21ST EUROPEAN ANNUAL CONFERENCE ON

Human Decision Making and Control

EDITOR: CHRIS JOHNSON

GIST TECHNICAL REPORT G2002-1,
DEPARTMENT OF COMPUTING SCIENCE, UNIVERSITY OF GLASGOW,
SCOTLAND.

TABLE OF CONTENTS

WORKSHOP TIMETABLE

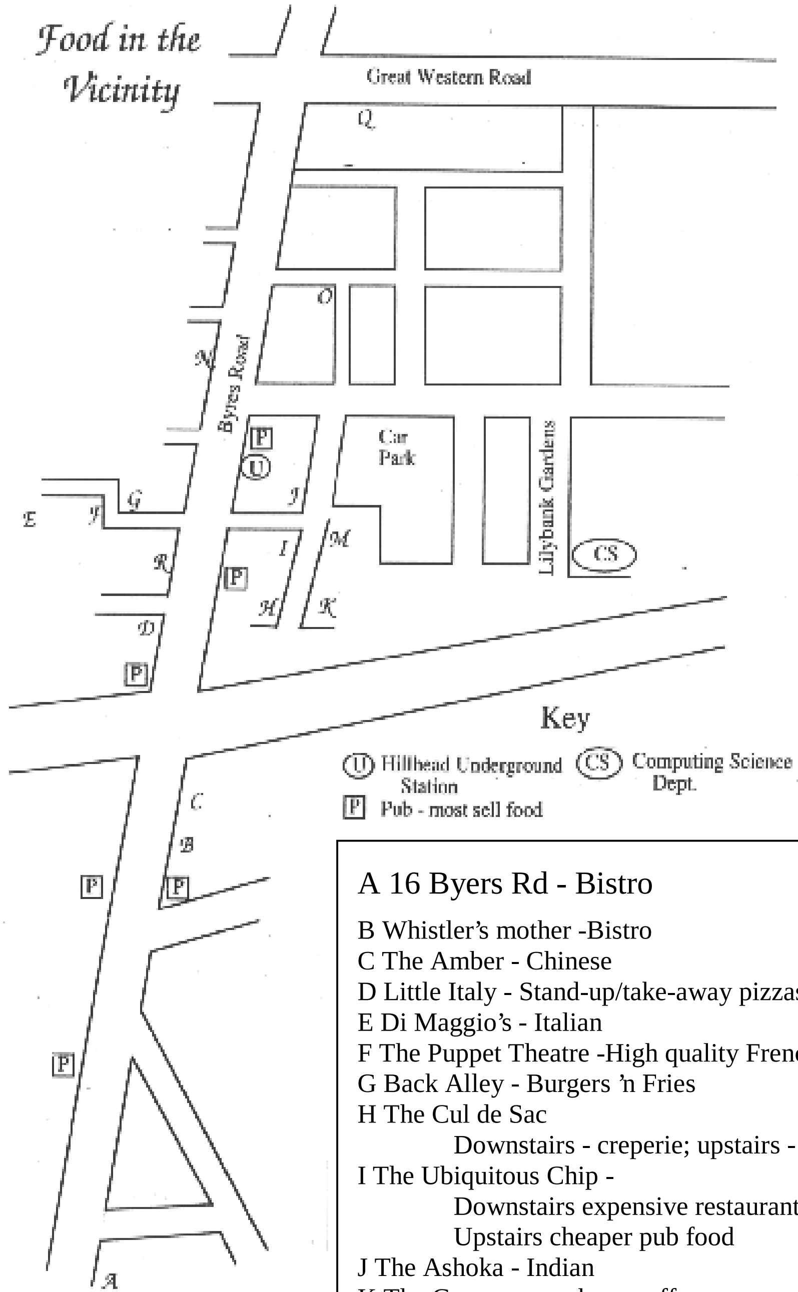
MONDAY 15TH JUNE

09.30-10.00 C. Johnson	Welcome and Introduction.
10.00-11.00 Chair: A. Pritchett, Georgia Institute of Technology	**Error Detection in Aviation** *Crossing the Boundaries of Safe Operation* N. Naikar & A. Saunders, Defence Science and Technology Organisation, Australia. *Activity Tracking for Pilot Error Detection From Flight Data* Todd J. Callantine, San Jose State University/NASA Ames Research Center, USA
11.00-11.30	*Coffee*
11.30-13.00 Chair: A. Sutcliffe, UMIST,	**Pilot Cognition** *Development and Preliminary Validation of a Cognitive Model of Commercial Airline Pilot Threat Management Behaviour* S. Banbury, Cardiff Univ, H. Dudfield, QinetiQ, M. Lodge, British Airways, UK. *Pilot Control Behaviour in Paired Approaches* Steven Landry and Amy Pritchett, Georgia Institute of Technology, USA. *Predicting Pilot Error: Assessing the Performance of SHERPA* N. Stanton, M. S. Young, P. Salmon, D. Harris, J. Demagalski, A. Marshall, T. Waldman, S. Dekker.
13.00-14.30	*Lunch*
14:30-15:30 Chair: T. J. Callantine, San Jose State Univ./NASA Ames Research Center	**Crew and Team-based Interaction in Aviation and Fire Fighting** *Coordination within Work Teams in High-Risk Environment, Effects of Standardisation* G. Grote and E. Zala-Mezö, Swiss Federal Institute of Technology, Zurich (ETH). *Assessing Negative and Positive Dimensions of Safety: A Case Study of a New Air Traffic Controller-Pilot Task Allocation.* L. Rognin, I. Grimaud, E. Hoffman, K. Zeghal, EUROCONTROL & CRNA, France. *Head-Mounted Video Cued Recall: A Methodology for Detecting, Understanding and Minimising Error in the Control of Complex Systems* M. Omodei, Latrobe Univ., J. McLennan, Swinburn Univ. of Technology, A. Wearing, Univ. of Melbourne, Australia.
15.30-16:00	*Tea*
16:00-17:30 Chair: E. Hollnagel, Univ. of Linkoping, Sweden	**Function Allocation and the Perception of Risk** *Tool Support for Scenario Based Function Allocation* A. Sutcliffe, J.-E. Shin, A. Gregoriades, UMIST, UK. *Time-Related Trade-Offs in Dynamic Function Scheduling* M. Hildebrandt, M. Harrison, University of York, UK. *An Examination of Risk Manager's Perceptions of Medical Incidents* M. Jeffcott, C. Johnson, University of Glasgow, UK

TUESDAY 16TH JUNE

09.00-09.30 C. Johnson	**Poster Summaries**
09.30-11.00 Chair: S. Bogner, Inst. for Study of Medical Error, USA	**Intensive Care and Surgery** *User Adaptation of Medical Devices: The Reality and the Possibilities* Rebecca Randell and C. Johnson, University of Glasgow. *Intelligent Systems in the Intensive Care Unit: A Human Centred Approach* M. Melles, A. Freudenthal, Delft Univ. of Technology, C.A.H.M. Bouwman, Groningen University Hospital, Netherlands. *Evaluation of the Surgical Process During Joint Replacements* J.P. Minekus, J. Dankelman, Delft University of Technology, Netherlands
11.00-11.30	*Coffee*
11.30-13.00 Chair: P. Sanderson, Univ. of Queensland, Australia.	**Constraints and Context Sensitivity in Control** *Preliminary Assessment of the Interface of an Anti-Collision Support System* P.C. Cacciabue, E. Donato, S. Rossano, EC Joint Research Centre, Italy *Designing Transgenerational Usability in an Intelligent Thermostat by Following an Empirical Model of Domestic Appliance Usage* A. Freudenthal Delft University of Technology, Netherlands. *Introduction in the Ecology of Spatio-Temporal Affordances in Airspace* An L.M. Abeloos, M. Mulder, M.M. van Paassen, Delft Univ. of Technology. *Modelling Control Situations for the Design of Context-Sensitive Systems* J. Petersen, Technical University of Denmark, Denmark.
13.00-14.30	*Lunch*
14:30-15:30 Chair: P. C. Cacciabue, EC Joint Research Centre, Italy	**Team Coordination and Competence** *A Formative Approach to Designing Teams for First-of-a-Kind, Complex Systems* N. Naikar, B. Pearce, D. Drumm, DSTO, P. M. Sanderson, Univ. of Queensland. *Crew Competence in Bulk Carriers* Steve Harding, UK Maritime and Coastguard Agency, UK. *Qualitative Analysis of Visualisation Requirements for Improved Campaign Assessment and Decision Making in Command and Control* C. Macklin, S. Cook, QinetiQ, M. Cook, C. Angus, C. Adams, R. Cooper, Univ of Abertay.
15.30-16:00	*Tea*
16:00-17:00 Chair: P. Wieringa, Delft University of Technology.	**Alarms, Barriers and Defences** *Model-Based Principles for Human-Centred Alarm Principles* S.T. Shorrock, Det Norske Veritas, UK, R. Scaife, A. Cousins, NATS, UK. *Toward a Decision Making Support of Barrier Removal* Z. Zhang, P. Polet, F. Vanderhaegen University of Valenciennes, France. *The Control of Unpredictable Systems* B. Johansson, E. Hollnagel & Å. Granlund, University of Linköping, Sweden.
17:00-17.15	*Close, hand-over to EAM 2003, presentation of best presentation award.*

Restaurants in the Local Area



A 16 Byers Rd - Bistro

B Whistler's mother -Bistro
C The Amber - Chinese
D Little Italy - Stand-up/take-away pizzas
E Di Maggio's - Italian
F The Puppet Theatre -High quality French
G Back Alley - Burgers 'n Fries
H The Cul de Sac
  Downstairs - creperie; upstairs - pub
I The Ubiquitous Chip -
  Downstairs expensive restaurant
  Upstairs cheaper pub food
J The Ashoka - Indian
K The Grosvenor - cheap caffe
M Jinty McGinty's - Irish pub & bar food
O The Metro - salads, hot dishes
Q The Parthenon - Greek food
R Burger King

Crossing the Boundaries of Safe Operation: Training for Error Detection and Error Recovery

Neelam Naikar and Alyson Saunders,

Defence Science and Technology Organisation
PO Box 4331, Melbourne, VIC 3001, Australia; neelam.naikar@dsto.defence.gov.au

Abstract: Widespread acceptance that human error is inevitable has led to the recognition that safety interventions should be directed at error management as well as error prevention. In this paper we present a training approach for helping operators manage the consequences of human error. This approach involves giving operators the opportunity to cross the boundaries of safe operation during training and to practise problem solving processes that enable error detection and error recovery. To identify specific requirements for training, we present a technique for analysing accidents/incidents that examines the boundaries that operators have crossed in the past and the problem solving difficulties they have experienced. This information can then be used to specify the boundaries that operators should be given the opportunity to cross during training and the problem solving processes they should practise. Initial applications of this approach have been encouraging and provide motivation for continuing further work in this area.

Keywords: human error, training, error management.

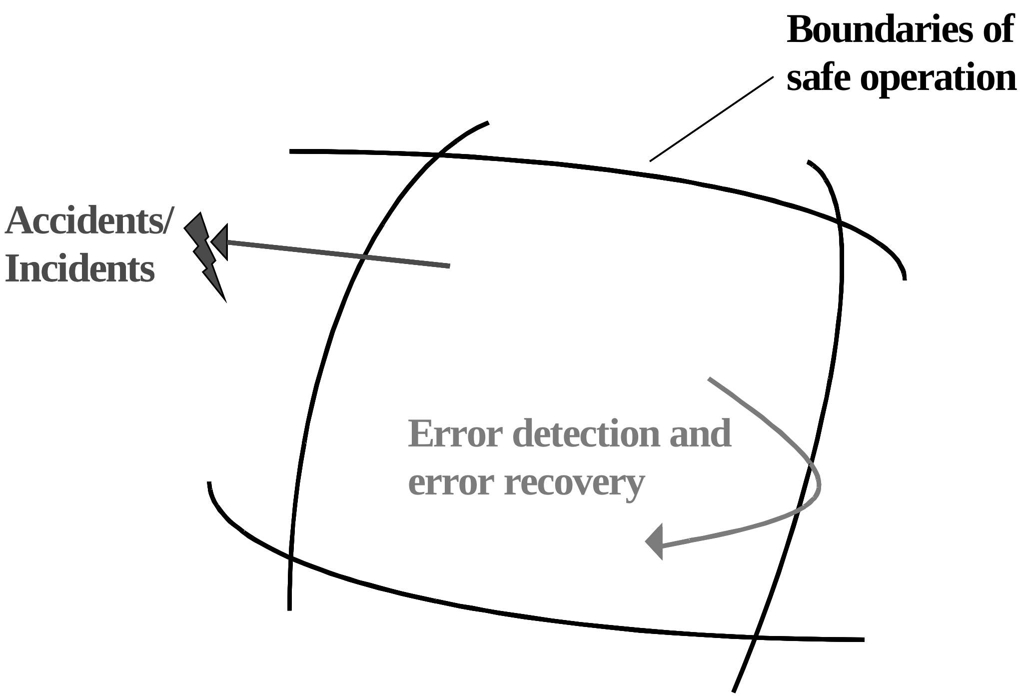
Introduction
There is widespread acceptance in the aviation community that human error is inevitable (Hollnagel, 1993; Maurino, 2001; Reason, 2000; 2001; Sarter & Alexander, 2000; Shappell & Wiegmann, 2000; Woods, Johannesen, Cook, & Sarter, 1994). An examination of any incident database will reveal a proliferation of errors involving, for example, incorrect switch selections, inadequate scanning of instruments, and inadequate cross-checking and monitoring. These kinds of errors are difficult to reduce or eliminate completely. The consensus therefore is that we must move beyond error prevention to helping aircrew manage the consequences of human error. Currently, safety interventions directed at error management include the design of error-tolerant systems (Noyes, 1998; Rasmussen, Pejtersen & Goodstein, 1994) and Crew Resource Management (Helmreich, Merritt & Wilhelm, 1999).

In this paper, we introduce a new approach for training aircrew to manage human error. This approach recognises, first, that although errors are inevitable, accidents are not. Second, although humans often make errors that threaten system safety, their ability to adapt to dynamic situations also makes them one of the most important lines of defence in averting an accident or incident once an error has occurred (Reason, 2000).

Interestingly, our observations of training in the Australian military indicate that they are rather good at training aircrew to manage *equipment* errors or failures (Lintern & Naikar, 2000). Many simulator-based training sessions involve the presentation of equipment failures to aircrew so that they can practise dealing with the malfunctions. Hence, aircrew develop well rehearsed processes for dealing with equipment failures if they occur on real missions. In contrast, little effort is directed at training aircrew to manage the consequences of human error. Instead, as is probably the case in many other organisations, the emphasis to date has been on training aircrew not to make errors in the first place. However, almost 80% of aircraft accidents are said to be caused by human error and many of these errors are difficult to eliminate completely.

The training approach that we have been developing to help aircrew manage the consequences of human error is based on Jens Rasmussen's conceptualisation of work systems as having boundaries of safe operation (Amalberti, 2001; Hollnagel, 1993; Rasmussen et al., 1994). Accidents or incidents can occur when operators cross these boundaries by making errors. However, crossing the boundaries is inevitable. That is, errors will occur on occasion. The emphasis in training therefore must be on error detection and error recovery. In particular, in the case of simulator-based training, operators must be given the opportunity to cross the boundaries of safe operation in the training simulator and to practise detecting and recovering from crossing these boundaries (Figure 1). Then, if the operators cross these boundaries during real operations, they are more likely to detect and recover from the error, and consequently avert an accident or incident.

Figure 1: Illustration of a work system as having boundaries of safe operation.



Applying this training approach can lead to novel ways of training. For example, consider the training of procedures in both commercial and military aviation. The most common practice is for aircrew to be drilled in executing the steps of a procedure until, it is hoped, they remember it and get it right every time. But a slip or lapse in executing some part of a procedure is inevitable, as an examination of any accident or incident database will show. Applying the training approach that we have developed implies that rather than simply drilling aircrew in executing procedures to minimise the chances of error, aircrew must also be given training in dealing with the situation that evolves if they make an error in executing a procedure. Thus, at least in a training simulator, aircrew should be given the opportunity to not follow a procedure or parts of a procedure, and to practise dealing with and recovering from the situation that evolves.

Some evidence of this training approach may be found in the Australian military, which indicates that it may have validity in operational settings. For example, aircrew are sometimes asked to place the aircraft in an unusual attitude and to practise recovering from this position. While this is very worthwhile, one of the problems in real operations is detecting when the aircraft is in an unusual attitude in the first place. So, our training approach would require that aircrew are also given practice at detecting unusual attitudes. In other cases, aircrew are given practice at detecting errors but not at recovering from errors. For example, a flying instructor acting as a navigator in a two-person strike aircraft may deliberately enter a wrong weapons delivery mode and check to see if the pilot detects the error. If the pilot does not detect the error, the flying instructor will alert the pilot to the situation and the weapons delivery mode will usually be corrected prior to the attack on the target. In contrast, the training approach we present here would require that the flying instructor leave the error uncorrected to give the pilot further opportunity to learn to recognise the cues that an error has occurred, and to practise dealing with the evolving situation.

In order to determine whether this training approach can be usefully applied to military aviation, we have started developing a systematic approach for identifying training requirements for managing human error. This approach relies on an analysis of accidents and incidents to examine the boundaries of a workspace that aircrew have crossed in the past, and the problem solving difficulties that aircrew have experienced in crossing these boundaries. Subsequently, this analysis can be used to identify requirements for training scenarios in terms of the boundaries that aircrew should be given the opportunity to cross in a training simulator, and the problem solving processes they should practise to enable error detection and error recovery.

Identifying Training Requirements for Managing Human Error

In this section, we present a technique for identifying training requirements for managing human error from an analysis of aircraft accidents. This approach involves three main steps. The first step is to identify the critical points in an accident. The second step is to use Rasmussen's decision ladder formalism to examine aircrew problem solving at each of the critical points. The third step is to generate training requirements to manage human error from the preceding analysis. We illustrate each of these steps by example below.

Identifying Critical Points: A critical point in an accident may be described as a crew action/non-action or a crew decision/non-decision, usually in response to an external or internal event, that threatens system safety. To illustrate, consider a hypothetical accident involving an F-111 aircraft in which the pilot executes a manoeuvre manually without first disengaging the autopilot. The pilot experiences difficulty executing the manoeuvre because the autopilot is fighting him for control of the aircraft. However, the pilot does not perform any corrective action. The autopilot disengages and produces an autopilot fail tone but the pilot fails to respond to the tone. As the autopilot disengages while the pilot is exerting high stick forces in order to gain control of the aircraft, the aircraft is thrown into a hazardous attitude and then hits the ground.

In this accident, the first critical point involves the pilot executing a manoeuvre manually without disengaging the autopilot and then failing to perform any corrective action. The second critical point occurs when the pilot does not respond to the autopilot fail tone. The third critical point occurs when the pilot is unable to recover from the hazardous aircraft attitude.

Examining Aircrew Problem Solving: This step involves using Rasmussen's decision ladder formalism to examine aircrew problem solving at each critical point. We chose the decision ladder over more traditional models of information processing because: all of the steps in the decision ladder need not be followed in a linear sequence; the decision ladder accommodates many starting points; and the decision ladder accommodates shortcuts, or shunts and leaps, from one part of the model to another. Thus, the decision ladder is a suitable template for modelling expert behaviour in complex work systems (see Vicente, 1999 for an extended discussion of these arguments). The decision ladder has also previously been used in accident analysis for classifying errors (O'Hare, Wiggins, Batt & Morrison, 1994; Rasmussen, 1982).

