coupling we discuss here refers to audio/visual and textual links. Frequency counts revealed the following:

text (given): audio/visual (returned) - sixteen events;

audio/visual (given): text (returned) - nine events.

Eleven of the sixteen text-audio/visual pairings occurred when there was an audio problem such as feedback being received, and the users were arranging, via the text link, how to handle the audio without involving the audio channel in so doing. The remaining five instances were of the following nature: Person A types (text): Person B interrupts (audio/visual). This tended to occur in scenarios in which the person using text embarked upon a detailed or lengthy description, which the other person then interrupted.

Not only did users couple different media types, they occasionally doubled up on, or overlapped their media use, for example talking whilst typing. This was true to a very large extent in three out of the ten pairs. The need to annotate the ongoing interaction is possibly linked to all three of these interactive sessions also showing the desire for some form of shared object.

2. Shared objects and spaces

A frequency count was made of the number of times a shared operational space was desired, either through verbal comment or evident physical activity. Five out of the ten pairs spontaneously displayed behavior indicating a need for some type of "shared activity" space. A shared activity facility was desired during tasks of the following nature: those requiring a shared record facility, those that were clarificatory in nature, and those requiring a shared authoring space.

DISCUSSION

Communication as subtle media matches

We observed that the audio/visual channel is used to override and interrupt a textual message from the other communication partner. This can be interpreted as a quick communication channel (audio) being used to draw the attention of the other individual using a slower communication channel (text).

Communication as sharing objects or spaces

Caldwell, Uang and Taha (1993) found contradictory findings in the literature, and the relation between task and media usage is still unclear. That the need for a shared record of events occurs, is a likely function of the fact that audio and video interactions are dynamic and transitory. Whereas a textual record, although it develops during the interaction, remains static enough on the screen to act as an external memory buffer. This explains the uptake of the textual channel in the form of a shared object. This is possibly indicative of a more general human propensity to share objects, especially non-textual, abstract representations and pictures and diagrams.

CONCLUSIONS

This paper presents late breaking work, which will lead to our characterizing the dimensions that are needed to make CMC systems support users well; particularly in terms of media, shared objects and spaces.

Our position is that even in informal communication users display a strong natural need for some form of shared external memory to record key decisions and points raised in their conversation. Shared authoring, editing and drawing/sketching tools are needed for creating documents such as letters or for explaining abstract concepts. The ability to show objects of discussion to a conversation partner also appears to be important.

In the next stage of our work we will replicate our study with more participants to see if these trends are generalizable. If so, then we will characterize the kinds of tasks for which such objects and spaces are required and develop prototypes which can be tested experimentally. We will also analyze the video in detail using play-back interview techniques (Monk et al., 1993) to elicit users' reasons for their media choices and need for sharing objects and spaces.

REFERENCES

Caldwell, B. S., Uang, S-T. and Taha, L. H. (1995) Appropriateness of communications media use in organizations: situation requirements and media characteristics. *Behavior and Information Technology*, 4, 199-207.

Isaacs, E. and Tang, J. (1994) What video can and cannot do for collaboration: a case study. *Multimedia Systems*, 2, 63-73.

Monk, A., Wright, P., Haber, J. and Davenport, L. (1993) *Improving your Human-Computer Interface: A Practical Technique*. New York: Prentice-Hall.

Whittaker, S. (1995) Rethinking video as a technology for interpersonal communications: theory and design implications. *International Journal of Human-Computer Studies*, 42, 501-529.

Effects of Noise on Driving Behaviour

Maria Teresa Bajo[1], José Juan Cañas[2], Francisca Padilla[3], Maria del Carmen Puerta[4]
Departamento de Psicología Experimental, Universidad de Granada,
Campus de Cartuja,
18071 Granada, Spain
+34 58 243767
[1]mbajo@goliat.ugr.es, [2]delagado@goliat.ugr.es, [3]fpadilla@platon.ugr.es, [4]mpuerta@goliat.ugr.es

ABSTRACT
Previous research has shown that noise may affect performance depending on the task cognitive demands. Driving is a complex task that encloses different subtasks that require a large amount of cognitive resources. In three experiments we explored the effect of different properties of noise on driving performance. Our results showed that intensity levels above 60 Db have a detrimental effect. This effect was more pronounced in detecting stimuli while driving the car. Furthermore, the presence of some type of music or conversation imposed on car's noise was less harmful than car's noise by itself.

Keywords
noise, driving, attention, music, safety

INTRODUCTION
The goal of the present study was to know the effect of intensity and meaningfulness of noise on driving behavior. While driving, different types of noises are present (those coming from the engine, the radio or the street). Previous studies have suggested that the most important parameters of the noise affecting performance are its intensity and variability. Intensities higher than 80 Db seem to impair performance in a variety of tasks. Also, low variability noises such as music or conversation seem to be less harmful than high variability noises (Broadbent, 1978; Loeb, 1986). But these effects are modulated by the type of task with more demanding tasks being more affected by the presence of noise (Kjellberg, 1990; Benke et al., 1993). Theoretical explanations assume that the presence of noise contributes to increase the level of activation. The difficulty of a task also contributes to increase activation. Therefore, performing a difficult task in the presence of noise may produce an increment in activation beyond the optimum level for performance (Ponsoda, 1983). This reasoning leads to predict that noise should produce an impairment in tasks as aircraft piloting or driving behaviour (Nuckells and Ng, 1995).

Driving is a complex task that requires a large amount of attentional resources, and therefore, may be affected by the noise present in our cars. When we are driving we perform different tasks simultaneously: maintain the car on the road, detect different stimuli that appear frequently while driving (e.g., traffic lights, a child running on the road), make a decision and respond to them. However, very little research has been done relating the presence of noise to driving performance.

We simulated the driving task in our laboratory. Our subjects simultaneously performed a tracking task and a detection task. Tracking involves skills similar to maintaining the car on the road. The detection task tried to simulate sign detection and detection of unexpected events while driving. Subjects had to follow a moving dot by manipulating an arrow with the computer's mouse. Below the dot a curved line depicted the trajectory to be followed by the dot. Depending on the condition, the dot and the line moved fast or slowly to simulate fast or slow driving. Simultaneously subjects had to detect and respond to a square or a circle that appeared randomly in a second computer screen at the right of the subject.