Using the decision ladder (Figure 2), the problem solving processes of aircrew at each critical point are analysed in terms of observation of information, situation analysis, goal evaluation, and planning and execution. The aim is to understand why the aircrew may have responded as they did. To do this, we have found it useful to prompt ourselves with the following questions about the aircrew's behaviour:

- Is it possible that the crew did not *observe* critical information?
- Is it possible that the crew had difficulty *diagnosing* the situation?
- Is it possible that the crew gave precedence to alternative *goals*?
- Is it possible that the crew had difficulty *defining the tasks and resources* required for dealing with the situation?
- Is it possible that the crew had difficulty *selecting or formulating procedures* for dealing with the situation?

- Is it possible that the crew did not *execute* the procedure as intended?

Figure 2: A decision ladder for a hypothetical accident.

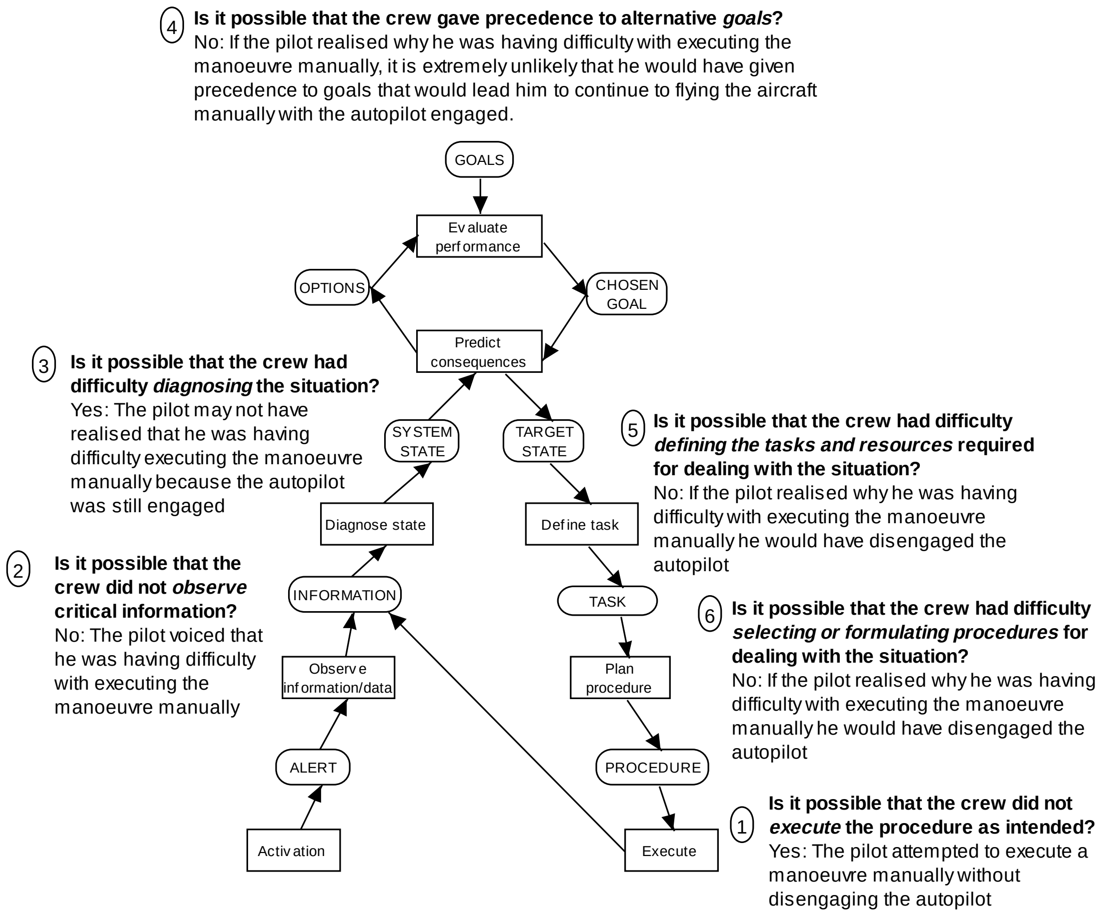


Figure 2 shows a decision ladder for the first critical point in the hypothetical accident we presented earlier. The annotations in the figure are presented as answers to the questions posed above, with the numbers representing the order in which the decision ladder should be followed. From this representation, we begin to understand that the pilot had difficulty in detecting the error he had made (i.e. that is that he had tried to execute a manoeuvre manually without disengaging the autopilot). More specifically, we can see that the difficulty he had in detecting the error was not because he did not *observe* the critical information (he was aware that he was finding it hard to execute the manoeuvre manually) but rather because he was unable to *diagnose* why he was finding it hard to execute the manoeuvre manually. If the pilot had made the correct diagnosis, he would have probably detected the error (that is, realised that he had forgotten to disengage the autopilot), and consequently corrected the error (that is, disengaged the autopilot).

Generating Training Requirements: The final step involves generating requirements for training on the basis of the preceding analysis. In particular, the elements of the preceding analysis that are relevant for informing training requirements include the boundaries that were crossed by aircrew, and the problem solving difficulties that they experienced. The requirements can then be structured in terms of the boundaries that aircrew should be given the opportunity to cross during training, and the problem solving processes that they should practise to enable error detection and error recovery.

From the analysis of the hypothetical accident presented above, the training requirement would be to give aircrew the opportunity to fly the aircraft manually with the autopilot engaged so that they can experience how the aircraft would respond. In addition, aircrew should practise disengaging the autopilot while they are controlling the aircraft manually. Then, if the aircrew forget to disengage the autopilot before executing a manoeuvre manually on a real mission, they are more likely to diagnose the error and recover from it successfully. This kind of training intervention is consistent with theories of naturalistic decision making which recognise that under time-critical, high-workload conditions, experts can make quick and effective decisions by matching situations to pre-existing templates of diagnoses and solutions that have worked in the past (Klein, 1993).

Application of Technique

Aircraft Accidents: So far we have applied the technique that we have developed to three F-111 accidents in the Royal Australian Air Force (the F-111 is a two-person strike aircraft). The accident data that was necessary for conducting the analysis was readily available in reports of the Accident Investigation Teams and Boards of Inquiry. Recordings of cockpit activity in the accident aircraft were particularly valuable for constructing the decision-ladder models of aircrew problem solving.

Examining the accident data was the most time consuming component of the analysis. It took between three to five days to examine the data for each accident (depending on the amount of information that was available about each accident). Once the accident data had been examined, it took approximately a day to complete the first step of the technique, two days to complete the second step, and a day to complete the third step.

Our analyses of the three aircraft accidents resulted in 6 training requirements. To assess the usefulness of the technique we interviewed 7 F-111 aircrew and 7 F-111 training instructors. Some of the questions we asked them included: (1) whether they already conducted the training suggested; (2) whether the training suggested was useful; and (3) whether they had been in an unsafe situation that was similar to the one that had resulted in the training requirement. We are still in the process of analysing the interview transcripts in detail, but from a preliminary examination of the transcripts it appears that they do not conduct the training suggested, that they thought the training suggestions were useful, and that they had previously been in similar unsafe situations.

Aircraft Incidents: We are also in the process of applying the technique we have developed to F-111 incidents. We have found that the data necessary for analysing the incidents is generally not available in the incident reports that have been filed by aircrew. To resolve this problem, we will interview aircrew about the incidents they have reported using a technique called Critical Decision Method (Klein, Calderwood & MacGregor, 1989). This technique allows interviewers to gradually shift aircrew from an operational description of the incident, which is the language that aircrew are most accustomed to speaking in, to a description of the problem solving processes that were behind the incident.

Our initial attempt at using the Critical Decision Method involved very little adaptation of the technique as it is described in Klein et al. (1989) and Hoffman, Crandall & Shadbolt (1998). Briefly, aircrew were asked to provide a general description of the incident followed by a more detailed account of the sequence of events in the incident. The interviewer and the aircrew then established a timeline for the incident and identified the critical points in the incident. Following that, the interviewer used a number of probes to elicit more detailed information from aircrew about the problem solving processes at each of the critical points in the incident. The probes were much the same as those described in Klein et al. (1989) and Hoffman et al. (1998).

On reviewing the interview transcripts we discovered that we had not fully captured the information we needed to develop decision-ladder models of the incidents, and consequently to identify training requirements. In addition, a significant difference in analysing incidents, as opposed to accidents, is that the aircrew who were involved can provide valuable information about how they actually detected and recovered from the error. Thus, we needed to interview aircrew not only about the problem-solving difficulties that led them to cross the boundaries of safe operation but also about the problem solving processes that enabled error detection and error recovery.

In other words, for each incident, the interviewer should focus on at least three critical points. These critical points involve the: (1) error; (2) error detection; and (3) error recovery. The interviewer should use general probes to prompt free recall of the aircrew's experiences at each critical point, followed by specific probes where necessary to elicit the information required for constructing decision-ladder models. Table 1 illustrates some of the general and specific probes that may be useful at each critical point. The specific probes are organised according to the parts of the decision ladder that the information elicited is relevant to. In the last two columns, the cells that are filled indicate that error detection processes are represented on the left side of the decision ladder whereas error recovery processes are represented at the top part and on the right side of the decision ladder.

Table 1: General and specific probes for interviewing aircrew about errors, error detection, and error recovery.

Parts of the decision ladder (except for first cell)	**Error:**	**Error detection:**	**Error recovery:**
General probes	What when wrong?	How did you detect the error?	How did you react or recover from the error?
Observation of information	What information did you have about the situation?	What cues alerted you that something was wrong?	
Diagnosis	What was your assessment of the situation at this point?	What was your assessment of what had gone wrong?	
Goal evaluation	• What were your specific goals at this time? • What other options did you consider? • Why did you select this option/reject other options?		• What were your specific goals at this time? • What other options did you consider? • Why did you select this option/reject other options?
Definition of tasks and resources	What was your plan for achieving your goals?		What was your plan for recovering from the situation?
Formulation and selection of procedures	Were there procedures for dealing with the situation? What were the steps of your plan?		Were there procedures for recovering from this situation? What were the steps of your recovery plan?
Execution	What tasks or actions did you carry out?		What tasks or actions did you carry out?

After the relevant information at each critical point has been obtained, the following probes may be useful for uncovering information for developing error management strategies: (1) In hindsight, could you have detected the error earlier and, if so, how?; (2) In hindsight, could you have recovered more effectively/efficiently from the error and, if so, how?; (3) In hindsight, is there anything you could have done to prevent the error from occurring and, if so, what?; (4) In hindsight, why do you think the error occurred?

We will trial this interview protocol in the near future. Further development and refinement of the probes may be necessary. Later, it may be worth considering how to design reporting templates for an incident database so that the information needed to develop training requirements for error detection and error recovery is captured.

Implementation of Training: Many of our training requirements must be implemented in training simulators rather than in real aircraft. This is not surprising because our training approach requires that aircrew ‘cross the boundaries of safe operation’ and it would be dangerous to do this in real aircraft. The challenge that we face is that the training simulators that are available may not have the necessary capability for supporting some of our training requirements. One solution is to document these training requirements for future simulation acquisitions.

Another option is to explore alternative techniques for training. Reason (2001) reported that the ability of surgical teams to deal with adverse incidents depended in part on the extent to which they had mentally rehearsed the detection and recovery of their errors. Thus, one possibility we are exploring is the use of mental rehearsal techniques, perhaps with the aid of PC-based visualisation tools, to give aircrew 'experience' in crossing the boundaries and in detecting and recovering from crossing the boundaries.

In addition, in a study examining the role of prior cases in pilot decision making, O'Hare and Wiggins (2002) found that written materials were an important source of remembered cases. They also suggested that PC-based simulations may also be effective candidates for case-based training systems. The approach we have presented in this paper may be useful for the preparation of cases for training, in particular, for providing information about the boundaries that were crossed by aircrew, the problem solving difficulties that they experienced, and the problem solving processes that would enable error detection and error recovery.

Conclusion

In this paper, we have described a new approach for training aircrew to manage human error. In addition, we have presented a technique for analysing aircraft accidents and incidents to identify specific requirements for training aircrew in error detection and error recovery. Initial applications of this approach have been encouraging and provide motivation for continuing further work in this area.

Acknowledgements

We thank the Strike Reconnaissance Group of the Royal Australian Air Force for sponsoring this work and the Directorate of Flying Safety of the Royal Australian Air Force for their support. We also thank Lee Horsington, Dominic Drumm, and Anna Moylan from the Defence Science and Technology Organisation for their assistance on this project. In addition, we thank Jim McLennan from Swinburne University of Technology for conducting some of the initial interviews with aircrew; and Gary Klein and Laura Militello of Klein Associates for their advice on the Critical Decision Method.

References

Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems. *Safety Science, 37*, 109-126.

Helmreich, R.L., Merritt, A.C., & Wilhelm, J.A. (1999). The evolution of crew resource management in commercial aviation. *International Journal of Aviation Psychology, 9*, 19-32.

Hoffman, R.R., Crandall, B., & Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors, 40(2),* 254-276.

Hollnagel, E. (1993). The phenotype of erroneous actions. *International Journal of Man-Machine Studies, 39*, 1-32.

Klein, G. A. (1993). *Naturalistic decision making: Implications for design*. Report CSERIAC SOAR 93-1. Ohio: Crew Systems Ergonomics Information Analysis Center.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method of eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics, 19*, 462-472.

Lintern, G., & Naikar, N. (2001). *Analysis of crew coordination in the F-111 mission*. DSTO Client Report (DSTO-CR-0184). Aeronautical and Maritime Research Laboratory: Melbourne, Australia.

Maurino, D. (2001). At the end of the parade. *Flight Safety Magazine, Jan/Feb 2001*, pp.36-39.

Noyes, J.M. (1998). Managing errors. In *Proceedings of the UKACC* International *Conference on Control*, pp.578-583. London: Institution of Electrical Engineers**.**

O'Hare, D., & Wiggins, M. (2002). *Remembrance of cases past: Who remembers what, when confronting critical flight events*. Unpublished manuscript.

O'Hare, D., Wiggins, M., Batt, R., Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics, 37*(1), 1855-1869.

Rasmussen, J. (1982). Human errors: A taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents, 4,* 311-333.

Rasmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). *Cognitive Systems Engineering*. New York: John Wiley & Sons.
Reason, J. (2000). Human error: Models and management. *British Medical Journal, 320,* 768-770.
Reason, J. (2001). The benign face of the human factor. *Flight Safety Magazine, Jan/Feb 2001,* pp. 28-31.
Sarter, N.B., & Alexander, H.M. (2000). Error types and related error detection mechanisms in the aviation domain: An analysis of aviation safety reporting system incident reports. *International Journal of Aviation Psychology, 10,* 189-206.
Shappell, S.A., & Wiegmann, D.A. (2000). *The Human Factors Analysis and Classification System – HFACS*. Report DOT/FAA/AM-00/7. Springfield: National Technical Information Service.
Vicente, K.J. (1999). *Cognitive Work Analysis*. New Jersey: Lawrence Erlbaum Associates.
Woods, D.D., Johannesen, L.J., Cook, R.I., & Sarter, N.B. (1994). *Behind human error: Cognitive systems, computers and hindsight*. Report CSERIAC SOAR 94-01. Ohio: Crew Systems Ergonomics Information Analysis Center.

Activity Tracking for Pilot Error Detection from Flight Data

Todd J. Callantine,

San Jose State University/NASA Ames Research Center, MS 262-4, Moffett Field, CA 94035, USA.
tcallantine@mail.arc.nasa.gov

Abstract: This paper presents an application of activity tracking for pilot error detection from flight data. It describes the Crew Activity Tracking System (CATS), in-flight data collected from the NASA Langley Boeing 757 Airborne Research Integrated Experiment System aircraft, and a model of B757 flight crew activities. It then presents an example of CATS detecting actual in-flight crew errors.

Keywords: human error detection, activity tracking, glass cockpit aircraft

Introduction

This paper describes an application of the Crew Activity Tracking System (CATS) that could contribute to future efforts to reduce flight crew errors. It demonstrates how CATS tracks crew activities to detect errors, given flight data and air traffic control (ATC) clearances received via datalink. CATS implements a so-called 'intent inference' technology, called activity tracking, in which it uses a computational 'engineering' model of the operator's task, together with a representation of the current operational context, to predict nominally preferred operator activities and interpret actual operator actions.

CATS was originally implemented to track the activities of Boeing 757 (B757) 'glass cockpit' pilots, with a focus on automation mode errors (Callantine and Mitchell, 1994). The CATS activity tracking methodology was validated as a source of real-time knowledge about B757 automation usage to support a pilot training/aiding system (Callantine, Mitchell, and Palmer, 1999). CATS has since proven useful as an analysis tool for assessing how operators use procedures developed to support new operational concepts (Callantine, 2000). It also serves as a framework for developing agents to represent human operators in incident analyses and distributed simulations of new operational concepts (Callantine, 2001a).

The research described here draws in large part from these earlier efforts. In particular, the CATS model of B757 flight crew activities has been expanded and refined. The representation of operational context used to reference the model to predict nominally preferred activities has similarly undergone progressive refinement. And, while the idea of using CATS to detect flight crew errors from flight data is not new, this paper presents an example of CATS detecting a genuine, in-flight crew error from actual aircraft flight data.

Using CATS to detect errors from flight data has several potential benefits (Callantine, 2001b). First, CATS provides information about procedural errors that do not necessarily result in deviations, and therefore would not otherwise be reported (cf. Johnson, 2000). Second, CATS enables airline safety managers to 'automatically' incorporate information about a detected error into a CATS-based training curriculum. Other pilots could 'relive' a high-fidelity version of the context in which another crew erred. Increasing the efficiency and fidelity of information transfer about errors to the pilot workforce in this way would likely yield safety benefits. A safety-enhancement program that uses CATS to detect errors would improve training by requiring safety and training managers to explicate policies about how an aircraft should preferably be flown.

The paper is organized as follows. It first describes the CATS activity tracking methodology, and information flow in CATS. The paper then describes a CATS implementation for detecting pilot errors. It first describes flight data obtained for this demonstration from the NASA Langley B757 Airborne Research Integrated Experiment System (ARIES) aircraft. It next describes two key representations. The first is a portion of a CATS model of B757 flight operations. The second is a representation of the constraints conveyed by ATC clearances that plays a key role in representing the current operational context (Callantine, 2002b). An example from the available flight data then illustrates CATS detecting pilot errors.

The paper concludes with a discussion of future research challenges. A lengthier report on this research appears in Callantine (2002a).

Activity Tracking

Activity tracking is not merely the detection of operational 'deviations' (e.g., 'altitude below glidepath'). The activity tracking methodology involves first predicting the set of expected nominal operator activities for the current operational context, then comparing actual operator actions to these predictions to ensure operators performed correct activities. In some situations, various methods or techniques may be acceptable; therefore the methodology also includes a mechanism for determining that, although operator actions do not match predictions exactly, the actions are nonetheless correct. In this sense, CATS is designed to 'track' flight crew activities in real time and 'understand' that they are error-free. As the example below illustrates, 'errors' CATS detects include those that operators themselves detect and rapidly correct; such errors may nonetheless be useful to examine.

CATS identifies two types of errors: errors of omission, and errors of commission. It further identifies errors of commission that result when the 'right action' is performed with the 'wrong value.' CATS does not base these determinations on a 'formulaic' representation of how such errors would appear in a trace of operator activities, nor attempt to further classify errors (e.g., 'reversals'). This is because the CATS model does not represent the 'steps' of procedures explicitly as 'step A follows step B;' instead it represents procedures implicitly by explicitly specifying the conditions under which operators should preferably perform each action. CATS predicts concurrent actions whenever the current context satisfies conditions for performing two or more activities. CATS interprets concurrent actions whenever the granularity of action data identifies them as such.

Like analysis techniques that rely on a 'reflection' of the task specification in a formal model of a system (e.g., Degani and Heymann, 2000), CATS relies on a correctly functioning system to reflect the results of actions (or inaction) in its state. CATS identifies errors by using information in the CATS model that enables it to assess actions (or the lack thereof, in the case of omissions) in light of the current operational context and the future context formed as a result of operator action (or inaction). Thus, one might view the CATS error detection scheme as 'closing the loop' between a representation of correct task performance and the controlled system, and evaluating feedback from the controlled system to ensure it 'jibes' with correct operator activities. Given that the system is operating normally and providing 'good data,' this is a powerful concept.

Crew Activity Tracking System (CATS): Figure 1 generically depicts information flow in CATS, between a controlled system and CATS, and between CATS and applications based on it. CATS uses representations of the current state of the controlled system and constraints imposed by the environment (including performance limits on the controlled system) to derive the current operational context. CATS then uses this representation to generate predictions from its model of operator activities. CATS compares detected operator actions to its predicted activities, and it assesses actions that it cannot immediately interpret as matching a prediction by periodically referencing the activity model until it receives enough new context

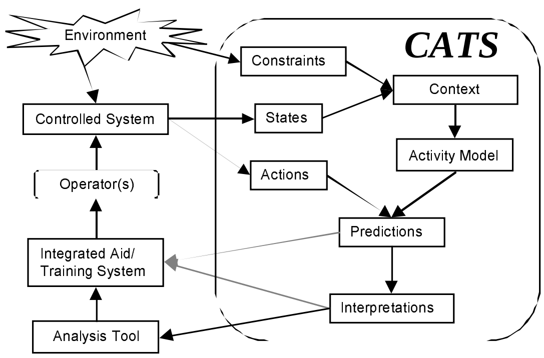


Figure 1 – Information flow within and between CATS and a generic human-machine system, with applications to error analysis, aiding, and training.

information to disambiguate possible interpretations.

CATS Implementation for Flight Data Error Detection

The following subsections specifically describe the implementation of CATS for detecting pilot errors from flight data. The first is devoted to the flight data itself. The second illustrates a portion of the CATS model, and the third describes how CATS generates the current operational context using a representation of ATC clearance constraints. The CATS model fragment includes portions relevant to an example of CATS detecting pilot errors presented in the fourth subsection. The following subsections all assume some knowledge of commercial aviation and a B757-style autoflight system. A detailed description of the Boeing 757 autoflight system mode usage is provided in Callantine, Mitchell, and Palmer (1999); see Sarter and Woods (1995), and Wiener (1989) for discussions of mode errors and automation issues.

B757 ARIES Flight Data: The NASA Langley B757 ARIES aircraft, with its onboard Data Acquisition System (DAS), provided the flight data for this research (Figure 2). The DAS collects data at rates in excess of 5 Hz, using onboard computers that perform sensor data fusion and integrity checking. In future applications such functionality may be required within CATS, so that data can be acquired directly from aircraft data busses. Table 1 shows the collection of values that comprise the data set. The data include information from important cockpit systems. The rightmost column of Table 1 shows data CATS derives from the sampled values using filters. Included are crew action events CATS derives from the values of control states. Target value settings on the MCP are derived with 'begin' and 'end' values, as in formal action specification schemes (cf. Fields, Harrison, and Wright, 1996). The present error-detection application focuses on interactions with the autoflight system MCP, so it only uses some of the available data. Also, for the present application, cockpit observations provide required clearance information.

CATS Model of B757 Navigation Activities: Figure 3 depicts a fragment of the CATS model used to detect errors from B757 ARIES data. The model decomposes the highest level activity, 'fly glass cockpit aircraft,' into sub-activities as necessary down to the level of pilot actions. Figure 3 illustrates eight actions. All actions derivable from the data are included in the full model. Each activity in the model is represented with conditions that express the context under which the activity is nominally preferred, given policies and procedures governing operation of the controlled system. The parenthesized numbers in Figure 3 refer to Table 2, which lists the 'and-or trees' that comprise these rules. For comparison to other work that considers human errors involved with CDU manipulations (e.g., Fields, Harrison, and Wright, 1997), the model fragment in Figure 3 shows just one of numerous FMS configuration tasks. However, because the B757 ARIES flight data do not include CDU data, modeling these tasks is not relevant to the present application.

Figure 2 – Data Acquisition System (DAS) onboard the NASA B757 ARIES aircraft (inset).

Table 1 – Available B757 ARIES data, including derived states and action events (rightmost column). The B757 ARIES DAS collects some variables from multiple sources.

Time variables	**NAV/COMM data**	**AFDS modes**	**FMC-A/T internal data**	***Derived states***
time	dme_range	fl_ch_engd	fmc_at_mach_mode_reqd	vert_speed
time1	left_dme_freq	hdg_hold_engd	fmc_at_airspeed_mode_reqd	alt_cap_engaged
time2	right_dme_freq	hdg_sel_engd	fmc_active_climb	spd_win_auto_chng
time3	left_dme_dist	land_2_green	fmc_climb_mode_reqd	ap_cmd_engd
Environmental information	right_dme_dist	land_3_green	fmc_active_cruise	***Derived MCP actions***
total_air_temp	left_vhf_freq	alt_hold_engd	fmc_con_mode_reqd	set MCP hdg
true_wind_dir	right_vhf_freq	vnav_armed_engd	fmc_crz_mode_reqd	set MCP alt
wind_speed	**FMC data**	lnav_armed_engd	fmc_active_descent	set MCP spd
AC position/attitude	fmc_target_airspeed	speed_mode_engd	fmc_display_annunc_on	set MCP mach
baro_alt	fmc_selected_altitude	thrust_mode_engd	fmc_eng_ident_1	set MCP vs
baro_corr	fmc_selected_airspeed	loc_engd	fmc_eng_ident_2	hdg sel press
flight_path_angle	fmc_selected_mach	vert_spd_engd	fmc_eng_ident_3	hdg hold press
ground_speed	fmc_crz_altitude	apprch_armed_engd	fmc_eng_ident_4	lnav press
computed_airspeed	fmc_eta	loc_armed_engd	fmc_eng_ident_5	vnav press
calibrated_airspeed	fmc_desired_track	back_course_armed_engd	fmc_eng_ident_6	spd press
mach	fmc_wpt_bearing	glideslope_engd	fmc_eng_ident_7	apprch press
magnetic_heading	fmc_cross_track_dist	**MCP Speed display status**	fmc_eng_ident_8	loc press
magnetic_track_angle	fmc_vert_dev	mcp_speed_display_blank	fmc_eng_ident_9	alt hold press
pitch_angle	fmc_range_to_alt	**Autothrottle**	fmc_eng_ident_10	vs mode press
radio_altitude	fmc_wide_vert_dev	at_armed	fmc_ga_mode_reqd	fl ch press
roll_angle	**AFDS states**	**MCP switches**	fmc_idle_thr_reqd	thrust mode press
true_track_angle	ap_cmd_ctr_engd	hdg_sel_reqd	fmc_msg_annunciated	mach toggled
iru_potential_vert_speed	ap_cmd_cen_gc_huh	hdg_hold_reqd	throttle_retard_reqd	c ap cmd switch press
hybrid_lat	ap_cmd_cen_gr_huh	lnav_reqd	pitch_speed_control_engd	l ap cmd switch press
hybrid_lon	left_ap_cmd_engd	vnav_reqd	vnav_operational	r ap cmd switch press
AC configuration/controls	ap_cmd_left_engd	spd_reqd	lnav_operational	arm autothrottles
left_engine_epr	right_ap_cmd_engd	apprch_reqd	tmc_valid	***Other derived actions***
right_engine_epr	ap_cmd_right_engd	loc_reqd	**VNAV submodes**	tune left VHF
flap_pos	ap_cmd_center_engd	alt_hold_reqd	fmc_vnav_speed_operational	tune right VHF
speed_brake_handle	ap_cws_center_engd	vs_mode_reqd	fmc_vnav_path_operational	set flaps
left_throttle_pos	ap_cws_left_engd	fl_ch_reqd	fmc_vnav_alt_operational	set spoilers
right_throttle_pos	ap_cws_right_engd	thrust_mod_reqd	**Thrust ratings**	
gross_weight	ap_in_control	**IAS/Mach toggle**	fmc_rating_1_reqd	
MCP target values	fd_c_on	mach_toggled	fmc_rating_2_reqd	
sel_mcp_altitude	fd_fo_on	**Crew Alert levels**	fmc_offset_annunciated	
sel_mcp_heading	fd_on_c	crew_alert_level_a	fmc_throttle_dormant_reqd	
sel_mcp_speed	fd_on_fo	crew_alert_level_b	fmc_thr_mode_reqd	
sel_mcp_vert_speed	**AFDS switches**	crew_alert_level_c	fmc_to_mode_reqd	
mcp_flare_retard_rate	ap_cmd_center_reqd	**Status data**	req_1_valid_resv	
sel_mcp_mach	ap_cmd_right_reqd	eec_valid	req_2_valid_resv	
MCP bank angle settings	ap_cws_center_reqd	engine_not_out		
bank_angle_lim_flaps_25	ap_cws_left_reqd			
bank_angle_lim_flaps_15	ap_cws_right_reqd			
bank_angle_lim_auto	ap_cmd_left_reqd			

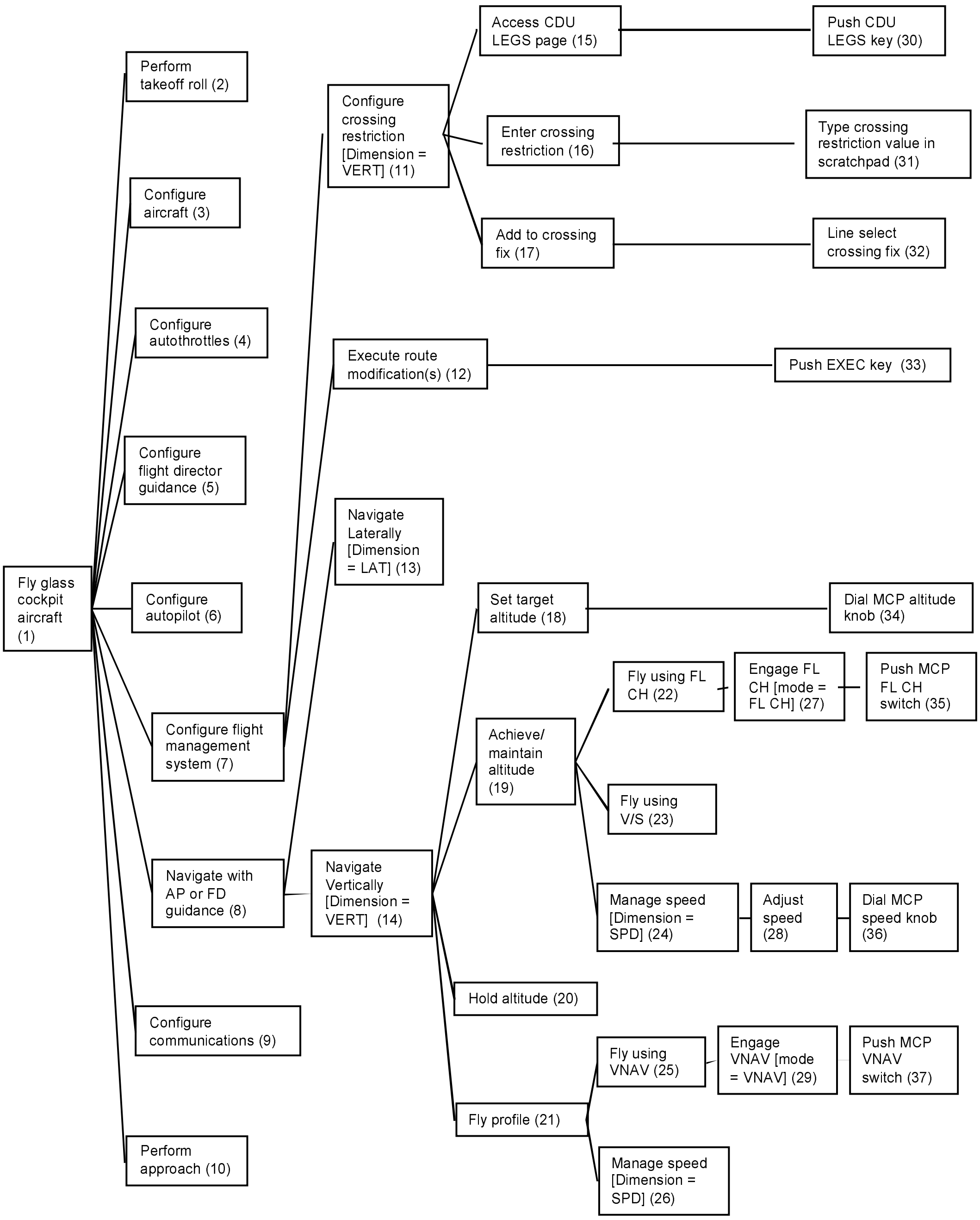


Figure 3 – Fragment of CATS model for B757 operations.

Table 2 – AND-OR trees of conditions under which the CATS model in Figure 3 represents activities as 'nominally preferred.' CATS predicts an activity when its conditions, plus all the conditions of its parent activities are satisfied by the current operational context.

(1) start-of-run
(2) (not above-runway-elevation)
(3) (and (not above-clean-speed) (not flight-surfaces-within-limits) (not gear-within-limits))
(4) (not autothrottle-armed)
(5) (not flight-director-on)
(6) [(and (not autopilot-cmd-mode-engaged) above-1000-feet-AGL)]
(7) (or (not programmed-route-within-limits) route-uplink-received)
(8) (and above-1000-feet-AGL (or autopilot-cmd-mode-engaged flight-director-on))
(9) (not comm-frequency-within-limits)
(10) (or approaching-glideslope-intercept-point approach-localizer-intercept-point)
(11) (not crossing-restriction-within-limits)
(12) route-modifications-within-limits
(13) (or autopilot-cmd-mode-engaged flight-director-on)
(14) (or autopilot-cmd-mode-engaged flight-director-on)
(15) (not cdu-page-LEGS)
(16) (and cdu-page-LEGS (not crossing-restriction-built))
(17) (and cdu-page-LEGS crossing-restriction-built)
(18) (not mcp-altitude-within-limits)
(19) (or (and (not current-altitude-within-limits) (not profile-within-limits-for-now)) expedite-needed)
(20) (and current-altitude-within-limits (not profile-within-limits-for-now))
(21) profile-within-limits-for-now
(22) (or (not altitude-close-to-target) expedite-needed)
(23) altitude-close-to-target
(24) (or fl-ch-engaged vs-engaged)
(25) profile-within-limits-for-now
(26) vnav-engaged
(27) (not fl-ch-engaged)
(28) (not target-speed-within-limits)
(29) (and (not vnav-engaged) (not capturing-required-altitude))
(30) (not cdu-page-LEGS)
(31) (not crossing-restriction-built)
(32) crossing-restriction-built
(33) route-modifications-within-limits
(34) (not mcp-altitude-within-limits)
(35) mcp-altitude-within-limits
(36) (not target-speed-within-limits)
(37) mcp-altitude-within-limits

Representation of ATC Clearance Constraints for Context Generation: Environmental constraints play a key role in defining the goals that shape worker behavior in complex sociotechnical systems (Vicente, 1999). CATS also relies on a representation of environmental constraints to construct a representation of the current operational context (see Figure 1). These factors motivated recent research on an object-oriented representation of the constraints ATC clearances impose on flight operations (Callantine, 2002b). Figure 4 shows the representation, which represents three key dimensions of constraints: vertical, lateral, and speed. CATS employs a rule base that enables it modify this constraint representation to reflect the constraints imposed (or removed) by each new ATC clearance.

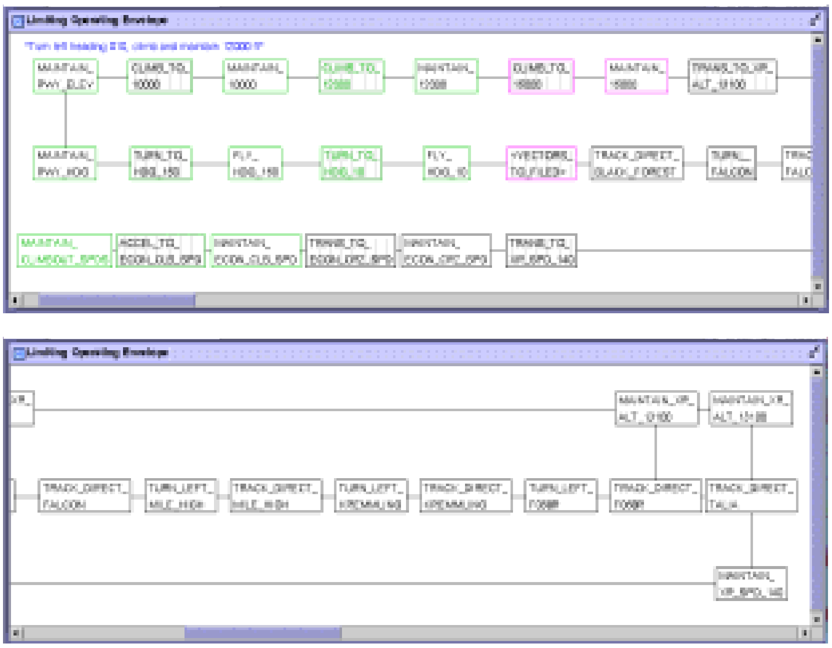


Figure 4 – Snapshot of a CATS representation of environmental constraints constructed from the filed flight plan and modified by ATC clearances.

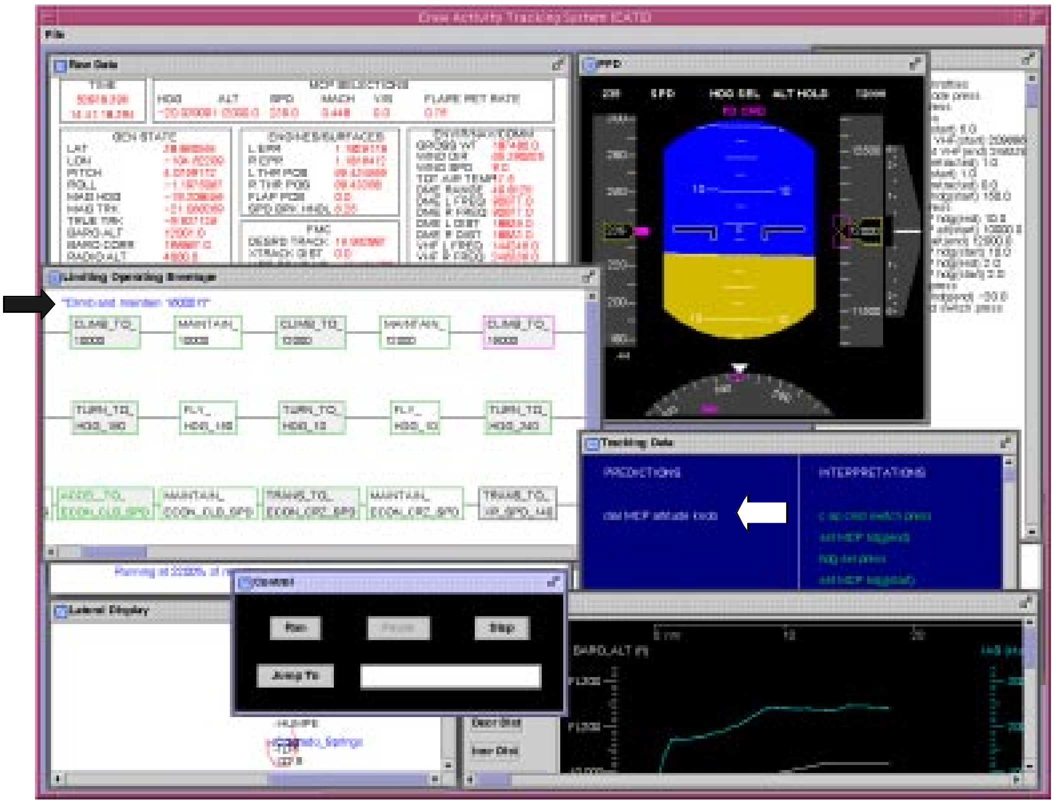


Figure 5 – Scenario Frame 1: In response to a clearance to climb, CATS predicts the crew should set the new target altitude on the MCP by dialing the MCP altitude knob.