GENERAL PROCEDURE
Noise was presented through earphones. Subjects in the silent condition also had the headphones on. Six blocks of three minutes conformed the experimental sessions. For experiments 1 and 2, in the tracking task, blocks 1, 3, and 5 defined the slow condition and blocks 2, 4, and 6 defined the fast condition. In experiment 3, all six blocks were fast. In the detection task in the three experiments, subjects had to respond by pressing "1" or "2" on the computer keyboard depending on whether the circle or the square appeared on the screen. These stimuli appeared in a random sequence intermixed with a blank screen. To assess the effect of practice, these blocks were grouped (1-2, 3-4 and 5-6) to defined the first, second and third practice conditions.

Previous to the experimental sessions, we recorded the car noise, the music and conversation within a car in normal driving conditions.

All subjects were psychology students at the University of Granada. Ten subjects participated in each of the between group conditions.

EXPERIMENT 1. EFFECTS OF NOISE INTENSITY.
In this experiment we explored the effect of several noise intensities (80, 70, 60 50, and silence/40 Db) at fast or slow speeds.

Results and Discussion

Tracking task.

The dependent variable was the Average Absolute Error. The effect of practice and speed were significant, $F(2,90)=7.48$, $p< 0.001$ and $F(1,45)=3.21$, $p<0.001$ respectively. Speed and practice interacted $F(2,90)=10.22$, $p<0.003$, so that practice had an effect only in the fastest conditions. The effect of intensity level tended to be significant with high intensities producing larger number of errors, $F(2,90)=10.22$, $p=0.13$.

Detection task.

Reaction times to the stimuli were averaged for each condition. The effect of practice, $F(2,90)=8.64$, $p<0.01$ and speed $F(1,90)=8.56$, $p<0.005$ were significant. More important was the interaction between practice, speed and intensity $F(8,90)=2.82$, $p<0.01$. Reaction times were slower when subjects were exposed to high intensities in the fast conditions, so that the normal effect of practice did not appear.

EXPERIMENT 2. TYPE AND CONTENT OF NOISE.

This experiment was designed to study the effect of several types of noise superimposed on the car noise (classical music, rap music, conversation, 70 Db noise and silence).

Results and Discussion

Tracking task.

The effect of practice and speed was significant $F(2,90)=9.66$, $p< 0.001$ and $F(1,90)=3.63$, $p<0.001$ respectively. There were more errors in the first block and in the faster speed condition. More important, type of noise was significant $F(4,45)=2.93$, $p<0.04$. Interestingly, the presence of music or conversation from the radio produced levels of performance similar to the silence condition and better than the performance on the 70 Db noise condition.

Stimulus detection task.

Practice and speed were significant sources of variance, $F(2,90)=4.06$, $p<0.03$ and $F(1,90)=7.42$, $p<0.01$. The more important effect was the interaction between speed and type of noise $F(8,90)=2.05$, $p<0.05$. At slow speed, all noise conditions produced slower responses than the silence condition, although the effect of classic music was much smaller than any of the other conditions. At fast speed, there was an important increment in RT in the rap music condition. Therefore, the presence of music or conversation might help to reduce the effect of the background noise when the task involves skills similar to tracking. This beneficial effect reverses when the task involved fast detection and response to a stimulus. This impairment was specially marked with rap music.

EXPERIMENT 3. MUSICAL PREFERENCES.

This experiment was designed to explore whether the effect of music would appear even when the music is the preffered by the subjects. In an interview previous to the experiment, subjects selected Beatles songs as their preffered type of music.

Results and Discussion

Tracking task.

The effect of noise conditions on tracking was not significant.

Stimulus detection task.

Again high variability music showed a detrimental effect on stimulus detection task. When the car noise had a intensity of 70 Dd, music impaired performance significantly, $F(4,42) = 2.49$, $p< .05$. With 80 Db showed a tendency that did not reach significance, $F(4,42) = 2.10$, $p = .09$.

GENERAL DISCUSSION

Results showed that intensity levels above 60 Db have a detrimental effect on performance. This effect was more pronounced in the detection task. In addition, the presence of classic music or conversation superimposed on car's noise was less harmful than car's noise by itself. However, high variability music (rap or Beatles songs) did not eliminate the detrimental effect of car noise. These results are important for theories that attribute the effect of noise to attentional demands; at the same time they suggest possible advice to drivers and cars' manufacturers to increase safety while driving.

REFERENCES

Benke, T., Koserenko, O., Watson, N.V. Gerstenbrand, F. (1993). Space and Cognition: The measurement of behavioral functions during a six-day space mission. *Aviation, space and Environmental medicine, 64*, 376-379.

Broadbent, D. (1978). The current state of noise research. *Psychological Bulletin. 85*, 1052-1057.

Kjellberg, A. (1990). Subjetive, behavioral and psychophsiological effects of noise. *Scandinavian Journal Enviromental Health. 16* , 29-38.

Loeb, M (1986). *Noise and Human Efficiency. Studies in Human performance*. London: John Wiley.

Nuckells, W. and Ng, Wf (1995). Fan noise reduction from a supersonic inlet during simulated aricraft approache. *J. of Engineering for Gas Turbines and Power Transactions of the ASME,*

Ponsoda, V. (1983). Explicaciones de los efectos del ruido en el rendimiento. *Revista General de Psicologia General y Aplicada. 38*, 1083-1115.

Extraction of Interface Quality Measurements from Dialogue Description Notations

Pedro M. Valero-Mora
Departamento de Metodología, Psicobiología y Psicología Social, Universidad de Valencia
C/ Blasco Ibáñez, 21 CP 46010. Valencia
Email: valerop@uv.es

ABSTRACT
An approach for extracting an interface complexity indicator from dialogue description notations, prior to construction of the latter, is demonstrated. This indicator is in turn tested in an empirical study, where a degree of predictive capacity is demonstrated.

Keywords
dialogue representation. interface complexity. methodology.