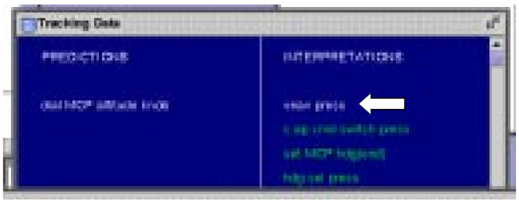


Figure 6 – Scenario Frame 2: CATS detects that a crew member pressed the VNAV switch instead of setting the MCP altitude.

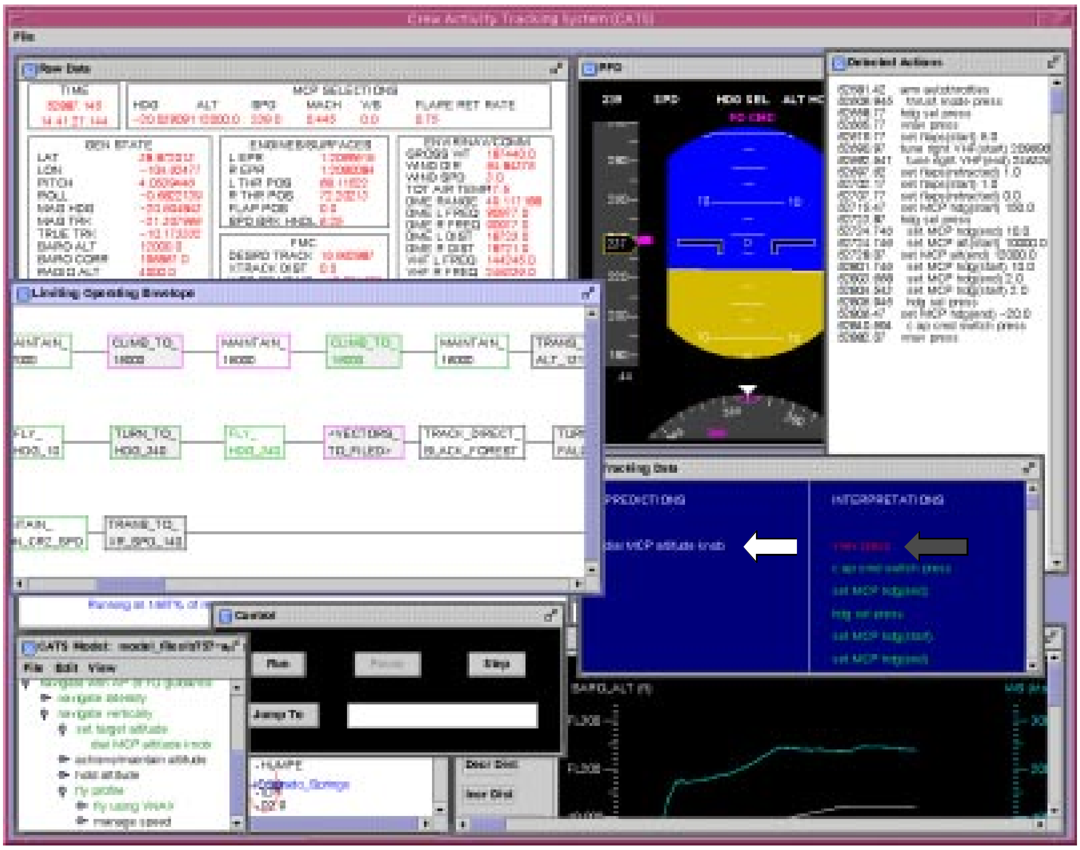


Figure 7 – Scenario Frame 3: CATS cannot reconcile the VNAV switch press with the current context, and therefore flags it as an error; CATS is still expecting the crew to dial the MCP altitude knob.

Figure 8 – Scenario Frame 4: CATS detects a pilot starting to dial the MCP altitude, and interprets it as matching its prediction, but with the wrong value (not an error, because the action is only the start of the altitude setting).

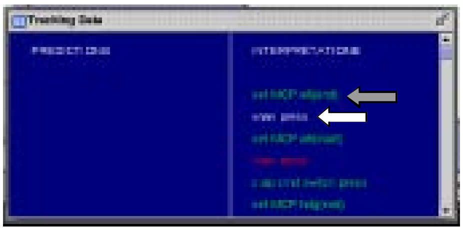


Figure 9 – Scenario Frame 5: A second VNAV switch press, before the altitude setting is finished.

Figure 10 - Scenario Frame 6: CATS detects that the crew has now opted to engage FL CH mode by pressing the FL CH switch. But because the altitude is now properly set, CATS now predicts the crew should push the VNAV switch to engage VNAV (the preferred mode according to the CATS model).

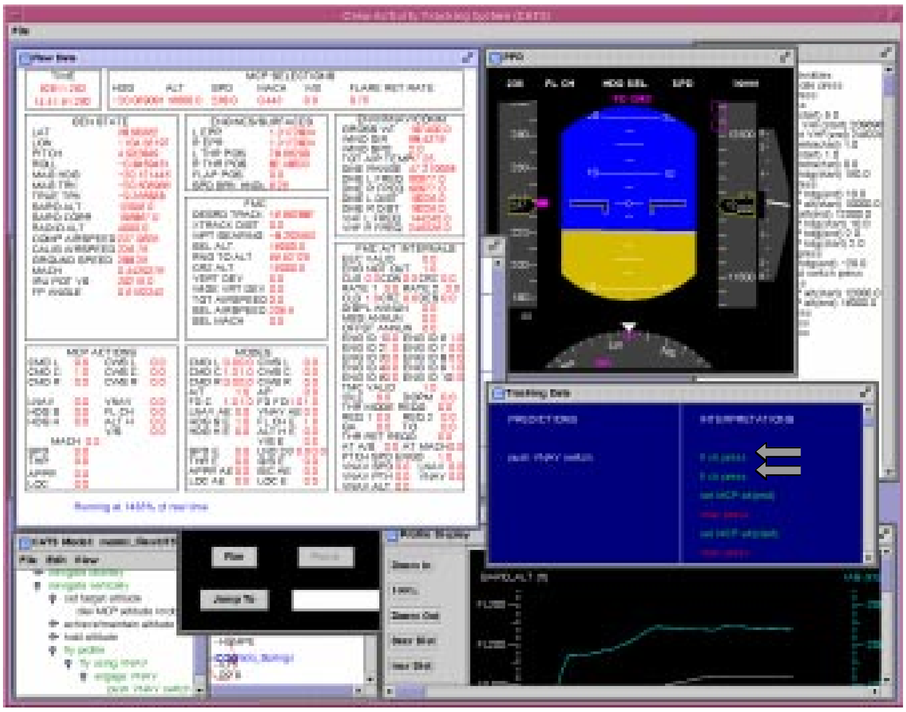


Figure 11 - Scenario Frame 7: CATS detects a second 'insurance' FL CH switch press, and interprets it as acceptable as it did the first FL CH switch press.

Figure 12 - Scenario Frame 8: The crew opts to engage VNAV; CATS detects the predicted VNAV switch press and interprets it as correct (elapsed time from Scenario Frame 1 is ~42 secs).

Error Detection Example: The paper now presents an example of CATS detecting errors from B757 ARIES flight data collected during actual flight test activities. (A series of snapshots, including some of the entire CATS interface, illustrate the example.) Although the data are real, in the flight test environment, strict procedures about how the pilots should preferably fly the airplane are unreasonable. Nonetheless, by imposing the model depicted in part in Figure 3, CATS was able to detect errors, and the errors were not contrived. While the errors CATS detects are insignificant, because they in no way compromised safety, the exercise nonetheless demonstrates the viability of CATS for error detection. On the SUN Blade1000™ test platform, the CATS Java™ code processes the flight data at approximately between twelve and twenty-two times real time.

Figure 5 shows the CATS interface at the start of the scenario (Scenario Frame 1). The crew has just received a clearance to "climb and maintain 16,000 feet." CATS modifies its representation of ATC clearance constraints accordingly, and using the updated context, predicts that the crew should set the new target altitude on the MCP by dialing the MCP altitude knob.

In Scenario Frame 2 (Figure 6), a pilot instead pushes the VNAV switch. Because CATS has not predicted this action, it cannot interpret the action initially. CATS instead continues processing data. In Scenario Frame 3 (Figure 7), CATS has received enough new data to interpret the VNAV switch press action. Had the action been correct, the autoflight system would have reflected this by engaging the VNAV mode and commencing the climb. However, VNAV will not engage until a new target altitude is set. To assess the VNAV switch press with regard to the current context, in which airplane is still in ALT HOLD mode at 12,000 feet, CATS searches its model to determine if any parent activities of the VNAV switch press contain information linking the action to a specific context. CATS finds that the ‘engage VNAV’ activity should reflect VNAV mode engagement in the current context (see Figure 3). Because this is not the case, CATS flags the VNAV switch press as an error. Meanwhile, CATS still expects the crew to dial the MCP altitude knob.

.

CATS detects a second FL CH switch press in Scenario Frame 7 (Figure 11). Perhaps a pilot performed this action as ‘insurance’ to engage a mode to begin the climb. Because FL CH mode engages, and this is reflected in CATS’ representation of the current context, CATS interprets both FL CH switch presses as correct acceptable alternative actions. By this time, CATS has also flagged the second VNAV switch press as an error. In the final frame of the scenario (Scenario Frame 8, Figure 12), the aircraft has begun climbing in FL CH mode. At this point the crew opts to engage VNAV mode. At last, CATS detects the predicted VNAV switch press and interprets it as correct.

Conclusions and Future Research

The above example demonstrates that CATS can detect errors from flight data. Although the errors CATS detects are inconsequential, this research indicates CATS can provide contextual information useful for disambiguating the causes of deviations or unusual control actions that arise in incident or accidents. Discoveries made using CATS can be incorporated into training curricula by connecting a CATS-based training system to a simulator and allowing pilots to ‘fly’ under conditions that correspond the actual context of an error-related event. Such capabilities are also useful outside the airline arena as they support both fine-grained cognitive engineering analyses and human performance modeling research.

Using CATS with flight data collected at ‘continuous’ rates results in better performance. Event-based data, such as those available from the NASA ACFS, require more complicated interpolation methods to avoid temporal ‘gaps’ in the CATS representation of context that can adversely affect CATS performance. Important directions for further research involve improving the coverage of flight data to include the FMS and CDUs, as well as work on methods to automatically acquire ATC clearance information. This research indicates that, if CATS has access to data with full, high-fidelity coverage of the controlled system displays and controls, it can expose the contextual nuances that surround errors in considerable detail.

Acknowledgements
This work was funded under the System Wide Accident Prevention element of the FAA/NASA Aviation Safety Program. Thanks to the NASA Langley B757 Flight Test team for their assistance with data collection.

References

Callantine, T. (2001a). Agents for analysis and design of complex systems. *Proceedings of the 2001 International Conference on Systems, Man, and Cybernetics,* October, 567-573.

Callantine, T. (2001b). The crew activity tracking system: Leveraging flight data for aiding, training, and analysis. *Proceedings of the 20th Digital Avionics Systems Conference*, 5.C.3-1—5.C.3-12 (CD-ROM).

Callantine, T. (2002a). Activity tracking for pilot error detection from flight data. NASA Contractor Report 2002-211406, Moffett Field, CA: NASA Ames Research Center.

Callantine, T. (2002b). A representation of air traffic control clearance constraints for intelligent agents. *Proceedings of the 2002 IEEE International Conference on Systems, Man, and Cybernetics*, Hammamet, Tunisia, October.

Callantine, T., and Mitchell, C. (1994). A methodology and architecture for understanding how operators select and use modes of automation in complex systems. *Proceedings of the 1994 IEEE Conference on Systems, Man, and Cybernetics*, 1751-1756.

Callantine, T., Mitchell, C., and Palmer, E. (1999). GT-CATS: Tracking operator activities in complex systems. NASA Technical Memorandum 208788, Moffett Field, CA: NASA Ames Research Center.

Fields, R., Harrison, M., and Wright, P. (1997). THEA: Human error analysis for requirements definition. Technical Report 2941997, York, UK: University of York Computer Science Department.

Fields, R., Wright, P., and Harrison, M. (1996). Temporal aspects of usability: Time, tasks and errors. SIGCHI Bulletin 28(2).

Johnson, C. (2000). Novel computational techniques for incident reporting. In D. Aha & R. Weber (Eds.), *Intelligent Lessons Learned Systems: Papers from the 2000 Workshop* (Technical Report WS-00-03), Menlo Park, CA: AAAI Press, 20-24.

Sarter, N., and Woods, D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors,* 31(1), 5-19.

Vicente, K. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work.* Mahwah, NJ: Lawrence Erlbaum Associates.

.

Development and Preliminary Validation of a Cognitive Model of Commercial Airline Pilot Threat Management Behaviour

Simon Banbury[1], Helen Dudfield[2] & Mike Lodge[3]

[1]School of Psychology, Cardiff University, Cardiff, CF5 1LP, UK. Email: *banburys@cardiff.ac.uk*
[2]QinetiQ, Farnborough / [3]British Airways.

Abstract: The present study extends previous models of threat management (e.g. Helmreich et al., 1999) by taking a cognitive approach to understand how pilots behave during threat situations. The CAPT-M model was developed by a group of psychologists and airline training captains to describe the behaviour of commercial airline pilots engaging in effective threat management. The present study attempted to identify the core behaviours that pilots use to manage threat situations successfully, and examine how important these behaviours are perceived to be in terms of both their importance to managing threat situations and their potential for training. These data were then used to validate the CAPT-M model. The present study used a variety of user-consultation techniques, such as group discussion and questionnaire-based methods on 34 commercial airline pilots. The findings revealed tentative support for the structure and content (i.e. component stages) of the CAPT-M model. Specifically, participants rated situation assessment and re-assessment to be the most important components of the model, as were having clearly defined goal and situation models. Theoretical considerations and practical implications are discussed.

Keywords: threat management / pilot cognition.

Introduction

The present study was conducted as part of European Commission Framework V funded project "Enhanced Safety through Situation Awareness Integration in Training" (ESSAI). The consortium partners consist of QinetiQ, NLR, DLR, Dedale, Thales, British Airways, Aero-Lloyd, Alitalia and the University of Berlin. The ESSAI project seeks to address problems that occur in commercial flight when pilots are confronted with non-normal and emergency situations, or *threats*, for which they do not have the appropriate procedures. These threats may occur because of lack of Situation Awareness (SA) on the part of the crew (there are procedures, but they do not recognise the situation), or may be the result of an unusual chain of events. A potential solution is thought to consist of enhanced training to provide strategies for effective threat management during non-normal or emergency flight operations.

Although a threat can be defined as either expected, such as terrain or predicted weather, or unexpected, such as ATC instructions or system malfunctions (Helmreich, Klinect and Wilhelm, 1999), we advocate that all threats are usually manifest in two ways; firstly, that there are no well-defined procedures that exist to resolve the threat[1]; and secondly, that even if a solution is found, its outcome is uncertain. Clearly, a strategy to manage a threat would be to seek an understanding of the event so that a solution can be found. Once this understanding is reached and the actions for resolving the situation are in place, the event is no longer termed a threat. This is consistent with strategies used in other domains to manage crises. For example, the first course of action by medical staff with patients with severe trauma is to stabilise the patient and assess the problem. Only when this has been achieved will remedial action be undertaken. Once again, the event ceases to become a threat when an understanding of the situation is reached and a successful course of action is instigated.

Situation assessment, or the process of acquiring and maintaining Situation Awareness (for a review see Endlsey, 1995), is an important step in threat management because it provides a state of awareness of the event, and a starting point for any decision-making undertaken to resolve it. Clearly, accurate situation assessment will lead to more effective threat management. Furthermore, the level of understanding of the event will dictate the behaviour that follows. Rasmussen's (1983) model of human performance provides a

[1] However, it also must be mentioned that some 'threats' do have established recovery procedures (e.g. reaction to a GPWS or TCAS warning), although the outcome is not always guaranteed (e.g. due to circumstances outside of design parameters). In addition, the pilot may decide not to instigate recovery procedures given their knowledge of other traffic or traffic information from ATC.

convenient framework to describe such behaviour during a threat situation. This framework assumes that behaviour can be represented at three levels; a skill-based level that utilises automated sensori-motor patterns that have been built up with practise; a rule-based level that operates on a recognition basis where the behaviours for known tasks are retrieved as required; and finally a knowledge-based level for which no 'know-how' rules are available from previous encounters, necessitating a higher conceptual level of control where goals and strategies are explicitly considered.
The very nature of a threat situation dictates that skill and rule-based behaviours are impossible given that the pilot has not encountered the situation before and there may not be formal procedures that can be used to deal with it. These limitations imply that pilots have to manage and solve the threat event through their own abilities using a knowledge-based strategy. An important consideration of these types of behaviour is the amount of cognitive effort required. On one hand skill-based behaviours rely heavily on highly practised, automatic processing, requiring little cognitive effort. On the other hand, knowledge-based behaviours require significant levels of cognitive effort in order to evaluate goals and decision options. This is rather unfortunate given that uncertain information, time pressure, high workload and high levels of tension and anxiety often typify threat situations.

A Cognitive Model of Commercial Airline Pilot Threat Management (CAPT-M): In order to develop a training package to enhance the effectiveness of threat management, it was first necessary to gain an understanding of how pilots should 'ideally' manage threat situations. A small working group of psychologists and airline training captains was assembled to develop a cognitive model to describe the behaviour of pilots engaging in effective threat management. This model was based on current training practices, anecdotal evidence (e.g. accident and incident case studies and personal experiences), theories of human cognition (e.g. situation awareness, decision making), and descriptive models of threat management (Helmreich, Klinect and Wilhelm 1999).

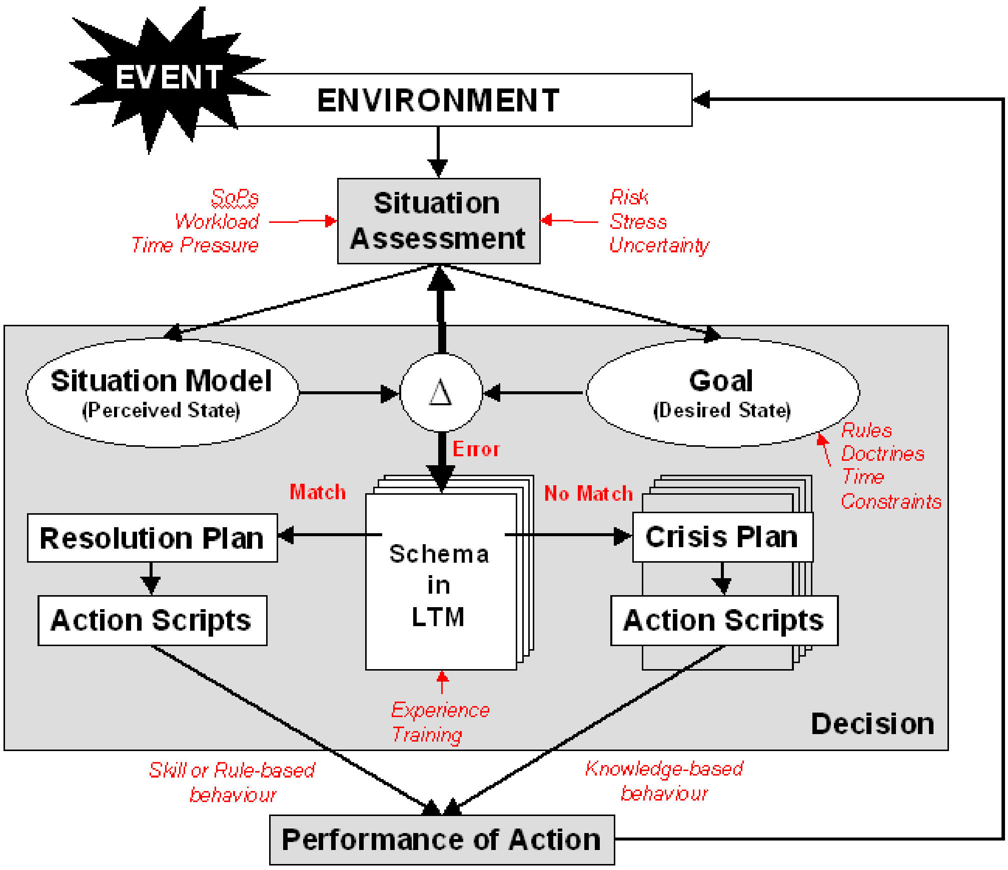


Figure 1 – A model of Commercial Airline Pilot Threat Management (CAPT-M) behaviour

The CAPT-M model was proposed to describe the relationship between Threat Management and Situation Awareness, and is based on the OODA (Observation, Orientation, Decision, Action) loop (Fadok, Boyd and Warden, 1995). The model also extends research by Trollip and Jensen (1991) on the cognitive judgement of pilots. They suggest that pilots utilise a eight-step process to solve problems on the flightdeck: vigilance, problem discovery, problem diagnosis, alternative generation, risk analysis, background problem (e.g. incidental factors), decision and action. The central tenets of the model are that the process is cyclical and adaptive (Neisser, 1976), and that the precursor to any decision-making behaviour is Situation Assessment (Endsley, 1995; Prince and Salas, 1997). The model is presented in Figure 1.

After the onset of an unexpected event in the environment, the process of Situation Assessment occurs. The result of which is a Situation Model comprising of the perceived state of the situation (i.e. Situation Awareness) and a Goal Model, the desired state of the situation as determined by procedures and doctrines (e.g. maintaining the safety of the aircraft and passengers). Workload, stress, time pressure and uncertainty mediate the quality of the Situation and Goal Models.

A comparison is then made between the Situation Model and the Goal Model to determine the extent of the threat. The level of discrepancy between the two also dictates the amount of intervention that is required to reach the Goal. In other words, no discrepancy means that the current course of action will reach the goal, whilst a large discrepancy indicates that intervention is required to ensure that the goal is reached. It is assumed that when an unexpected event occurs in the environment, and is accurately assessed by the pilot, its effect would be to cause a large discrepancy between the perceived and desired state. This is concept is similar to Finnie and Taylor's (1998) IMPACT (Integrated Model of Perceived Awareness ConTrol) model which argues that the acquisition and maintenance of SA is derived from behaviour directed to reduce the mis-match between the perceived level of SA and the desired level of SA.

The 'problem' (i.e. the intervention needed to resolve the discrepancy between desired and actual state) is now represented in memory and is compared to existing schema in long term memory. If the pilot feels that he or she has insufficient information to form a Situation Model and/or Goal Model, further Situation Assessment is then undertaken.

Experienced decision-makers working under time pressure report that they use recognition-based rather than evaluation-based decision making strategies; acting and reacting on the basis of prior experience rather than comparing decision options through formal or statistical methods. Recognition-primed decision making fuses two processes; situation assessment and mental simulation. Situation assessment generates a plausible plan of action, which is then evaluated by mental simulation (Klein, 1993). In line with Klein, the model proposes that schema attempt to fit expected perceptions and are fine-tuned by experience, in both a bottom-up and top-down fashion. In bottom-up processing, information of perceived events is mapped against existing schema on the principle of best fit. Whereas in top-down processing, anomalies of fit are resolved by the fine-tuning of the evoked schema in the light of perceived evidence, or by initiating searches to fit the newly changed schema structure (Klein, 1993).

As the event is only classed as a threat if there are no well-defined procedures, clearly no schema will exist to assist the pilot in resolving the threat situation (i.e. bottom-up). With no match to schema in memory, the pilot is faced with producing a bespoke Crisis Plan in order to stabilise the situation (i.e. top-down). However, although they may share surface features with the problem, they may not be appropriate to use. A course of action is decided and then acted upon. Once again, the situation is re-assessed and the resultant Situation Model is compared to the desired goal. This process is repeated until the bespoke crisis plan creates an intervention that does have a schema-match in memory. At this point, the event ceases to be a threat and can be managed using existing procedures, or Resolution Plans. The difference between Crisis and Resolution plans can be couched in terms of Rasmussen's model of human performance. On the one hand, Resolution plans operate on a rule-based, recognition basis where behaviours for known tasks are retrieved as required. On the other hand, Crisis plans operate on a knowledge-based level for which no 'know-how' rules are available from previous encounters, necessitating a higher conceptual level of control where goals and strategies are explicitly considered.

The Present Study: The present study attempted to identify the core behaviours that pilots use to manage threat situations successfully, and examine how important these behaviours are perceived to be in terms of both their usefulness to threat management and their importance to training. The present study used a variety of user-consultation techniques, such as group discussion and questionnaire-based methods on 34 commercial airline pilots. Participants were asked to respond to questions relating to the strategies that they use to manage threat situations. These responses were used to help validate the proposed model of Threat Management (CAPT-M). In addition, a number of demographic measures were taken.

Method
For the purposes of this study, a formal definition of the term 'threat' was used. Preliminary studies had shown little consensus between pilots when asked to define a 'threat'. To ensure consistency of responses between the participants in this study the following definition was used; a threat is an unexpected and potentially life-threatening chain or combination of events, causing uncertainty of action and time-pressure. This can range from a hazardous situation to a major crisis.

A group discussion between all participants was held at the beginning of the experimental session to clarify what is meant by 'threat' and 'threat management'. Participants were then asked to answer the questionnaires in light of a situation they had experienced which was representative of this agreed definition.

Demographics: Participants were asked to give the following information:

- Flying Hours – Specifically, the number of commercial, military and private flying hours.
- Types of Aircraft flown – Specifically, the aircraft model and type, and the number of hours (both as Captain and First Officer) they had flown in each. From these data, the number of hours flown in glass, hybrid and steam cockpits were calculated.

Validating the Threat Management model: Participants were asked to answer questions relating to threat management strategies they have used in the past to manage threat events. They were asked to indicate their agreement or disagreement with a number of statements (see below) in light of a recent event they had encountered before that was representative of our definition of a threat. The scale was constructed in a 5 point Likert format: Strongly Disagree, Disagree, Neutral, Agree and Strongly Agree. Participants' responses to these questions were also used to validate a model of threat management. Unbeknownst to the participants, these questions mapped directly on to the components of the threat management model. Thus, participant responses to each of these statements were used as evidence for, or against, the stages of the proposed model of threat management.

Stage of Model	Statement
Situation Assessment	It is important to make an assessment of the current state of the situation.
Situation Model (or perceived state)	It is important to hold a representation of the current state of the situation in my mind.
Goal Model (or desired state)	It is important to hold of representation of my current goal or goals in my mind.
Comparison of Goal and Situation Model	It is important to reflect on the differences between where I am now and where I want to be.
Schema in LTM	It is important to compare my perception of the current situation with past experiences.
Resolution Plan	It is important to take a course of action that I have encountered before, rather than embark on an unknown one.
Crisis Plan	It is important to take a course of action that I have not encountered before, rather than do nothing.
Action Scripts	It is important to formalise the details of the action before instigating them.
Iterations	It is important to re-assess the current situation once any action has been performed.

Skills important to Threat Management: In addition, participants were asked to rate their agreement with a number of skills (see below) for how important they are to threat management. Once again, the scale was constructed in a 5 point Likert format: Strongly Disagree, Disagree, Neutral, Agree and Strongly Agree.

1. Situation Assessment

2. Risk Assessment
3. Risk Taking
4. Experience
5. Team-work
6. Inter-personal communication
7. Leadership
8. Communication
9. Checklist Management
10. Systems Knowledge
11. Task Management
12. Attention Management
13. Aircraft Energy Awareness
14. Option generation
15. Option selection

'Trainability' of Threat Management skills: Finally, participants were asked to rate their agreement with a number of skills that could be trained to improve threat management (see below). Once again, the scale was constructed in a 5 point Likert format: Strongly Disagree, Disagree, Neutral, Agree and Strongly Agree.

1. Situation Assessment
2. Risk Assessment
3. Team-work
4. Verbal Communication
5. Leadership
6. Non-verbal Communication
7. Checklist Management
8. Systems Knowledge
9. Task Management
10. Attention Management
11. Aircraft Energy Awareness
12. Option generation
13. Option selection
14. Task Prioritisation
15. Workload Management

Results
Demographics: Participants were 34 flight-crew employees of a major international airline (33 male and 1 female). Of the sample, 18 were Captains, 10 were First Officers and 6 were Flight Engineers. All participants were English speaking, British nationals.

The number of hours flown as Captain, First Officer and Flight Engineer were as follows:

Position	Total Hours of Sample	Mean Hours of Sample
Captain	82320	4573 (3495.36)
First Officer	156730	6269 (2599.79)
Flight Engineer	59580	9930 (4108.21)

(Standard Deviations in brackets)

The amount of hours flown in steam, hybrid and glass cockpits were as follows:

Cockpit	Total Hours of Sample	Mean Hours of Sample
Steam	237135	7411 (4790.81)
Hybrid	11050	1842 (867.42)
Glass	41220	2576 (1981.71)

(Standard Deviations in brackets)

The flying hours of the participants were as follows:

Type	Total Hours of Sample	Mean Hours of Sample[2]
Commercial	293650	8637 (3961.05)
Military	9303	274 (762.86)
Private	22152	651 (955.60)

(Standard Deviations in brackets)

Validating the Threat Management model: Participants were asked to indicate their agreement or disagreement with a number of statements relating to threat management, specifically in the context of a recent threat event. Unbeknownst to the participants, these questions mapped directly on to the components of the threat management model. Participant responses to the statements (and stage of the model they represent) are presented below. The scoring of these items were SD=1, D=2, N=3, A=4, SA=5.

Stage of Model	Statement	Mean	Standard Deviation
Situation Assessment	1. I made an assessment of the current state of the situation	4.3	0.95
Situation Model (or perceived state)	2. I consciously thought through the current state of the situation	3.9	0.93
Goal (or desired state)	3. I held a representation of my current goal or goals in my mind	4.0	0.69
Comparison of Goal and Situation Model	4. It is important that I reflected on the differences between where I was and where I wanted to be	3.9	0.80
Schema in LTM	5. I compared my perception of the current situation with past experiences of similar situations	3.8	1.02
Resolution Plan	6. It was important that I took a course of action that I had encountered before, rather than embarked on an unknown one	2.8	1.04
Crisis Plan	7. It was important that I took a course of action that I hadn't encountered before, rather than doing nothing	3.2	1.19
Action Scripts	8. I formalised the details of the action before I instigated them	3.2	1.00
Iterations	9. I re-assessed the current situation once any action had been performed	4.3	0.68

A one-way within-subjects analysis of variance showed significant differences between the ratings for the nine statements, $F(8,305)=10.87$, $p<0.001$. Post hoc Newman Keuls showed that participants agreed with statements 1, 2, 3, 4, 5 and 9 significantly more than statements 6, 7 and 8 ($p<0.05$).

Skills important to Threat Management: Participants were asked to rate their agreement with a number of skills for how important they are to threat management. The scoring of these items were SD=1, D=2, N=3, A=4, SA=5.Participant responses were as follows:

[2] If those without any military or private flying experience are excluded from this analysis the mean 'Military' flying hours was 1760 hours and the mean 'Private' flying hours was 852 hours.

Factors	Mean	Standard Deviation
1. Situation Assessment	4.8	0.39
2. Risk Assessment	4.4	0.66
3. Risk Taking	2.3	1.06
4. Experience	4.5	0.56
5. Team-work	4.9	0.29
6. Inter-personal Communication	4.9	0.33
7. Leadership	4.4	0.54
8. Communication	4.8	0.41
9. Checklist Management	4.3	0.57
10. Systems Knowledge	4.3	0.63
11. Task Management	4.4	0.50
12. Attention (i.e. arousal) Management	4.2	0.61
13. Aircraft Energy Awareness	4.5	0.71
14. Option generation	4.2	0.48
15. Option selection	4.5	0.62

A one-way within-subjects analysis of variance showed significant differences between the ratings for the 15 factors, $F(14,509)=38.04$, $p<0.001$. Post hoc Newman Keuls showed that participants agreed with factor 3 (risk taking) significantly less than all of the other factors ($p<0.05$), and agreed with factors 5 (team-work) and 6 (inter-personal communication) significantly more than all of the other factors ($p<0.05$).

'Trainability' of Threat Management skills: Participants were asked to rate their agreement with a number of skills that could be trained to improve threat management. The scoring of these items were SD=1, D=2, N=3, A=4, SA=5. Participant responses were as follows:

Factors	Mean	Standard Deviation
1. Situation Assessment	4.2	0.67
2. Risk Assessment	4.1	0.55
3. Team-work	4.2	0.74
4. Verbal Communication	3.7	0.97
5. Leadership	3.5	0.99
6. Non-verbal Communication	3.0	0.89
7. Checklist Management	4.4	0.50
8. Systems Knowledge	4.7	0.45
9. Task Management	4.1	0.64
10. Attention Management	4.2	0.74
11. Aircraft Energy Awareness	4.4	0.55
12. Option generation	3.9	0.66
13. Option selection	3.9	0.69
14. Task Prioritisation	4.2	0.89
15. Workload Management	4.2	0.72

A one-way within-subjects analysis of variance showed significant differences between the ratings for the 15 factors, $F(14,509)=10.73$, $p<0.001$. Post hoc Newman Keuls showed that participants agreed with factor 6 (non-verbal communication) significantly less than all of the other factors ($p<0.05$), and agreed with factor 8 (systems knowledge) significantly more than all of the other factors ($p<0.05$).

Correlation Analyses: A number of correlation analyses were conducted between:
Flying Hours (Commercial, Military and Private)
Ratings for the nine Threat Management strategies representing the stages of the CAPT-M model.

The results of the correlation analyses are presented in the table below:

Variable 1	Variable 2	Pearson's (r)	Probability
Commercial Hours	HM1 – "Assess current state"	-0.533	0.001
Private Hours	HM9 – "Reassess situation"	-0.566	0.001

Correlation analyses indicated that the more commercial flying experience participants had, the less they agreed with the statement "I made an assessment of the current state of the situation". In addition, the more private flying experience participants had, the less they agreed with the statement "I re-assessed the current situation once any action had been performed".

Discussion

Summary of findings: Results from the preliminary validation of the threat management model revealed tentative support for its structure and content (i.e. component stages). Specifically, participants rated situation assessment and re-assessment to be the most important components of the model, as were having clearly defined goals and situation models. However, participants were less enthusiastic about the components of the model relating to selecting the course of action. There may be a number of explanations for the latter finding. Firstly, it may be difficult for participants to introspect about their own cognition during decision making. Secondly, it is plausible that the list of statements about the model did not include a number of other decision options. For example, doing nothing and letting the situation develop is sometimes an effective strategy when an accurate situation assessment is not possible at the onset of the threat situation. In addition, the results suggest that the pilots in the present study were not adverse to taking an unknown course of action if required.

Analysis of the participant responses to what factors were important in threat management indicated that there was most agreement with Situation Assessment and Teamwork. There also agreement, albeit to a lesser extent, with Communication (both inter-personal and with air-traffic control) and Option Selection. These results are consistent with our assumption that situation assessment is the precursor to effective threat management. In addition, participants also reported that team communication and option selection were important for threat management. This view can be supported by the Sioux City DC10 incident in which two pilots concentrated on stabilising the situation and flying the aircraft while the other two concentrated on generating options to solve the problems caused by a loss of hydraulic flying controls. A controlled crash was thus achieved at a planned airfield with emergency services standing by, resulting in 184 survivors. Finally, there was common disagreement with Risk Taking being an important factor in threat management. This is consistent with airline procedural attitudes towards risk taking where it is not advocated except in extremis when no viable option presents itself (e.g. in a ditching or crash landing).

Analysis of the participant responses to what factors can be trained indicated that there was most agreement with Systems Knowledge and the least agreement with Non-verbal Communication. These results suggest that pilots in the present study believe that technical skills (i.e. Systems Knowledge) training is more important to managing threat situations than non-technical skills training (e.g. co-operation, leadership and management skills, situation awareness and decision making; Avermaete, 1998). This perceived bias may originate in the type of NTS training that the majority of flightcrew have historically received. The early CRM training centred largely on interpersonal activity and became widely discredited among the flying fraternity since it was seen as being remote from real operational issues. Current NTS education has begun to make progress in redressing this complaint.

It is also interesting to note their resistance to non-verbal communication training, despite it being an important information channel between team members. Clearly, training that focusing on understanding non-verbal cues will enhance communication and shared awareness of goals and intentions, especially in glass cockpits where non-verbal communication cues are already impoverished.

Correlation analyses indicated that the more commercial flying experience participants had, the less they agreed with the statement "I made an assessment of the current state of the situation". This finding may reflect differences in training between commercial and ex-military pilots insofar as traditional civilian training emphasises knowledge of existing procedures and the choice of the correct one, whereas military training emphasises situation assessment more. In addition, the more private flying experience participants had, the less they agreed with the statement “I re-assessed the current situation once any action had been performed”. This may reflect the fact that most private flying is solo, thus making it harder to reassess the situation.

Implications: When discussing the results of the present study in terms of SA it is important to differentiate between SA the ‘product’ and SA the ‘process’. Very simply, the process of SA refers to how SA is developed and maintained during flight, while the product is the resultant, elusive thing we call SA itself (Dominguez, 1994). Indeed, Endsley (1995) distinguishes the term situation awareness as a state of knowledge from the processes used to achieve that state. These processes, which may vary widely among individuals and contexts. Endsley refers to the process of achieving, acquiring and maintaining SA as ‘situation assessment’. Indeed, the results of the study support indicated that situation assessment was rated as an important component of threat management. Further, participants also rated the accurate formulation of situation and goal models to be equally as important.

The high ratings by participants for the importance of situation re-assessment, support our view that the process of threat management is reliant on an active and cyclical process. In addition, this process is also adaptive. This is in line with Neisser (1976) who formulated the concept of a ‘Perceptual Cycle’, where the interaction between human and environment shapes the human’s perceptions and actions. Neisser argued that the structure of our knowledge and expectations of some aspect of the world (i.e. schema) are always activated, but that the activation of particular schema is as an oriented response to the environment. This oriented response selects new information for attention that in turn activates appropriate schema and so on. In terms of threat management, we believe that pilots’ awareness of the environment or situation is actively modified by their changing appreciation of information gleaned from that environment. One key to successful threat management then, is to ensure that the pilots are oriented to the critical threats and pertinent information in the environment, through technological and/or training solutions (e.g. ESSAI).