INTRODUCTION
After determining the requirements of a human operator-computer interface, the designer needs to specify his or her conceptions by means of a representation capable of allowing evaluations for the introduction of posterior improvements. Many different notations and representation methods are available as tools in application to these two activities (i.e., representation and evaluation), based on the following classification (Dix et al., 1993): system models, dialogue representation methods, and user models. Of these, the dialogue representation methods are based upon the representation of an interactive system, while user models evaluate it from the user's perspective.

While a number of attempts have been made to take advantage of the best of both worlds (e.g., Kieras and Polson (1985) proposed two notations that could be executed in parallel, thus providing exchanges of actions and responses), such a distinction would oblige the designer interested in using formal notations to learn and use one of each type. Thus, the user models (e.g., PUM, Ayn) analyze an interface specified by some other - indeterminate - method. On the other hand, the dialogue representation methods are limited to either formal evaluations (reachability, completeness, etc. Olson et al., 1995), or to step-by-step execution in obtaining an impression of their function. Another possibility for carrying out evaluations based on dialogue descriptions is to employ guidelines with which the system may be contrasted. However, these guidelines pose the inconvenience of possessing a qualitative and static form that leaves the evaluation of their importance and applicability to the intuition of the designer. This contrasts with the efforts made to perform quantitative predictions, as in user models such as NGOMS.

The present study explores an approach by which quantitative indicators of the complexity of an interface may be derived based on dialogue description, thereby informing the designer of those situations that may be regarded as "critical" and thus susceptible to redesigning. In addition, an empirical study is carried out in which the execution of a group of subjects is predicted using these indicators.

Number of options as an indicator of complexity
As proposed by Norman (1988), we may presume that a given task possesses a broad decision structure when the number of options to be considered for taking a decision is large. In contrast, in the case of a narrow structure, the decisions are sequenced, and each action is chosen from a group of more reduced alternatives. The number of options may thus be regarded as a complexity index for each decision, and can be used as a predictor - for example, of the time that will be required by the subjects to initiate the correct action.

However, the determination in a graphic interface of the number of user-available options is a rather complicated matter. This is particularly so when the interface has not yet been constructed. This problem was addressed by means of a tool known as an "Action Simulator" (Monk and Curry, 1994), which implies an extension of the "Action-Effect" rules proposed by Monk (199). This tool causes the different states in which a dialogue is found to correspond with the global set of actions available in that state.

METHOD
The program analyzed
A program was analyzed, consisting of a simulation of five tasks that may be encountered in a word-processing context (saving files, modifying a header, etc.). In order to carry out these tasks, a succession of actions was employed, such as pressing keys, dropping menus and responding to dialogue boxes, in accordance with the Macintosh style of interface.

Likewise, minimal documentation was elaborated, indicating how to perform these tasks. Although in fact each task constituted a different module, the documentation made reference to the program as a whole. The correct sequence of actions was pre-programmed for each of these modules; as a result, when the user specified an incorrect action the system either failed to respond, or (in the event of repeated incorrect actions) exhibited a message specifying the correct action.

Action-Effect analysis

The program (specifically, each of the program modules) was described by means of an Action Simulator, which allowed the counting of available actions in each state.

The test situation

Twelve students in their first year at the Faculty of Psychology (Valencia, Spain) assimilated the performance of all five tasks. The subjects read a small manual that introduced them to the different tasks contemplated by the program. The section corresponding to each individual task was read prior to its accomplishment.

Each task was carried out in its basic aspects (varying only points such as the text to be typed) on at least three consecutive occasions by each subject. If upon concluding the three attempts the subject was still unable to carry out the sequence of actions without error (i.e., the performing of an action out of the correct sequence), he or she was instructed to attempt the task on further occasions to secure at least one correct execution per subject and task. The interactions were recorded along with the times required by the users to specify each action.

RESULTS

Data transformation and statistical analysis

As the distribution of the dependent variable exhibited a marked positive asymmetry, its logarithm was used to the effect of analysis. A variant of intra-subject regression analysis (Pedhazur, 1982) was employed in our study, known as "pooled time series analysis", which contemplates the utilization of subjects as random factors rather than as fixed factors - a practice well known in Econometrics (Maddala, 1971; Green, 1993). This model also makes it possible to eliminate the effects of intra-subject autocorrelation.

Predictor variables

The predictor variables employed were the subjects themselves (introduced as a fictitious variable), the number of task attempts (generally one to three, but as many as four or five in some cases), the order in which the task was performed (one to five), whether the attempt in question was the first or not (to control for the fact that the first attempt was considerably more difficult than the rest - thereby introducing a curved line effect) and, finally, the number of options available for each action performed by the subject.

Model fitting

The variance accounted for by a model addressing the subjects as random component was approximately 10%. This indicates that the model considered is incomplete and that it lacks important variables if our aim is to completely account for the dependent variable under study. However, this consideration is less important than the demonstration that the variable "number of options" possesses statistical significance ($t = 9.271$, $p < 0.001$). Figure 1 shows the prediction fit to the logarithms of the empirical points (above) and residuals' distribution (below). The residuals are seen to exhibit a satisfactory behaviour.

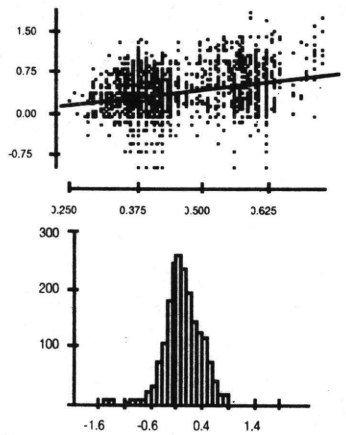


Figure 1. Diagnostic graphs testing the fit of the model from the analysis of variance.

CONCLUSIONS

The derivation of usability indicators by means of a dialogue description notation of an interface still being designed may constitute a useful strategy, as it is more neutral than user's models about the possible analyses to be performed.

The present study shows that a measurement as simple as the number of options to consider in order to specify a given action may be calculated from a dialogue description, and correlated to the time required by the subjects to specify each action.

REFERENCES

Dix, A., Finlay, J., Abowd, G., & Beale, R. (1993). *Human-Computer Interaction.*: Prentice Hall.

Green, W. H. (1993). *Econometric Analysis*: MacMIllan.

Kieras, D., & Polson, P. G. (1985). An approach to the formal analysis of user complexity. *International Journal of Man-Machine Studies, 22*, , 365-394

Maddala, G. S. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica, 29*, 2, 341-358.