Given the importance of situation assessment and re-assessment to successful threat management, the ESSAI project is developing and evaluating a training package that enhances threat management through SA training. Particular importance is placed on promoting high levels of SA (i.e. projection) to allow pilots to ‘avoid’ threats, rather than ‘trap’ or ‘mitigate’ them (see Helmreich et al., 1999). This training is designed to augment, and not replace, existing technical skill-based training.

Further Development of the Model: The present model describes the threat management at an individual level only and does not take into account the interaction between crewmembers. However, it is relatively straightforward to imagine how this might be achieved. Although the threat and the environment are the same for each crewmember, but the ‘perception’ of the situation (i.e. situation assessment) will be unique for each crewmember. Thus, the sharing of information between team members is critical for successful threat management. Specifically, it is desirable that crewmembers share information pertaining to their own situation model and goal (i.e. where they think they are and where they think they should be). Although past experiences and training background are also unique for each individual, the results of any matching to schema in memory should be communicated between crewmembers. At the very least, this ensures that all parties are cognisant of each other’s line of thinking, but this information may also cue others to recall similar experiences. Once the action or resolution plan has been decided, the performance of the appropriate actions must be also co-ordinated to ensure the correct execution of the plan.

The advantage of extending the model in such a way is that it identifies communication links between crewmembers at all stages of the threat management process. When these communication links are broken (e.g. due to high workload or poor CRM), it is possible to predict the outcome of these failures on the threat management process.

Finally, the model has been developed into the Index of Threat Management (ITMS) questionnaire through the inclusion of questions relating to threat anticipation, prioritisation and communication. The goal is to use the CAPT-M model as a benchmark for the evaluation of ESSAI training. The efficacy of the training intervention can be, in part, evaluated through the self-report of participants on the ITMS questionnaire, directly after each simulator session.

Conclusions

The present study extends previous models of threat management (e.g. Helmreich et al., 1999) by taking a cognitive approach to understand how pilots behave during threat situations. As discussed, the model will be used in the design and evaluation of training to improve the effectiveness of commercial airline pilot threat management (e.g. ESSAI). However, further empirical research is needed to fully validate the model. Such empirical work should include the generation and testing of predictions from the model about behaviour during threat situations.

References

Avermaete, J. A. G. (1998). NOTECHS: Non-technical skill evaluation in JAR-FCL. NLR technical report NLR-TP-98518.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37, 65-84.

Fadok, D. Boyd, J., and Warden, J. (1995). Air Power's Quest for Strategic Paralysis. Maxwell AFB, Ala.: Air University Press.

Finnie, S. and Taylor, R. M. (1998). The Cognitive Cockpit. Flight Deck International. UK: International Press.

Helmreich, R. L., Klinect, J. R., and Wilhelm, J. A. (1999). Models of threat, error and CRM in flight operations. In Proceedings of the 10th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Klein, G. A. (1993). A recognition-primed decision (RPD) model for rapid decision-making. In: G. A. Klein, J. Orasanu, R. Calderwood and C. Zsambok (Eds.). Decision making in action: Models and methods. New Jersey, USA: Ablex.

Neisser, U. (1976). Cognitive Psychology. New York: Appleton-Century-Crofts.

Oppenheim, A. N. (1992). Questionnaire design, interviewing and attitude measurement. Pinter: London.

Prince, C., Salas, E. (1997) The Role Of Situation Assessment In The Conduct Of Flight And In Decision Making, In: Harris, D. (Ed.) Engineering Psychology and Cognitive Ergonomics, Vol. 1, pp. 291-297.

Trollip, S., and Jensen, R. (1991). Human Factors For General Aviation. Jeppesen Sanderson Inc.: Englewood.

Pilot Control Behavior in Paired Approaches

Steven J. Landry and Amy R. Pritchett

School of Industrial and Systems Engineering, Georgia Institute of Technology,
765 Ferst Drive, Atlanta, GA 30332-0205

Abstract: A piloted flight simulator study investigated pilot control behavior while self-separating from a proximate aircraft during an instrument landing system approach. This "paired approach" operational concept pairs two aircraft, with the trail aircraft of the pair remaining within a safe zone calculated to be free (for a specified period of time) from collision danger and wake turbulence. Enabling this operation would allow aircraft to land simultaneously on closely spaced parallel runways in poor weather. Pilots in this study flew the trail aircraft and were shown either the current safe zone, a "worst-case" safe zone based on the approach procedure, or both. The lead aircraft's behavior was varied so that it either followed the approach procedure, changed its speed unexpectedly, or blundered across the trail aircraft's approach course. Pilots were expected to track the safe zone, resulting in different behavior depending on the type(s) of safe zone displayed. The results instead suggest pilots attempted to match the lead aircraft speed. Since procedures and displays to date have not anticipated this type of pilot behavior, changes to support it, or training to overcome it, may be necessary.

Keywords: Paired approaches, self-separation, design of human-integrated systems.

Introduction

In order to reduce air traffic controller workload, reduce communications requirements between controller and pilot, or to eliminate communications delays between controller and pilot, moving some aircraft to aircraft separation responsibility to the flight deck has been widely proposed. This is referred to as "self-separation", and requires additional new monitoring and control behaviors from pilots. Pilots must monitor their position relative to other aircraft, and maneuver their aircraft to remain at the proper distance from those aircraft.

One example of self-spacing is found in a proposed operation called "paired approaches", which places two aircraft on instrument approaches to closely spaced parallel runways with one aircraft offset behind the other. The trail aircraft maintains a position relative to the lead aircraft (Stone, 1996; Pritchett, 1999; Hammer, 1999) that guarantees that neither aircraft will be in danger of loss of separation within a certain time window should the other "blunder" (i.e. depart its approach path), and that neither aircraft will be affected by the other's wake. This range of positions is called the safe zone (shown in Figure 1), and is the range of current positions that allows the trail aircraft to remain on its approach and still have 500-foot separation with a blundering lead aircraft, and also pass in front of the wake vortex. The calculations for the position of the safe zone are based on the aircraft positions, speeds, the amount of time for which protection is provided, and the crosswind speed.

Two different underlying bases can be used to determine the safe zone (Pritchett and Landry, 2001). The first uses procedural information; i.e. a "predicted" safe zone can be calculated assuming that the aircraft are following a pre-specified approach procedure, thereby presenting a spatial boundary which is predictable, small and stable, but which does not account for either aircraft not complying with the approach procedure. The position of the front of the predicted safe zone is pre-calculated by using the worst-case position and speed that is allowed under the procedure. The second is based on real-time information; i.e. the "actual" safe zone is recalculated throughout the approach based on the current states of both aircraft, thereby presenting a spatial boundary which is as large as possible for the immediate context, and constantly (sometimes rapidly) changing in size and location.

From an implementation viewpoint, these two bases for a safe zone are important considerations because each may have different equipment and procedural requirements. The actual safe zone requires a broadcast of the lead aircraft's position and speed. The trail aircraft must have the capability to receive this information, and also have the means to rapidly calculate and display the safe zone. The predicted safe zone could be calculated in advance, and would not, in theory, require any special equipment except an indication of the longitudinal separation from the lead aircraft in the flight deck of the trail aircraft.

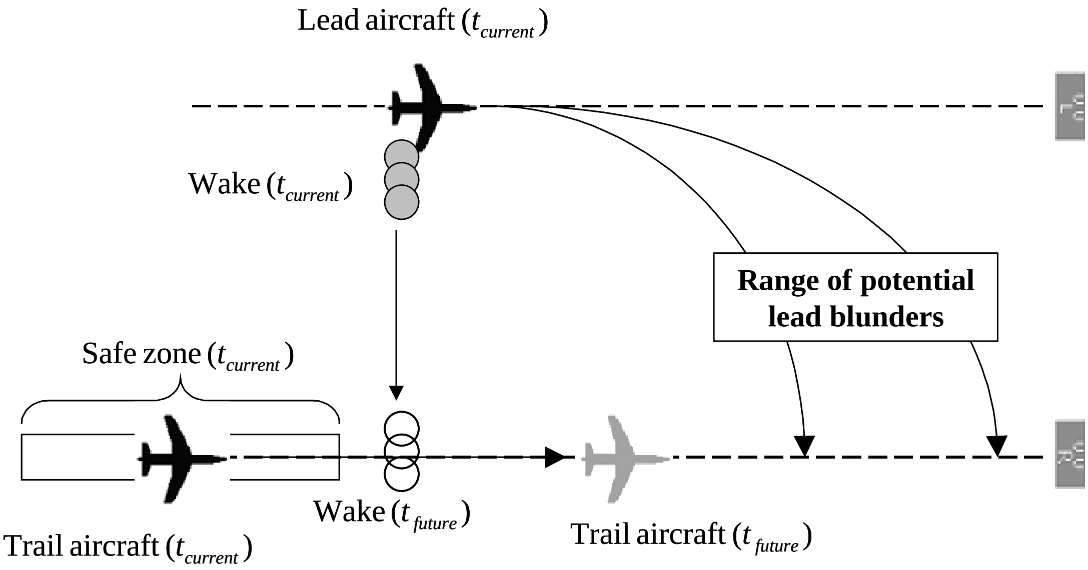


Figure 1 – Schematic of the safe zone

The approach procedures would also have to be different. The trail aircraft, if given only the actual safe zone, would have to remain within the safe zone regardless of the behavior of the lead aircraft, and follow a missed approach procedure if the safe zone were departed. This missed approach procedure would probably be a predetermined maneuver consisting of a turn away from the lead aircraft's approach path, a climb, and acceleration. If given the predicted safe zone, the trail aircraft would also have to remain within the safe zone, and would have to perform a missed approach if those limits were exceeded. However, the trail aircraft would also have to execute a missed approach if either aircraft violated the assumptions of the predicted safe zone.

Both safe zones are defined as a range of positions relative to the lead aircraft. As the lead aircraft changes position, both safe zones move with it. Movement of the trail aircraft relative to the lead aircraft is therefore also movement relative to the safe zone. For example, if the trail aircraft is closing on the lead aircraft, then it would also be closing on the front of the safe zone (and moving away from the back of the safe zone).

For the actual safe zone, the safe zone has a second source of movement relative to the trail aircraft. Since the safe zone is continuously updated based on the current speeds and positions of the two aircraft, its position relative to the lead aircraft can be changing. For example, as the trail aircraft increases (or the lead aircraft decreases) its airspeed, the safe zone needs to be further in trail of the lead aircraft. So, if the lead aircraft slows, not only will the trail aircraft begin closing on the lead aircraft (and the front of the safe zone), but the front of the safe zone would be moving away from the lead aircraft (and back towards the trail aircraft). The movement of the front of the safe zone may therefore be based on several factors, which could be difficult for the pilot to understand.

Pilot control and monitoring strategies are implicit in the type of safe zone displayed to the pilot, as summarized in Table 1. If given the actual safe zone, pilots may choose to remain at a particular position relative to the front of the safe zone, with control movements consistent with this tracking. Since the actual safe zone is dynamic, control movements may be frequent. Since the actual safe zone could potentially change faster than the pilot could react, control movements may be somewhat severe as well. If given the predicted safe zone, pilots may also try to remain at a given distance from the front of the safe zone. However, since the predicted safe zone is relatively stable, fewer control movements would have to be made. In addition, these control movements need not be severe due to the relatively static nature of the predicted safe zone.

Table 1 - Monitoring strategies.

	Predicted Only	Actual Only	Both
General control strategy	Stay within safe zone (predicted safe zone is small but stable).	Stay within safe zone (actual safe zone is large but dynamic).	Generally stay within predicted safe zone, but can briefly depart if within actual safe zone.
Measures of control strategy	Position within safe zone, correspondence of throttle movements to deviations of position within safe zone	Position within safe zone, correspondence of throttle movements to deviations of position within safe zone.	Position within predicted safe zone (or actual if predicted departed). Correspondence of throttle movements.
General monitoring strategy	Occasional checks on position in safe zone. Conformance monitoring of lead aircraft.	Frequent checks on position within safe zone.	Occasional checks on position within predicted safe zone. More frequent if outside of predicted safe zone.
Measures of monitoring strategy	Stable position maintained within safe zone. Able to detect lead aircraft noncompliance.	Stable position maintained within safe zone.	Stable position maintained within predicted and/or actual safe zone. Able to detect lead aircraft noncompliance.
Reaction to noncompliance	Should recognize noncompliance that will invalidate the safe zone	Should execute a missed approach only upon departing safe zone.	Should execute a missed approach upon departing actual safe zone.

In addition to the control strategies, pilots would have additional monitoring tasks. When given the actual safe zone, pilots would have to frequently monitor their position within the safe zone. If given the predicted safe zone, pilots would not have to monitor their position within the safe zone as frequently (since the predicted safe zone is fairly static), but would also have to monitor the lead aircraft for conformance to the procedure (since the predicted safe zone would be invalid if the lead aircraft does not comply with the approach procedure).

If both safe zones are displayed, pilots may choose to utilize the better features of each. The pilots may be able to track the predicted safe zone, resulting in less monitoring of the safe zone and less control movements, while also monitoring their position relative to the actual safe zone to reduce the need to monitor the lead aircraft's compliance.

The reaction of pilots to a lead aircraft that was not conforming to the approach procedure would also be different for each of the safe zones. If the actual safe zone is displayed, and the lead aircraft did not comply, the trail aircraft would not have to take any action except to try to remain within the safe zone, and perform a missed approach if they depart the safe zone. If the predicted safe zone is displayed, the trail aircraft would have to consider whether the safe zone was still valid, and perform a missed approach if it is not.

This study examined whether pilots would be able to fly a stable instrument approach when given the additional task of acting as the trail aircraft and tracking the safe zone. In addition, the two different underlying conceptual bases of the safe zone were studied, since they may require different control strategies and foster different types of monitoring for unusual situations. Finally, the study may give general insight into both paired approaches and, more generally, self-separation tasks.

Stability of the approach was measured by both control movements (where smaller and generally fewer control movements would indicate stability), and approach error (as given by error about the desired glidepath). Control strategy was evaluated by examining whether there was a desired position within the safe zone that the pilot tried to maintain. Monitoring was studied by questioning the pilots concerning the compliance of the lead aircraft (which would be varied during the experiment) and evaluating their responses.

Method

Apparatus: Participating pilots (12 male airline pilots current or previously qualified in glass cockpit aircraft) were asked to fly approaches using Georgia Tech's Reconfigurable Flight Simulator (RFS) (Ippolito and Pritchett, 2000). The RFS is a medium fidelity simulator running on a Pentium III desktop computer. The simulator was configured with a dynamic model and cockpit systems representing a Boeing 747-400. The navigation display (ND) included an overlay of traffic information about the aircraft on the other approach and the safe zone presentations, which were displayed as staple shaped brackets (Figure 2).

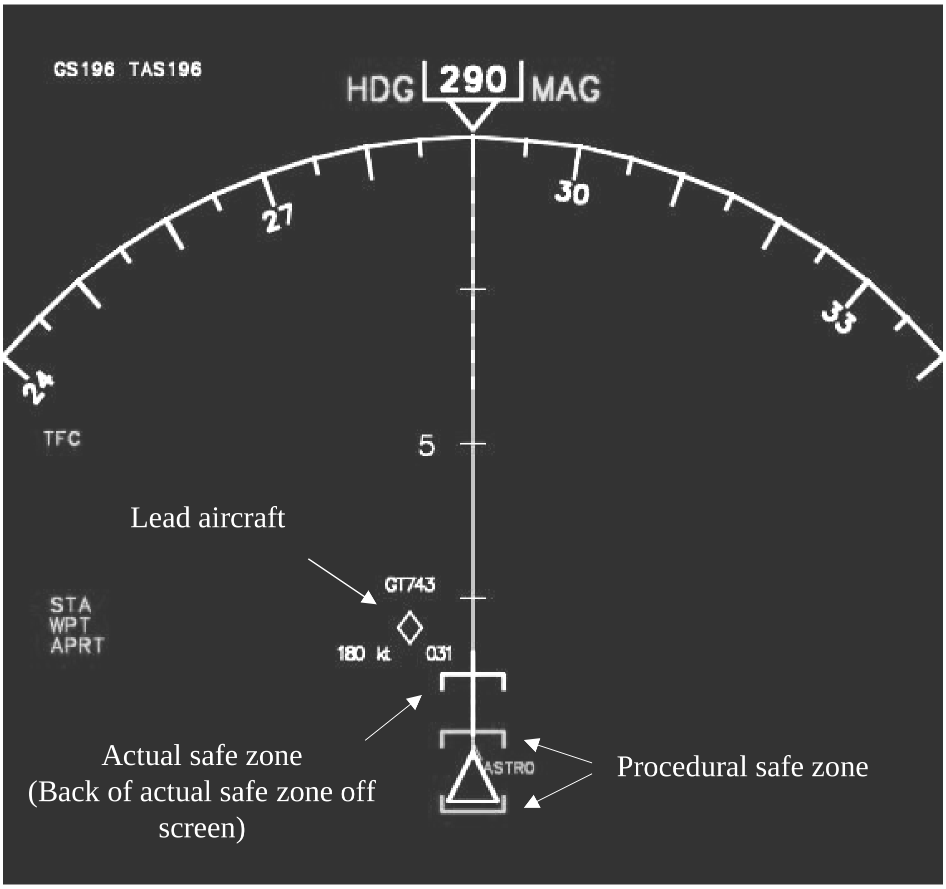


Figure 2 - Navigation display

Procedure: Pilots were given detailed briefings on the simulator and the procedure, and given an opportunity to practice with each until they felt comfortable. In the briefing on the safe zone, it was stressed that a position within the actual safe zone was safe for the next 30 seconds from collision and wake turbulence regardless of the actions of either aircraft, while a position within the predicted safe zone was similarly safe, but only in the 30 seconds following noncompliance from the approach procedure by either aircraft. If the safe zone was departed, this protection was no longer guaranteed, and it was recommended that a missed approach be executed. The missed approach procedures were provided on the approach plate, and indicated both a climb and a turn away from the other approach path.

The pilots were instructed to fly an instrument landing system approach, while remaining within the safe zone. This type of approach relies on a broadcast signal indicating the extended runway centerline (localizer), and a separate signal indicating the proper vertical profile (glideslope). The pilot is given a display of his or her deviation from those ideal trajectories and must make corrections to return to the proper course and glideslope.

The pilots flew the trail aircraft, with the lead aircraft being a scripted pseudo-aircraft. Each run began at approximately 20 miles from runway threshold on the localizer and at approximately 200 knots true air speed (KTAS). The participants were instructed that ATC had told them (and the lead aircraft) to maintain 180 KTAS, plus or minus 10 knots, until 5 miles from runway threshold, where they could slow to their normal approach speed of 148 KTAS.

Experiment Design and Independent Factors: Each participant pilot flew 10 data collection runs. The first nine runs represented a two-factor design with three safe zone displays and three noncompliance types. The three displays refer to the conceptual basis of the safe zone, as follows:

- Predicted safe zone display: The predicted safe zone was shown on the ND.
- Actual safe zone display: The actual safe zone was shown on the ND.
- Both safe zones display: Both safe zones were shown on the ND, allowing the pilot to directly compare the two types of information. In this case the pilots were briefed that they could depart the predicted safe zone as long as they remained within the actual safe zone. This display is shown in Figure 2.