Monk, A. (1990). Action-Effect rules: a technique for evaluating an informal specification against principles. *Behaviour and Information Technology, 9*, 2, 147-155

Monk, A., & Curry, M. B. (1994). Discount dialogue modelling with Action Simulator. In G. Cockton, S.W. Draper and G. R. S. Weir (Ed.), *People & Computers, 9-Proceedings of HCI'94* . Cambridge University Press.

Norman, D. A. (1988). *La Psicología de los objetos cotidianos*. Nerea.

Olsen, D. R., Monk, A., & Curry, M. B. (1995). Algorithms for automatic dialogue analysis using propositional production systems. *Human Computer Interaction, 10*, 1, 39-78.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston.

Painters' Attitudes Towards Internet Art Presentations

Alexander E. Voiskounsky
Psychology Department, Moscow University after M.V.Lomonosov
8 Mokhovaya str., Moscow 103009
Russia
+7 095 203 3409
vae@motiv.cogsci.msu.su

ABSTRACT
Art presentations are becoming more and more common on the Internet. For the classical art, the providers are responsible for the adequacy of presentations. For the modern art, the creators themselves keep all the responsibility. Predispositions of 14 modern Russian painters towards presentations of their art pieces on Internet are empirically investigated. The research method used is free interview with fixed questions. The subjects had no idea of Internet presentations, and the explanations provided by the experimenter were of solely verbal nature. This is due to the fact that the Internet services are not common in Russia. The results show that the painters' basic attitudes towards presentations of their art works on Internet tend to be positive. The motivations for this tendency are the aesthetic motivation, the prestigious motivation, and the motivation of self-expression. The reasons for negative attitudes expressed include the need for a feedback and the need to control the adequacy of the viewers' projections of the pieces of art.

Keywords
attitudes, motivation, painting, internet, art presentation

INTRODUCTION
Internet services are developing rapidly, and art presentations are now an integral part of Internet. Separate pieces of art and/or art collections are presented by museums, by smaller galleries or studios, and by painters themselves. Among the pieces presented one might have access both to old and to modern art. The OASYS Network, for example announces it has over 1000 freelance professionals in the "fully queryable database", and provides free promotional cyberspace to every professional artist or graphic designer (see http://www.oasysnet.com). The exhibition sites are becoming more and more innumerous. The research and analytical work in the field is in progress; for example, the research article by M. McLaughlin is issued ("published") in the electronic *Journal of Computer-Mediated Communication*, which is available at
[http://www.huji.ac.il/www_jcmc/jcmc.html]
or else at
[http://cwis.usc.edu/dept/annenberg/announce.html].

To restrict the review to only Russian and to only modern art, several newest sites might be named. The Sovereign Collection Fine Art Gallery introduces the Website "for Netscape browsers with high resolution screens" with the 19th & 20th centuries paintings, sculpture, etc. of American/Russian/Dutch/French/ German/Scandinavian/British origin (see http://www.teleport.com/~art4sale/index.html).

The Institute of Commercial Technology (Moscow) presents an exhibition of painting and graphics of Andrei Bil'zho who is named among the best in caricature graphics (see http://www.fe.msk.ru/). The Lili Brochetain Collection of Nonconformist Russian Art presents the new Internet site for the exhibition "The New Russian Avant-Garde (from Brejnev to Gorbachev's Perestroika)" - see [http://www.stria.ca/Brochetain]. The list of Websites might be continued.

Recently, an ambitious project of presenting the best of modern art produced in Russia has been initiated in Moscow: museums, galleries. and painters personally are welcomed to present slides from their best pieces and textual comments - all this to be digitized and put into World Wide Web sites located in Moscow. The project is in progress just now: the technical problems are mostly resolved, but the organizational, the art expertize and the human factor problems seem to be the most actual ones.

To restrict the discussion with the human factors problems, we might note specifically that quite a sample of modern (living now) painters/sculptors/ designers/folk-art craftsmen seem to believe that art making and presentation techniques have nothing to do with the informatics and computer science. For the most part, they are ignorant about the WWW, and their unwillingness (sometimes expressed rather eloquently) to present their pieces in the Internet has much in common with the technostress or computerophobia (Brod, 1984) problem.

RESEARCH AIMS AND METHOD
In Russia, there are no massive research projects dealing with the attitudes towards informatics, or Internet, or computers in general. Some sporadical cross-culture studies hint that the attitudes towards computers are very close, taken Russian and American children (Martin et al., 1992). The Russian population is found to express quite strong readiness towards the advance of information database systems (Voiskounsky & Neumann, 1993). . In the Russian human factors and cognitive ergonomics paradigm the research field on positive/negative consequences of computerization is rather popular. The research is usually restricted with separate human activities (managerial, or

productive, or teaching, or recreational, etc.) to be computerized: unwillingness of the sample members to use the systems, or else the benefits of the computer systems introduced are analyzed in order to come to valuable recommendations aimed to better the design and the introductive procedures of the information systems (Yaroshevsky & Tikhomirov, 1986).

Close to this paradigm, to work out the methodology needed to convince the painters in the usefulness of the Internet presentations, we started with the case study of the painters' attitudes towards presentations of their pieces in the WWW. The attitudes were measured while interviewing 14 painters, all Muscovites, aged between 27 and 52, composed of females (5) and males (9). All of them are professional artists, and 11 earn their living with art and/or applied art exclusively. The interviews were held during Spring-Fall, 1995. All the subjects were interviewed in a gallery in the downtown of Moscow, individually.

RESULTS AND DISCUSSION

Results show that there are 3 separate groups of subjects: those with positive attitudes (8 subjects), those with negative attitudes (4 subjects), and intermediates (2 subjects). To express it differently, non-positive attitudes are characteristic for about a half of painters.

Positive Attitudes

The reasons and motivations for the positive attitudes are like the following. Some subjects (5) consider the Internet presentations to be alike a new media for ads: they believe advertising is always useful and needed. Besides, they are fond of a possibility to change presentations easily, to expose both older and newer pieces of art, all the art manners characteristic for a painter. Two painters believe the WWW presentations are highly prestigious (the elite of the world community will have access to the pieces of art presented). One of the subjects believes that art is supposed to aesthetise each segment of the Universe: if the WWW forms such a segment, painters have to insert pieces of art into it to make this segment more beautiful.

Negative and Ambivalent Attitudes

The reasons and motivations for the negative attitudes are like the following. One of the subjects believes the Internet presentations lack the feedback channel, i.e. he has no chance to learn the viewers' comments and estimations, which form the most precious part of all the traditional expositions and presentations for him. Though there are fairly good technical possibilities to express the comments (via e-mail or each of traditional communication channels), the subject has expressed his dislike for mediated communications and an obvious preference for face-to-face discussions with the viewers. The other 3 subjects do not believe in proper color presentations of their pieces. Taken each potential WWW subscriber who might have access to the modern Russian art pages, no one is able to control the monitor characteristics of his workstation. Thus, improper monitors will give inadequate color images, and this is really an unbearable situation for the subjects. In order to avoid the situation, they express negative attitudes towards presenting art pieces on the Internet.

The two intermediates expressed almost the same contradictions, and confessed that they can give their graphics pieces for Internet presentations, but not their paintings. Color combinations they think are irrelevant for graphics, but significant for oil painting. Thus, they express only partly positive attitudes.

Competence in Informatics

The other result is that the more competent in informatics are the subjects, the less positive attitudes they express. That is, those who have no (or almost no) background knowledge and personal experience with computers usage, unlike those competent in informatics, are not able to advance any realistic obstacles and drawbacks that might lead to negative effects of the Internet exhibitions. That is, computer literacy/illiteracy is a major factor in the methods of introduction of new information technologies and the Internet services into artistic communities.

Thus the results show that the painters' basic attitudes towards presentations of their art pieces on the Internet tend to be positive. The motivations for positive attitudes are: the aesthetic motivation, the prestigious motivation, and the motivation of self-expression. The reasons for negative attitudes include the need for a feedback and the need to control the adequacy of the viewers' projections of the pieces of art. The subjects' background knowledge in informatics gives rise to realistic reasons that may lead to negative attitudes towards the art presentations on the Internet.

ACKNOWLEDGEMENTS

The research was supported by the Russian Foundation for Fundamental Research, project "Creative Activity and Information Technologies".

REFERENCES

Brod, C. (1984). *The Human Cost of the Computer Revolution*. Reading, Mass.

Martin, C.D., Heller, R.S. & Mahmoud, E. (1992). American and Soviet children's attitudes towards computers. *Journal of Educational Computing Research*, 8(2), 155-185.

M.G.Yaroshevsky & O.K.Tikhomirov (Eds.) (1986) *Psikhologicheskie Problemy Avtomatizatsii Naucno-Issledovatel'skikh Rabot* (Psychological Problems of R&D Informatization) Moscow: Nauka Publ. (In Russian).

Voiskounsky, A.E. & Neumann, M.G. (1993). Psychological readiness of Russians for information technologies usage in press. *Proc. of the International Conference on Information Technology and People*, Pt. 1. Moscow, 59-64.

Margaret L. McLaughlin. The Art Site on the World Wide Web. In: *Journal of Computer-Mediated Communication*, vol. 1, #4; available at:
`http://www.usc.edu/dept/annenberg/vol1/issue4/vol1no4.html` or
`http://shum.huji.ac.il/jcmc/vol1/issue4/vol1no4.html`

Human Factors Engineering with the PUMA (Performance and Usability Modelling for Air Traffic Control) Tool

Clive P. Warren
Sowerby Research Centre
British Aerospace plc
FPC 267, PO Box 5
Bristol, BS12 7QW, UK
tel +44 117 936 6988
fax +44 117 9363733
Clive.Warren@src.bae.co.uk
http://bel.avonibp.co.uk/studios/clive

Paul O. Day
Roke Manor Research Ltd.
Roke Manor
Romsey
Hampshire, SO51 0ZN, UK
tel +44 1794 833216
fax +44 1794 833433
Paul.Day@roke.co.uk

Mike K. Hook
Roke Manor Research Ltd.
Roke Manor
Romsey
Hampshire, SO51 0ZN, UK
tel +44 1794 833303
fax +44 1794 833433
Mike.Hook@roke.co.uk

ABSTRACT
The authors have developed, under contract to the UK Civil Aviation Authority, a tool which supports the construction of predictive models of UK en-route air traffic controller workload. The output of this work was a methodology and supporting tool. The PUMA tool allows the rapid assessment and first-cut filtering of proposed changes to ATC procedures and computer assistance to the Air Traffic Controller prior to more detailed real-time simulation. The paper describes our approach and demonstrates its applicability to UK en-route ATC workload assessment and prediction.

Keywords
air traffic control (ATC); human factors; computer assistance; system development; performance modelling; workload; PUMA.

INTRODUCTION
The domain of ATC is one area in which Human Factors has much to offer in terms of the integration and development of computer systems intended to support operator's tasks. With the planned harmonisation of European ATC systems and procedures, and rapid advances in technology, ATC workstations of the future will introduce increasing computer assistance to many aspects of controllers' tasks. This paper describes a new approach to the evaluation of future ATC systems which can be carried out either before or after their implementation in the workplace. By incorporating Human Factors expertise both before and during the design process of new ATC systems, it is hoped that the cognitive compatibility between controllers and their supporting systems will be maximised. The new approach, Performance and Usability Modelling technique for Air Traffic Management (PUMA), and its supporting tool are described in this paper. The work on the PUMA tool was funded by the United Kingdom (UK) Civil Aviation Authority (CAA), and is being used in the evaluation of future en-route ATC systems and working practices. The PUMA tool has been implemented initially on a Macintosh™ computer system, supported by the Network Modelling Software Environment (NMSE) software developed by Roke Manor Research Ltd (RMRL) in the UK.