The noncompliance type refers to the type of noncompliance committed by the lead aircraft:

- No noncompliance: a baseline in which the lead aircraft complied with all procedural restrictions.
- Speed noncompliance: The lead aircraft slowed substantially below the approach procedure's minimum allowed speed, as if this aircraft were configuring and attaining final approach speed 5-10 miles before allowed by approach procedures.
- Lateral noncompliance: The lead aircraft turned toward and crossed the participant's approach path, in the form of a turn to a new heading commonly used as a noncompliance model.

Once the participant completed these nine runs, he flew a tenth run with one of the three safe zone displays in a combined noncompliance scenario: specifically, the lead aircraft first slowed below the minimum allowed procedural speed, and then the lead aircraft also turned toward and crossed the trail aircraft's approach path. If the pilot did nothing in this situation, he was likely to first exit the front of the safe zone and then be in risk of collision when the lead aircraft turned towards him.

Basic aircraft parameters (position, speed, heading) were recorded throughout the data runs. In addition, pilot control movements (elevator and throttle) were recorded, as was glideslope deviation. Because the aircraft was laterally stable (no disturbances were introduced into the simulator scenarios) and most pilots were able to remain on the localizer without any aileron or rudder movements, measures of lateral-directional control, although recorded, were not used in the data analysis.

Results

For each subject and each experimental run, the control movement and glideslope deviation data were aggregated, providing a mean and standard deviation. For each of the measures, data collected after the pilot initiated a missed approach were removed. Since large changes in throttle and elevator, and large deviations from the glideslope, are undesirable and indicative of an unstable approach, the standard deviation of these measures was used to examine the stability of the approaches. In addition, the number of throttle movements was examined to compare the number of discrete control changes, both by display and noncompliance type, and before and after noncompliance occurred.

An ANOVA was then performed on the standard deviations for the three responses (throttle setting, elevator position, and deviation from glideslope) and on the number of throttle movements using a general linear model with three main factors: subject, display type, and noncompliance type.

For glideslope error and elevator position standard deviation, there were significant differences across subjects, but no significant differences across display or noncompliance type, as shown in Tables 2 and 3.

Table 2- Analysis of Variance for Glideslope Error, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	11	127.0	130.5	11.9	2020.00	0.021
Blunder	3	14.9	15.1	5.0	0.94	0.426
Display	2	5.8	5.8	2.9	0.54	0.583
Error	91	489.8	489.8	5.4		
Total	107	637.5				

Table 3 - Analysis of Variance for Elevator Standard Deviation, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	11	51.5	49.7	4.5	11.55	0.000
Blunder	3	1.1	0.9	0.3	0.79	0.504
Display	2	1.2	1.2	0.6	1.48	0.233
Error	91	35.6	35.6	0.4		
Total	107	89.4				

Table 4 - Analysis of Variance for Throttle Standard Deviation, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Subject	11	6884.1	6871.2	624.7	4.50	0.000
Blunder	3	6281.9	6014.3	2004.0	14.44	0.000
Display	2	223.6	223.6	8.0	0.81	0.450
Error	103	14295.3	14295.3	111.8		
Total	119	27684.9		138.8		

The ANOVA for throttle standard deviation is shown in Table 4, with the main effects shown in Figure 3. Significant main effects were found for subject and noncompliance type, but not for display type. Pairwise comparisons showed that, except for between lateral and speed noncompliance, all pairwise differences between noncompliance conditions were either significant or marginally significant (Both-Lateral, None-Speed). A chi-squared test on the number of discrete throttle movements indicated no significant difference before and after noncompliance occurred.

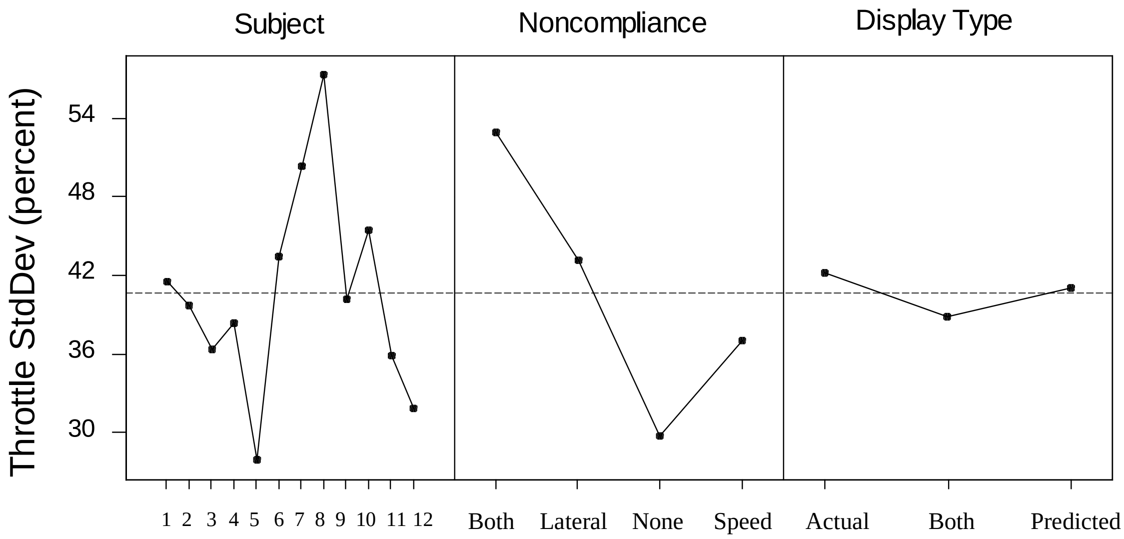


Figure 3 - Main effects plot - throttle setting

Table 5 - P-values for regressions of speed changes against throttle movements

		Subject											
Scenario type	Overall	1	2	3	4	5	6	7	8	9	10	11	12
No Noncompliance	**0.000**	**0.004**	**0.014**	**0.020**	**0.000**	**0.001**	**0.000**	**0.005**	**0.040**	0.289	**0.000**	**0.000**	**0.025**
Pre Speed Noncompliance	**0.001**	0.890	**0.002**	**0.083**	0.148	**0.016**	0.112	0.917	0.119	0.524	**0.068**	**0.013**	0.403
Post Speed Noncompliance	**0.000**	0.485	**0.060**	0.526	**0.002**	**0.055**	**0.000**	**0.002**	**0.012**	0.331	**0.000**	**0.001**	**0.019**
Pre Lateral Noncompliance	**0.010**	**0.002**	**0.062**	**0.079**	**0.001**	**0.000**	0.438	**0.000**	**0.025**	0.544	**0.000**	**0.000**	0.239
Post Lateral Noncompliance	0.142	**0.086**	**0.012**	0.936	**0.037**	**0.000**	0.622	**0.000**	**0.025**	0.538	**0.028**	**0.001**	0.717

Regression examined whether throttle changes could be predicted by either changes to position within the safe zone, changes to the trail aircraft's relative position with respect to the lead aircraft, or to changes in the speed difference between lead and trail aircraft. There was no linear relation of throttle changes to changes in safe zone position or lead aircraft relative position.

However, in many cases there was an inverse linear relation of throttle changes to speed difference between the lead and trail aircraft. Table 5 shows the probabilities that this relation does not exist for each subject, with the highlighted cells being significant to $\alpha<0.10$. These results are shown in five conditions: the non-compliance scenarios; the speed non-compliance scenarios separated by behavior before and after the lead aircraft's change in speed; and the lateral non-compliance scenarios separated by behavior before and after the lead aircraft's change in lateral direction. Four pilots were found to have throttle control behavior correlated with speed differences in all conditions; all other pilots except Pilot 9 were found to have this correlation in behavior in at least two conditions. These results suggest that many of the pilots were using speed differences between themselves and the lead aircraft as a primary determinant of their throttle movements, except, in some cases, during lateral noncompliance from the lead aircraft.

Pilots successfully identified 79% of all noncompliance cases. Pilots did not detect 4% (or a total of 4) of the noncompliance conditions and misidentified 18%. Misidentifications included cases where:

- no noncompliance occurred but pilots indicated it had occurred,
- speed noncompliance occurred and the pilot indicated either no noncompliance or lateral noncompliance had occurred, or
- lateral noncompliance occurred and the pilot indicated either no noncompliance or speed noncompliance had occurred.

An ANOVA found no significant differences for the missed detections and misidentifications by display or noncompliance type.

Discussion

The pilots did not appear to follow the anticipated control or monitoring strategies. The pilots did not maintain a particular position within whichever safe zone was displayed, nor were the throttle movements consistent with deviations of position with respect to the safe zone. Instead, these throttle movements corresponded with a difference in speed between the lead and trail aircraft, suggesting that the pilots were attempting to match the lead aircraft's speed, regardless of which safe zone was displayed.

The pilots also did not appear to follow the expected monitoring behavior. Pilots made frequent mistakes about whether noncompliance had occurred, and of what type it was. Moreover, noncompliance detection performance was the same regardless of display type, even though monitoring needs to increase when given only the predicted safe zone. Pilots did, however, appear to be monitoring the speed of the other aircraft quite closely, making frequent mention of when the text display of the lead aircraft's speed was obscured by other symbology, regardless of display type. This latter finding further supports the idea that the pilots' control strategy was to null differences in speed between themselves and the lead aircraft.

In addition, pilots' monitoring of the safe zone appeared to be similar to the monitoring of a "red line" on an engine instrument. Pilots did not want to exceed this limit, but otherwise made little attempt to track it. When the pilots exceeded the safe zone, they performed a missed approach.

Pilot reaction to noncompliance of the lead aircraft also was partly unanticipated. Although pilots did execute a missed approach as specified by the procedure if they departed the safe zone, they did not react properly to lead aircraft noncompliance when they were given the predicted safe zone. This suggests that they did not (or could not) interpret the consequences of the lead aircraft noncompliance.

Conclusions

The pilots in this study did not follow the position-keeping strategy that is implicitly expected in studies of paired approaches. In fact, they appeared to make little attempt to maintain a static position within the safe zone. Instead, they appeared to favor a strategy of matching speed with the lead aircraft. This strategy would keep them in a static position with respect to the safe zone only if the safe zone were static with respect to the lead aircraft. However, this is not the case for the actual safe zone, which is updated using real-time information, as the position of the safe zone behind the lead aircraft changes as speeds and lateral separation change.

The design of procedures and displays for paired approaches has not considered the possibility of this strategy. If this behavior is to be supported, then future analysis of the system must incorporate it into the modeling of the pilot and into the design of procedures and displays. For example, in addition to providing a text indication of lead aircraft speed, presenting relative speed and/or command information to the trail aircraft may provide direct support to matching the lead aircraft's speed. Similarly, procedures could be adapted to allow the trail aircraft to match the lead aircraft's speed.

This type of support may also require that the information be presented on the primary flight display, rather than on the ND as used in this experiment. The displays used in the experiment were based on prototypes (Bone, 2001; Pritchett, 2000). These displays reflected the assumption that the pilot of the trail aircraft would monitor the position and behavior of the lead aircraft, behavior that is best supported by a map-like depiction of the lead aircraft, such as provided by the navigation display and/or traffic situation display. However, pilots do not typically monitor the navigation display on final approach. Support for the pilots' apparent strategy (matching the lead aircraft's speed) may be better integrated with the other information used and tasks performed by the pilot during the approach, which are centered on the Primary Flight Display (PFD) rather than ND.

If the pilots' strategy is instead deemed to have inadequate performance for this operation, then, in addition to training designed to adapt the pilot's behavior, changes to displays and procedures would be required. The changes may include displaying a speed cue (e.g. a range of speeds on the speed tape of the PFD) that is calculated to keep the trail aircraft within the safe zone for a specified period of time. Alternatively, a "desired position" cue could be added to the display to indicate to the pilot where in the safe zone he or she should be.

In a general sense it is difficult to know a priori what strategies operators will bring to a novel operation. Assumptions about these strategies are often adopted from similar systems, and may be incomplete or inaccurate. These assumptions often have significant implications for how procedures and technologies are designed, and mismatches between these and operator strategy can cause poor overall performance.

Careful up-front analysis and ecological experimentation can catch these assumptions before too much time and effort are expended on a poor design.

References

Bone, R., Olmos, O., and Mundra, A., 2001. Paired approach: a closely-spaced parallel runway approach concept. The International Symposium on Aviation Psychology, Columbus OH.

Hammer, J., 1999. Case study of paired approach procedure to closely spaced parallel runways. Air Traffic Control Quarterly, Vol. 8(3), 223-252.

Ippolito, C. and Pritchett, A., 2000. Software architecture for a reconfigurable flight simulator. AIAA Modeling and Simulation Technologies Conference: Denver, CO.

Ockerman, J. and Pritchett, A. (in press). A review and reappraisal of task guidance: Aiding workers in procedure following. International Journal of Cognitive Ergonomics.

Pritchett, A. and Landry, S., 2001. Two studies of paired approaches. Presented at The 2001 Air Traffic Management Research and Design Workshop: Sante Fe, NM.

Pritchett, A., 1999. Pilot performance at collision avoidance during closely spaced parallel approaches. Air Traffic Control Quarterly, Vol. 7(1), 47-75.

Pritchett, A., 2000. Display effects on shaping apparent strategy: A case study in collision detection and avoidance, International Journal of Aviation Psychology, Vol. 10(1), 59-83.

Stone, R., 1996. Paired approach concept. In Waller, M.C. and C.H. Scanlon (Eds.) Proceedings of the NASA Workshop on Flight Deck Centered Parallel Runway Approaches in Instrument Meteorological Conditions: NASA Langley Research Center.

Predicting Pilot Error: Assessing the Performance of SHERPA

Neville A. Stanton, Mark S. Young, Paul Salmon (1), Don Harris (2), Jason Demagalski (2), Andrew Marshall (3), Thomas Waldman (4), Sidney Dekker (5).

(1) Department of Design, Brunel University, Egham, Surrey, TW20 OJZ, UK, +44 (0)1784 431341
(2) Cranfield University, UK
(3) Marshall Associates, London, UK
(4) University of Limerick, Ireland
(5) Linkoping University, Sweden

Abstract: This paper introduces SHERPA (Systematic Human Error Reduction and Prediction Approach) as a means for predicting pilot error. SHERPA was initially developed for predicting human error in the nuclear industry about 15 years ago. Since that time validation studies to support the continued use of SHERPA have been encouraging. Most research shows that SHERPA is amongst the best human error prediction tools available. Yet there is little research in the open literature of error prediction for cockpit tasks. This study attempts to provide some evidence for the reliability and validity of SHERPA in aviation domain.

Keywords: SHERPA, errors, reliability, validity

Introduction

Human error is an emotive topic. Psychologists and Ergonomists have been investigating the origins and causes of human error since the dawn of the discipline (Reason, 1990). Traditional approaches suggested that error was an individual phenomenon, the individual who appears responsible for the error. Indeed, so-called 'Freudian slips' were treated as the unwitting revelation of intention: an error revealed what a person was really thinking but did not wish to disclose. More recently, error research in the cognitive tradition has concentrated upon classifying errors within taxonomies and determining underlying psychological mechanisms (Senders & Moray, 1991). The taxonomic approaches by Norman (1988) and Reason (1990) have led to the classification of errors into different forms, e.g. capture errors, description errors, data driven errors, association activation errors and loss of activation errors. Reason (1990) and Wickens (1992) identify psychological mechanisms implied in error causation, for example the failure of memory retrieval mechanisms in lapses, poor perception and decision-making in mistakes and motor execution problems in slips. Taxonomies offer an explanation of what has happened, whereas consideration of psychological mechanisms offer an explanation of why it has happened. Reason (1990), in particular, has argued that we need to consider the activities of the individual if we are able to consider what may go wrong. This approach does not conceive of errors as unpredictable events, rather as wholly predictable based upon an analysis of an individual's activities. Since the late 1970's much effort has been put into the development of techniques to predict human error based upon the fortunes, and misfortunes, of the nuclear industry. Despite this development, many techniques are poorly documented and there is little in the way of validation studies in the published literature.

Validating Human Error Prediction

Whilst there are very few reports of validation studies on ergonomics methods in general (Stanton and Young, 1999a), the few validation studies that have been conducted on HEI are quite optimistic (e.g. Kirwan, 1992a, b; Stanton and Baber, 1996). It is encouraging that in recent years the number of validation studies has gradually increased. Empirical evidence of a method's worth should be one of the first requirements for acceptance of the approach by the ergonomics and human factors community. Stanton and Stevenage (1998) suggest that ergonomics should adopt similar criteria to the standards set by the psychometric community, i.e. research evidence of reliability and validity before the method is widely used. It may be that the ergonomics community is largely unaware of the lack of data (Stanton and Young, 1998) or assumes that the methods provide their own validity (Stanton and Young, 1999b).

The development of HEI techniques could benefit from the approaches used in establishing psychometric techniques as two recent reviews demonstrate (Bartram et al, 1992, Bartram et al, 1995). The methodological concerns may be applied to the entire field of ergonomics methods. There are a number of issues that need to be addressed in the analysis of human error identification techniques. Some of the judgments for these criteria developed by Kirwan (1992, b) could be deceptive justifications of a technique's effectiveness, as they may be based upon:

- User opinion
- Face validity
- Utilisation of the technique.

User opinion is suspect because of three main reasons. First it assumes that the user is a good judge of what makes an effective technique. Second, user opinion is based on previous experience, and unless there is a high degree of homogeneity of experience, opinions may vary widely. Third, judgments may be obtained from an unrepresentative sample. Both Kirwan (1992, b) and Baber & Stanton's (1996) studies used very small samples. Face validity is suspect because a HEI technique might not be able to predict errors just because it looks as though it might, which is certainly true in the domain of psychometrics (Cook, 1988). Finally, utilisation of one particular technique over another might be more to do with familiarity of the analyst than representing greater confidence in the predictive validity of the technique. Therefore more rigorous criteria need to be developed.

Shackel (1990) proposed a definition of *usability* comprising effectiveness (i.e. level of performance: in the case of HEI techniques this could be measured in terms of reliability and validity), learnability (i.e. the amount of training and time taken to achieve the defined level of effectiveness) and attitude (i.e. the associated costs and satisfaction). These criteria together with those from Kirwan (1992, b: i.e., comprehensiveness, accuracy, consistency, theoretical validity, usefulness and acceptability) and the field of psychometrics (Cronbach, 1984; Aiken, 1985) could be used to assess HEI techniques (and other ergonomics methods) in a systematic and quantifiable manner.

Systematic Human Error Reduction and Prediction Approach (SHERPA)

SHERPA (Embrey, 1986) uses Hierarchical Task Analysis (HTA: Annett *et al*. 1971) together with an error taxonomy to identify credible errors associated with a sequence of human activity. In essence the SHERPA technique works by indicating which error modes are credible for each task step in turn, based upon an analysis of work activity. This indication is based upon the judgement of the analyst, and requires input from a subject matters expert to be realistic.

The process begins with the analysis of work activities, using Hierarchical Task Analysis. HTA is based upon the notion that task performance can be expressed in terms of a hierarchy of goals (what the person is seeking to achieve), operations (the activities executed to achieve the goals) and plans (the sequence in which the operations are executed). Then each task step from the bottom level of the analysis is taken in turn. First each task step is classified into a type from the taxonomy, into one of the following types:

- Action (e.g. pressing a button, pulling a switch, opening a door)
- Retrieval (e.g. getting information from a screen or manual)
- Checking (e.g. conducting a procedural check)
- Selection (e.g. choosing one alternative over another)
- Information communication (e.g. talking to another party)

This classification of the task step then leads the analyst to consider credible error modes associated with that activity. From this classification the associated error modes are considered. For each credible error (i.e. those judged by a subject matter expert to be possible) a description of the form that the error would take is given. The consequence of the error on the system needs to be determined next, as this has implications for the criticality of the error. The last four steps consider the possibility for error recovery, the ordinal probability of the error, its criticality and potential remedies.