TECHNOLOGY AND THE DOMAIN OF ATC
Surprisingly, the ATC workplace has not had the same level of introduction of new technology as has been seen on the flight decks of modern passenger aircraft. This situation is set to change, mainly due to the large growth in air traffic since the 1980s. World-wide estimates of future air traffic levels made in the 1980s were very conservative. It is now estimated that, by the end of the century, air traffic levels in Europe will be double those of the late 1980s, with similar estimates for the USA (Cadogan, 1990). ATC authorities have now developed new strategies to deal with the current bottlenecks which constrain the movement of air traffic (e.g. the concept of a European Air Traffic Management System, (EATMS, 1992) . These strategies include the modernisation of computer and radar equipment, and the introduction of new routing systems such as the Central Control Function (CCF) in the UK, and area navigation (RNAV) throughout Europe.

CONTROLLER TASKS AND WORKLOAD
For the task of controlling en-route or Terminal Manoeuvring Area (TMA) airspace, controller workload is generally seen to be the primary limitation on ATC system capacity. Controller workload is made up both of the tasks required directly for control of individual aircraft, and also associated tasks such as communications, co-ordination and liaison functions. In ATC systems running at or near their planned capacity, the individual controller can become burdened by high task demands during the controlling activity (see e.g. Hopkin, 1991). As air traffic continues to grow over the next decade, without measures to increase system capacity, or constrain traffic growth, there will be increasing pressure on controllers.

In order to understand the implications of enhanced computer assistance, new equipment, or new procedures on controller workload, it is essential to have both a reasonable means of representing the controller's tasks, and also some means of modelling workload, such that reasonable predictions may be made as to how proposed changes in equipment and procedures affect controller workload and performance.

Task Modelling using PUMA
The approach taken in PUMA to representing the controller's activities is similar to that adopted by Lenorovitz & Phillips (1987). The controller is modelled as "an event-sensitive multitasking information processor". This approach has a clear advantage for subsequent modelling and analysis activities, in that a series of *task events* can be identified which together form the controllers' task (Warren, 1992). In the PUMA methodology, this approach is reflected in the Hierarchical Task Analysis (HTA) activity which is one of the first steps in the production of a PUMA model. Furthermore, it is possible to both validate and supplement this approach not only by consulting with controllers and other domain experts, but also by conducting an Observational Task Analysis (OTA) of controllers performing their tasks, and then relating observed tasks to predicted tasks. The observational task analysis approach used in PUMA involves the controller talking through a videotape recording of his task activity immediately

after the shift has ended, thereby allowing an insight into exactly what was carried out, and the underlying cognitive activity associated with that part of the task. This technique is known as a "cognitive walk-through". The cognitive walkthrough is also recorded on video tape. The controllers' activities are then analysed in terms of the *actions* and *objects* being acted upon, additional data provides the time at which the activities occurred, and their duration.

Evaluating Workload

While the methods described above will provide a good understanding of controller tasks, it is also necessary to understand the workload associated with elements of the task. The study and measurement of workload has absorbed the attention of psychologists and human factors engineers for many years. While an intuitively obvious concept, its measurement has been approached in a variety of ways, including subjective measures, secondary-task measures, and physiological measures (see Gopher & Donchin, 1986 for a review). We use the concept of workload in those circumstances where task demands may exceed a person's capacity to deliver, where individual aspects of the task are normally within that person's capacity to deliver. In terms of modelling workload, at one time it was thought that the human operator could be viewed as a single processor with a limited information processing capacity, and that different tasks loaded this processor to different degrees, leaving in some cases some spare capacity. While this seemed inherently reasonable, and was attractive from an information theory viewpoint, a growing body of evidence during the 1960's and 1970's led to the conclusion that this was too simple a view. This prompted the development of what are now classed as *multiple-resource pool models* (eg Wickens, 1983), where people are seen as possessing a number of processing mechanisms, which compete for resources in order to undertake tasks. BAe have specialised in the modelling of workload over a period of years, and developed a methodology that incorporates this approach, known as the Aircrew Workload Assessment System (AWAS) which is embodied in the PUMA methodology and tool.

OVERVIEW OF THE PUMA METHOD

The broad approach involves five phases of activity. These phases are:

Phase 1: Hierarchical task analysis (to understand the structure of the tasks undertaken by controllers);

Phase 2: Observational task analysis (to verify how the task is carried out in practice);

Phase 3: Task-workload assignment (to understand how each task phase can be viewed in terms of controller workload);

Phase 4: Model consolidation and validation (to ensure that the model is complete, usable, and predictive);

Phase 5: Use of the model to predict workload in hypothetical ATC scenarios.

In order to make the models of ATC behaviour easy to use, our approach incorporates the use of computer-based simulation techniques to generate appropriate task sequences for any hypothetical ATC scenario. These task sequences can then be fed into the workload assessment component to generate predicted workload profiles against time for that scenario. The PUMA tool displays colour graphics depicting how workload varies as the sequence of ATC tasks is undertaken. In this way, it is possible to assess the effects of the introduction of new computer assistance for the task, or new organisations of controlled airspace.

ACKNOWLEDGEMENTS

The PUMA tool was developed in the UK Civil Aviation Authority funded "FEATS Case Studies Project" by Roke Manor Research Ltd and British Aerospace plc, Sowerby Research Centre. The views of the authors do not necessarily reflect those of the UK Civil Aviation Authority.

REFERENCES

Cadogan, A. (1990) - Clearing the Traffic Jam. *Flight International*, 17-23 January 1990, pp 33-35, Reed Business Publishing, England.

EATMS (1992) - European Air Traffic Management System Concept Document, Eurocontrol Document, Issue 1.1, Dated 16 June 1992, Brussels.

Gopher, D. and Donchin, E. (1986) - Workload - an examination of the concept. In Boff, KR, Kaufman, L, and Thomas JP. (Eds), *Handbook of Perception and Human Performance - Volume II: Cognitive Processes and Performance*, Wiley, UK.

Hopkin, V.D. (1991) - The impact of automation on Air Traffic Control. In Wise, JA, Hopkin VD, and Smith ML. (Eds), *Automation and Systems Issues in Air Traffic Control*, Springer- Verlag

Lenorovitz, DR and Phillips, MD (1987) Human factors requirements engineering for air traffic control systems. In Salvendy, G., (Ed), *Handbook of Human Factors*, John Wiley, New York.

Warren, C.P. (1992) - Automation in ATC and aviation: incorporating Human Factors in the system development process for safety critical systems. In Proc. The Limits of Automation in ATC and Aviation (Sienna, November 25-27, 1992), in Press.

Wickens, C.D. (1983) - Processing resources in attention. In Beatty, J. and Davies R., (Eds), *Varieties of Attention*, Parasuraman, John Wiley and Sons, New York.

Modelling of Operator Team Behavior in Nuclear Power Plants

Seiichi Yoshimura and Naoko Hasegawa
Human Factors Research Center
Central Research Institute of Electric Power Industry
2-11-1 Iwado Kita Komae-shi
Tokyo 201
Japan
+81 3 3480 2111
yoshi@criepi.denken.or.jp, naoko-h@criepi .denken.or.jp

ABSTRACT
An operator team of a nuclear power plant behaves based on the alarms and related parameter indicators on a control panel during off-normal conditions. Therefore, the human machine interface (HMI) is very important from human error's point of view. Human modeling may be one of the good ways for the analysis and evaluation of human errors. The modeling in this paper was done including the human characteristics such as characters, arousal level, knowledge and so on. In addition, emotions which influence the human behaviors were investigated for the future improvement of the human model.

Keywords
nuclear power plant, human machine interface, human modeling

INTRODUCTION
A nuclear power plant is a large system, and is supported by computer systems from plant start up to shutdown, especially supported during off-normal conditions. The computer technologies are progressing rapidly. Therefore, the task of operators is changing from operation to watching because of the introduction of such computer technologies. This tendency is expected to rise in the future as the automation proceeds and the human errors are also expected to change in quality under such situations.

The importance of HMI has been recognized recently, for the necessity of accordance between humans and machine systems. The more the accordance is desired, the more the analysis of the humans and the efforts to bring the machine systems close to the humans need to be conducted. It is considered that one of the good ways to analyse the humans is human modeling. Traditional human modeling (Cacciabue, 1991) was conducted by modeling a human. However, a nuclear power plant is operated by an operator team, i.e. by a few operators who have different tasks in a control room. Therefore, considering the human error analysis in an actual environment, human modeling should be done not only from individual point of view but also from team's point of view. In addition, human characteristics including emotions are necessary for the precise modeling of a human because human errors are thought to be influenced by such characteristics (Scheier and Carver, 1981).

This paper describes the outline of an operator team behavior model (Yoshimura et al., 1994) with human characteristics except emotions and also describes the investigation results of emotions and their influences to the cognitive process of an operator.

OPERATOR TEAM BEHAVIOR MODEL
Operator team behavior model is for simulating the behavior of a team. The model consists of an operator team, HMI and a plant. As for the operator team, it includes three operator behavior models, i.e. a leader (shift supervisor) and two followers (reactor and turbine operators). They obtain plant information through HMI and they make a team's decision by communicating each other through human human interface (HHI). Finally, they operate the HMI based on the decision. Thus, the plant situations are changed. In this way, the operator behavior model simulates the behavior of an operator and the HHI model simulates the communication between operator behavior models and the decision making process as a team.

Operator Behavior Model
The operator behavior model has four functions as an operator's cognitive process, i.e. "attention", "thinking", "action" and "utterance". A leader does not watch the HMI, nor does he touch it by himself because it is defined that he does not have the action function. He only hears the alarms through the attention function and he orders the followers to gather the relevant information. He thinks based on the information using the thinking function, where two kinds of processings - skill and knowledge - are included. Skill processing is defined as a fixed cognitive processing based on rules or defined as a reflective cognitive processing. On the other hand, knowledge processing is defined as a deeper cognitive processing in which situation understanding is done and counter operations are created based on the understanding.

The followers have action function and they move to watch the HMI according to their task allocation. They think based on the plant information obtained by themselves. They make a conversation with a leader when necessary and finally control the HMI.

Human Human Interface Model
The functions of the HHI model are utterance control, conflict resolution and task allocation. Task allocation of three operator behavior models is as follows, i.e. a leader considers the whole plant and a follower considers the reactor or the turbine side. Conflict resolution is for making a

team's decision when decisions by a leader's and a follower's are different.

Human Machine Interface Model
A control panel of a simplified BWR plant is included in the model.

Plant Model
Plant model is a simplified BWR type and consists of three main systems, i.e. reactor, turbine and condenser/feedwater systems. The model can simulate about 20 malfunctions. The scope of simulation is from 100% steady state to reactor scram.

SIMULATION TEST RESULTS

Simulation tests using subjects (experts and students) were conducted to verify the performance of an operator team behavior model. The model and the subjects coped with the same malfunctions, i.e. "increase of water temperature" and "leak at outlet of condensate water pump B".

The team behavior model took every action as it was taught. For instance, the model opened condensate water make-up valve after Hotwell water level low alarm was activated. However, the subjects both experts and students opened the make-up valve just after the condensate water pump B stopped. These results mean that the humans behave more predictively than the model does. Anyway, the team behavior model obtained information from a control panel autonomously and acted based on each operator's thinking, and took appropriate counter operations. It can be concluded that the team behavior model is almost appropriate from the above results.

EMOTIONS AND THEIR INFLUENCES

Emotions are considered to influence the cognitive process of an operator, thus the human errors may happen due to the emotions. Therefore, it is necessary to include the emotional function in the operator behavior model for the precise modeling of an operator. The function is currently supposed to be activated by at least one of such inputs as "information from a plant","operation conducted by an operator" and "human relations between operators". Investigation was conducted by answering the expected emotions and actions corresponding to a given input. Subjects were the people who had operation experience in nuclear power plants.

The results are as follows; (1) "Fear" is the most frequently expecteded emotion during off-normal conditions. (2) Human error in the thinking process such as "Cannot decide what should be done" was the most frequently expected. Thus, the kind of emotion which can be expected during off-normal conditions and the relation between inputs and emotions, emotions and expected human errors were made clear.

CONCLUSIONS

An operator team behavior model which is for simulating the behavior of a team was developed. The appropriateness of the model was verified by the comparison with subjects through some simulation tests.

Emotions and their influences were made clear. The emotional function in addition to other human characteristics will be included in the operator behavior model based on the investigation results in the near future.

REFERENCES

Cacciabue, P.C. (1991). COSIMO, A Cognitive Model in Blackboard Architecture. *Symposium of AI and Psychology, Reliability Engineering & System Safety*, Special Issure on Cognitive Science Approaches to Process Control.

Scheier, M. F., and Carver, C.S. (1981). Cognition, Affect, and Self-Regulation. *Seventeenth Annual Carnegie Symposium on Cognition*, 157-183.

Yoshimura, S., Sasou, K. and Takano, K.(1994). Team Behavior Simulation Model. *Journal of Scientific & Industrial Research*, 53,574-578.

INDEX OF KEYWORDS

activities 23
activity analysis 149
activity theory 103
air traffic control 149, 155, 161
alarm systems 133
art presentation 191
artefacts 63
artificial intelligence 173
attention 187
attitudes 191
automation 161

business goals 83

CAER 161
case-based reasoning 109
categorisation 109
cognition 155
cognitive architecture 97
cognitive demand 149
cognitive empathy 41
cognitive ergonomics 97, 103
cognitive modelling 95, 97
cognitive semantics 103
cognitive simulation models 97
commenting tool 1
complex systems 35, 97
computer use 77
computer-mediated communication 185
conceptual model 183
control 41
controller's picture 161
conversation patterns 63
cooperative dialogues 47
coordination 29, 41
critiquing 1

databases 11
design 5
design activity 89
desktop video conferencing 185
diagnosis 133
dialogue representation 189
disabled 179
distributed cognition 5
driving 187
dynamic situation 41
dynamic task allocation 155

education 167
emergency management 29, 57
employment service 11
epistemological analysis 127
ergonomics 173
error 97
error tolerance 35
ethnography 115
evaluation 5, 89
evaluation methods 181
experientialism 103
experiments 133
expertise 89, 181

framework revision 127

goal orientation 109
groupware 115
GTA 115

H-CI modelling 5
help system 173
human cooperation 35
human machine interface 195
human modelling 195
human reliability 35
human-computer interaction 71, 179
human-machine cooperation 155
hypermedia 71

icon design 181
information needs 11
information presentation 133
interactive media 167
interface complexity 189
interface design 167
interlocutionary logic 47
internet 191
intonation 173

knowledge 23
knowledge elicitation 95

learning programming 77
learning support 163

media-coupling 185
medical records 181
meeting 63
meeting facilitation 63
memory impairment 179
mental workload 1, 161
methodology 95, 189
modelling 195
motivation 191
multimedia 167
multiple tasks 57
music 187

narrative 167
noise 187
nuclear power plant 47, 133, 195

painting 191
paradigm shift 83
participatory design 95
PD1 161
perceived usefulness 1
PHARE 161
phonetics 173
pictorial aids 77
pilot task 149
planning and control 57
problem solving 139

regulation mechanisms 35
requirements engineering 63
reuse-support tool 109
revising 139

safety 187
shared object 185
shared representation 47
shift changeover 47
simulation 29
situation model 139
smart card 183
socio-technical systems 53
software design 139
software documenting 139
software quality 121
software reuse 139
spatial ability 71
stakeholder analysis 53
stress (in prosody) 173
stress (mental) 149
support systems 89
system design 53, 103

task analysis 115
task appropriateness 127
telephone-traffic monitoring 109
text processing 139
textbase 139
training 133
training needs 29

usability 11
usability engineering 127
usability evaluation 121
usability goals 83
usability testing 183
use of automation 149
user interaction model 183
user interface design support 1
user interface guidelines 1
user participation 23
user performance 181
user-centered design 183
utility analysis 83

ward monitoring systems 53
Wizard-of-Oz 1
word-processing 179
work organisation 23
worksystem 149

INDEX OF AUTHORS

Allwood, C. M. 11
Amat, A.-L. 149
Artman, H. 41
Axtell, C. 23

Bajer, J. A. 179
Bajo, M. T. 77, 187
Barabesi, A. 185
Bates, J. 53
Baurén, M. 1
Bellorini, A. 149
Benyon, D. 103
Bonnardel, N. 89, 181
Boy, G. A. 95
Bright, B. P. 63
Burkhardt, J.-M. 139
Burmistrov, I. 69

Cabrera, A. 83
Cañas, J. J. 77, 187
Chaudet, H. 181
Clarke, A. A. 17
Clegg, C. 23

Dahlbäck, N. 71
Day, P. 193
Deleuze-Dordron, C. 139
Denoeud, B. 173
Détienne, F. 139
Dowell, J. 29

Ericsson, M. 1

Falzon, P. 147
Fields, B. 5
Foti, A. 173

Goillau, P. J. 161
González, J. 83
Grant, S. 97
Grusenmeyer, C. 47

Harrison, M. D. 5
Hasegawa, N. 195
Hill, B. 57
Hoc, J.-M. 155
Hoeve, M. 115
Hollnagel, E. 133
Höök, K. 71
Hook, M. 193

Imaz, M. 103

Joblet, L. 89

Kane, R. 53
Kelly, C. J. 161
Kirby, M. A. R. 53
Kirwan, B. 133

Lemoine, M.-P. 155
Lenting, B. F. 115
Leonova, A. 69
Long, J. 57
Löwgren, J. 1

Maiden, N. A. M. 63
Méjean, V. M. 181
Merino, J. 183
Morvan, E. 35

Navarro, R. 77

Oppermann, R. 163

Padilla, F. 187
Patrick, J. 145
Pellegrin, L. 181
Perron, L. 109
Plowman, L. 167
Preece, J. 185
Puerta, M. del C. 187

Quesada, M. 183

Rae, J. 23
Ramsay, J. 185
Raquillet, M. 173
Rincón, M. 83
Rognin, L. 35
Rouet, J.-F. 139

Sauvagnac, C. 147
Sikorski, M. 121
Sjölinder, M. 71
Smith, W. 29
Soufi, B. 17
Sperandio, J.-C. 35
Stary, C. 127
Stoodley, K. D. C. 53

Thomée, S. 11
Totter, A. 127
Tyrer, J. R. 17

Valero-Mora, P. M. 189
van der Veer, G. C. 115
Vassie, L. H. 17
Visser, W. 109
Voiskounsky, A. E. 191

Wærn, Y. 1, 41
Warren, C. P. 193
Waterson, P. 23
Wright, P. C. 5

Yoshimura, S. 195