Studies of SHERPA

Kirwan (1992b) conducted a comparative study of six potential HEI techniques. For this study he developed eight criteria on which to compare the approaches. In his study, Kirwan recruited 15 HEI analysts (three per technique, excluding group discussion). Four genuine incidents from the nuclear industry were used as a problem to focus the analysts' effort. This is the main strength of the study, providing a high level of ecological or face validity. The aim of the was to see if the analysts could have predicted the incidents if the techniques had been used. All the analysts took less than two hours to complete the study. Kirwan presented the results for the performance of the techniques as both subjective judgments (i.e.: low, medium and high) and rankings (i.e. worst and best). No statistical analysis was reported in the study, this is likely to be due to methodological limitations of the study (i.e. the small number of participants employed in the study). From the available techniques, SHERPA achieved the highest overall rankings and Kirwan recommends a combination of expert judgement together with the SHERPA technique as the best approach.

A study by Baber & Stanton (1996) aimed to test the hypothesis that the SHERPA technique made valid predictions of human errors in a more rigorous manner. In order to do this, Baber & Stanton compared predictions made by an expert user of SHERPA with errors reported by an observer. The strength of this latter study over Kirwan's is that it reports the use of the method in detail as well as the error predictions made using SHERPA. Baber & Stanton's study focuses upon errors made during ticket purchasing on the London Underground, for which they sampled over 300 transactions during a non-continuous 24-hour period. Baber and Stanton argue that the sample was large enough as 90% of the error types were observed within 20 transactions and after 75 transactions no new error types were observed. From the study, SHERPA produced 12 error types associated with ticket purchase, nine of which were observed to occur. Baber & Stanton used a formula based upon Signal Detection Theory (Macmillan & Creelman, 1991) to determine the sensitivity of SHERPA in predicting errors. Their analysis indicated that SHERPA produces an acceptable level of validity when used by a expert analyst. There are, however, a two main criticisms that could be aimed at this study. First, the number of participants in the study was very low; in fact only two SHERPA analysts were used. Second, the analysts were experts in the use of the technique; no attempt was made to study performance whilst acquiring expertise in the use of the technique.

Stanton & Stevenage (1998) conducted two experimental studies to test the learnability of SHERPA with novice participants. In the first study, the error predictions of 36 participants were compared to those who had no formal error methodology, to see if people using SHERPA performed better than heuristic judgement. Similar to the Baber & Stanton (1996) study, these predictions were made on the task that required people to make a purchase from a vending machine. Participants using the SHERPA technique correctly predicted more errors and missed fewer errors than those using the heuristics. However, they also appeared to incorrectly predict more errors. There appears to be a trade-off in terms of training such that a more sensitive human error identification is achieved, at the cost of a greater number of false positives. This is probably a conservative estimate, as no doubt if the observation period was extended indefinitely, more error types would be observed eventually. In the second study, 25 participants applied SHERPA to the vending task on three separate occasions. The data reported by Stanton & Stevenage show that there is very little change over time in the frequency of hits and misses however, the frequency of false alarms appears to fall over time and consequently, the frequency of correct rejections appears to increase. In terms of the overall sensitivity of error prediction, this shows remarkable consistency over time.

Predicting Pilot Errors in the Autopilot Task Using SHERPA

The purpose of this study was to evaluate the SHERPA methodology applied to the analysis of the flight deck for the autoland task. There are many limitations on this study. For starters, there is no attempt to evaluate the dialogue between the pilot, the co-pilot and air traffic control. It is already assumed that autoland will be used. There are also limitations with regard to the collection of error data from pilots, which largely relied upon self-report to a questionnaire survey. Nevertheless, within these limitations, some insight into the success with which SHERPA can be applied to an aviation domain can be gleaned

Eight graduate engineering participants aged between 22 and 55 years took part in this study. All participants were trained in the SHERPA methodology. The training comprised an introduction to the key stages in the method and a demonstration of the approach using a non-aviation example, using an in-car task from Stanton & Young (1999a). Participants were then required to apply the method to another non-aviation task with guidance from the instructors from a public technology task from Stanton & Stevenage (1998). The purpose of this was to ensure that they had understood the workings of the SHERPA method. A debriefing followed, were participants could share their understanding with each other. When the instructors were satisfied that the training was completed, the main experimental task was introduced. This required participants to make predictions of the errors that pilots could make in the autoland task.

To make their error predictions, participants were given a HTA of the autoland task developed by the authors (comprising some 22 subtasks under the main headings: setting up for approach, lining up for the runway, and preparing the aircraft for landing), a demonstration of autoland via Microsoft flight simulator, the SHERPA error taxonomy, and colour photographs of: the autopilot panel; levers for flaps, landing gear and speed brake; the primary flight displays; and an overview of the cockpit.

Participants were required to make predictions of the pilot errors on two separate occasions, separated by a period of four weeks. This enabled intra-analyst reliability statistics to be computed. The predictions were compared with error data reported by pilots using autoland. This enabled validity statistics to be computed. The signal detection paradigm provides a useful framework for testing the power of HEI techniques. In particular, it identifies type I errors (a miss: when the error analyst predicts the error will not occur and it does) and type II errors (a false alarm: when the error analyst predicts that there will be an error and there is not) in the judgement of the analyst.

Analysis of the data revealed the mean reliability of analysts between time one and time two using SHERPA as approximately 0.7 and mean validity, expressed as an mean of the hit and false alarm rates as approximately 0.6. These values are moderate, but it should be noted that this was the first time the participants had applied the SHERPA method in anger and that they were not aviation experts. The pooled error predictions are compared to the errors reported by pilots in table one. If the error predictions are pooled, the validity statistic rises to approximately 0.9 which is very good indeed.

Table 1. Pooled error data

		Errors Observed	
		Yes	No
Errors Predicted	Yes	Hits = 52	F. A. = 4
	No	Misses = 5	C. R. = 179

Conclusions

In conclusion, the results are promising for the use of SHERPA in predicting pilot error. Whilst more studies are needed to investigate different tasks, the current study shows that novices were able to acquire the approaches with relative ease and reach acceptable levels of performance within a reasonable amount of time. This supports the investigation by Stanton & Stevenage and is quite encouraging. The study also shows that HEI techniques can be evaluated quantitatively.

Human error is a complex phenomenon, and is certainly far from being completely understood. Yet in attempting to predict the forms in which these complex behaviours will manifest themselves armed only with a classification systems and a description of the human and machine activities it is amazing what can be achieved. Despite the gaps in our knowledge and the simplicity of the techniques, the performance of the analysts appears surprisingly good. This offers an optimistic view of the future for human error identification techniques. There are a number of other criticisms that need to be addressed, however. Stanton and Stevenage (1998) propose that clearer documentation on the methodologies needs to be provided, and that cross validation studies should be undertaken.

Acknowledgement

This research is supported by a grant from the Department of Trade and Industry as part of the European EUREKA! research programme.

References

Aitkin, L. R. (1985) Psychological Testing and Assessment. Allyn & Bacon: Boston.

Annett, J.; Duncan, K. D.; Stammers, R. B. & Gray, M. J. (1971) Task Analysis. Training Information No. 6. HMSO: London.

Baber, C. & Stanton, N. A. (1996) Human error identification techniques applied to public technology: predictions compared with observed use. Applied Ergonomics. 27 (2) 119-131.

Bartram, D.; Lindley, P.; Foster, J. & Marshall, L. (1992) Review of Psychometric Tests (Level A) for Assessment in Vocational Training. BPS Books: Leicester.
Bartram, D.; Anderson, N.; Kellett, D.; Lindley, P. & Robertson, I. (1995) Review of Personality Assessment Instruments (Level B) for use in Occupational Settings. BPS Books: Leicester.
Cook, M. (1988) Personnel Selection and Productivity. Wiley: Chichester.
Cronbach, L. J. (1984) Essentials of Psychological Testing. Harper & Row: New York.
Embrey, D. E. (1986) SHERPA: A systematic human error reduction and prediction approach. Paper presented at the International Meeting on Advances in Nuclear Power Systems, Knoxville, Tennessee.
Kirwan, B. (1992a) Human error identification in human reliability assessment. Part 1: overview of approaches. Applied Ergonomics, 23 pp. 299-318.
Kirwan, B. (1992b) Human error identification in human reliability assessment. Part 2: detailed comparison of techniques. Applied Ergonomics, 23 pp. 371-381.
Macmillan, N. A. & Creelman, C. D. (1991) Detection Theory: a user's guide. Cambridge University Press: Cambridge.
Norman, D. A. (1988) The Psychology of Everyday Things. Basic Books: New York.
Reason, J. (1990) Human Error. Cambridge University Press: Cambridge.
Senders, J. W. & Moray, N. P. (1991) Human Error. LEA: Hillsdale, NJ.
Shackel, B. (1990) Human factors and usability. In: Preece, J. & Keller, L. (eds) Human-Computer Interaction. Prentice-Hall: Hemel Hempstead.
Stanton, N. A. & Baber, C. (1996) A systems approach to human error identification. Safety Science, 22, pp. 215-228.
Stanton, N. A. & Stevenage (1998) Learning to predict human error: issues of reliability, validity and acceptability. Ergonomics 41 (11), 1737-1756
Stanton, N. A. & Young, M. (1998) Is utility in the mind of the beholder? A review of ergonomics methods. Applied Ergonomics. 29 (1) 41-54
Stanton, N. A. & Young, M. (1999a) A Guide to Methodology in Ergonomics: Designing for Human Use. Taylor & Francis: London.
Stanton, N. A. & Young, M. (1999b) What price ergonomics? Nature 399, 197-198
Wickens, C. D. (1992) Engineering Psychology and Human Performance. Harper Collins: New York.

Coordination within work teams in high risk environment

Gudela Grote and Enikö Zala-Mezö

Swiss Federal Institute of Technology, Zurich (ETH) Institute of Work Psychology
Nelkenstr. 11 8092 Zürich, Switzerland
zala@ifap.bepr.ethz.ch

Abstract: In the following we outline a research project with the main question: how do standardization and work load influence coordination processes. In the long run we focus on two high risk systems, aviation and medicine, where work processes in the former are much more standardized than in the latter. In the present article, however we will concentrate on the data from the field of aviation. We show some theoretical background from both traditional and more recent research perspectives and introduce our methodology, where we analyse the communication processes quantitatively based on observational categories related to information flow and leadership as well as more qualitatively using indicators for heedful interrelating (Weick and Roberts, 1993). The aim of this part of the study is to identify different types of coordination patterns under conditions of high versus low work load which also vary in the degree of standardisation.

Keywords: coordination, standardization, heedful interrelating, explicit – implicit coordination.

Introduction

Coordination defined as tuning of interdependent work processes to promote concerted action towards a superordinate goal (Kieser & Kubicek, 1989) is needed for any activity which cannot be carried out by one person and which cannot be subdivided into independent parts (Hackman & Morris, 1975). Coordination is therefore a core activity in any work organization. As Tesluk et. al (1997) formulate: "The essence of organizational action is the coordination, synchronization, and integration of the contributions of all organizational members into a series of collective responses." Crucial in this respect is the type of interdependence created by the chosen division of labour in combination with the general demands of the task and the task environment. Generally, three types of interdependence of work activities are distinguished (e.g. Tesluk et al., 1997; Thompson, 1967) according to the type and degree of interdependence: pooled, sequential, and reciprocal interdependence. Tesluk et al. (1997) in view of special demands created by task performance in high risk situations like flying an airplane or operating a patient, have added a forth form of interdependence called *intensive work situations* where team members work very closely together and work flow is poly-directional, flexible and very intensive, because the team repeatedly faces novel situations with new problems which have to diagnosed and solved within the team.
In the following, the focus will be on such intensive work situations and the specific requirements for coordination in these situations, especially discussing standardization, i.e. coordination via centrally determined and highly formalized programs and plans, as a widely used form of coordination in high risk work systems.

Standardized coordination in high risk organisations

In order to understand the reasons for and effects of standardized coordination in high risk organizations, it is helpful to conceptualise organizational activities in terms of the management of uncertainty. The kinds of uncertainty an organisation has to deal with and how these uncertainties are handled by the organisation has been a core issue in organisation theory. Prominent authors in this field like Thompson (1967), Perrow (1967) and Susman (1976) have helped to systematize the nature of uncertainties relevant to organisations and the ways organisations deal with them.
There are two extreme approaches to handling uncertainty. The first one tries to minimize uncertainty or at least the effects of uncertainty in the organization using mainly feedforward control based on high standardization and programming of work flows, providing minimal degrees of freedom to the people in charge of carrying out the plans.
The other approach having been advertised by organisation theorists and work scientists for several decades now is - instead of trying to minimise the uncertainties themselves - to enable each and every member of an organisation to handle uncertainties locally. From this perspective, planning is understood primarily as a resource for situated action (Suchman, 1987).

Standardisation can be regarded as the key element in the minimising uncertainty approach, while the competent coping with uncertainty relies much more on personal and lateral co-ordination mechanisms. There are some critical voices regarding the usefulness of high levels of standardization, mainly pointing to the system's reduced capability to adequately act in the face of requirements stemming from internal and external disturbances of normal operation (e.g. Amalberti, 1999; Perrow, 1984; Grote, 1997).
Following, some newer concepts will be presented that may help in shaping standardization in a way more conducive to safe operation in high-risk systems.

New directions for thinking about standardization
In most high risk systems, standardization in the form of standard operating procedures has been developed with ever increasing detail in order to streamline human action and to reduce its influence as a risk factor. While generally there is an understanding that rules are useful guides for safe behaviour, there is also an increasing concern that too many rules incrementally developed will not make up a good system to help human actors do the right thing especially in states of abnormal operation where they would need strong, but also flexible guidance (e.g. Amalberti, 1999).
Another basic problem with standardization is that especially in non-routine situations reliance on common standards may turn into an overreliance, impeding switches to more explicit coordination and with that switches to higher levels of common action regulation, i.e. switches from skill-based to rule-based or from rule-based to knowledge-based behaviour.[3]
Making a similar distinction between minimizing uncertainties vs. competently coping with uncertainties, as was suggested before in this article, Rasmussen has argued that "rather than striving to control behaviour by *fighting deviations* from a particular pre-planned path, the focus should be on the control of behaviour by *making the boundaries explicit and known* and by giving opportunities to develop *coping skills at boundaries*" (Rasmussen, 1997: 191; italics in the original).
In line with this approach to rules, some authors (e.g. Hale & Swuste, 1998; LePlat, 1998) have begun to develop typologies of rules in order to help the design of rule systems directly tailored to the needs for guidance as well as for autonomy and control arising in different stages of action regulation. From an action regulation perspective, rules can concern goals to be achieved, define the way in which decisions about a course of action must be arrived at, or prescribe concrete actions.
Hale and Swuste (1998) also suggest some criteria to help decide at which level of the organisation these rules should be defined: predictability of the system; innovation rate in the system; interaction requirements; local expertise. From an organisational perspective, rules should also be discussed as elements of the coordination mechanisms operating within and between parts of an organization.
During the last decade, coordination in high-risk environments has been addressed in an increasing number of studies. Usually, coordination on team level has been analysed with no explicit reference to organisational coordination mechanisms and the types of rules the teams have to adhere to, however.
The vast majority of the studies have been carried out in aviation settings, taking for granted a high level of standardization. Following, the evidence on coordination requirements for successful performance provided by these studies will be reviewed.

Studies on coordination in work teams in high-risk environments
Given the definition of work teams as "...two or more people with different tasks who work together adaptively to achieve specified and shared goals" (Brannick & Prince,1997), coordination is one of the team's main activities.
A core concept in many of the studies on team coordination is the distinction between explicit and implicit coordination in relation to coping with high levels of workload. Explicit coordination is considered necessary when an agreement must be arrived at about how an action should be organised. It occurs typically during new tasks and new situations or when a new group of people make up a team to accomplish a job. People have to devote extra resources (very often communication) to organize the activities. Implicit coordination occurs when every one in a team knows his/her job, the actions harmonise

[3] It is to be noted that rule-based behaviour refers to a special kind of action regulation effort, i.e. the selection of behaviour based on choices between fairly prescribed alternative courses of action. Rules in the meaning used above as standard operating procedures can be related to this level of action regulation, but also to the other two levels, depending on the type of rule (LePlat, 1998).

with each other based on some kind of shared understanding (Cannon-Bowers & Salas, 2001), and therefore little noticeable effort for coordination is required.
Another theory which could give us a new perspective is from Weick and Roberts (1993). They have provided case-study based and more qualitative accounts of similar phenomena of more or less effective team coordination in their analyses of high-reliability organizations mentioned previously. In order to explain effective team coordination, they suggest the concept of "heedful interrelating". A core idea of this concept based on Asch's theory on group interaction is that safety operations in highly complex situations require deliberate efforts by all actors to constantly (re-)consider effects of their own actions in relation to the goals and actions of others, or in Weick and Roberts' words: "... (to) construct their actions (contribute) while envisaging a social system of joint actions (represent), and interrelate that constructed action with the system that is envisaged (subordinate)" (Weick & Roberts, 1993: 363; see also Table 1 for tentative indicators of heedful/heedless interrelating).

Indicators for	
Heedful interrelating	Heedless interrelating
Detailed representation of others	Less detailed representation of others
Contributions shaped by anticipated responses	Contributions shaped less by anticipated responses
Broad boundaries of envisaged system	Narrow boundaries of envisaged system
Attention focus on joint situation	Attention focus on local situation
Good comprehension of the implications of unfolding events	Little comprehension of the implications of unfolding events

Table 1- Tentative indicators for heedful vs. heedless interrelating
(adapted from Weick and Roberts, 1993)

As was stated already, research on team coordination in high-risk environments usually has not explicitly addressed which organizational coordination mechanisms (which level of standardisation) provide the framework for the observed team behaviours. A more theoretically guided approach to what coordination is and how different kinds of communication can contribute to fulfilling different demands on coordination is needed in order to develop more systematic indicators of coordinated action.
A more qualitative and systematic approach to team coordination seems also warranted because situational demands can vary drastically within the generally used classification of high vs. low workload, potentially requiring very different communication and coordination strategies.
In the following, we will use the term task load, when we describe objective difficulties connected to the properties of a task and we will use the term workload, when referring to how a situation is perceived by the people facing the task.

Standardization and coordinated action: Study design

As a starting point for developing a more systematic and theory-guided account of team coordination in high-risk and high workload situations, Weick and Roberts's (1993) concept of heedful interrelating appears to be most promising because of its roots both in systemic organization theory and social psychology. Unfortunately, up to now, it has remained sketchy and no attempts have been made to derive measurable indicators for coordinated action from it.
Team coordination as measured by indicators for heedful interrelating should be studied in task environments with different degrees and types of standardization as evidenced by different sets of rules laid out by the organization, using Hale and Swuste's (1998) classification scheme of safety rules as an operational framework.
In order to do this, we have chosen the comparison between cockpit crew coordination and coordination in the emergency room.

Keeping in mind all important differences characterising these groups, analysing advantages and disadvantages of different degrees of standardization on team coordination and team performance in these settings should be very beneficial from a theoretical and also from a very practical point of view.
To summarize, in our study we are asking the following four questions:

- Are there differences in patterns of coordination behaviours between teams in cockpits and emergency rooms as a result of differences in degrees of standardization?
- Can these differences be linked to team performance under varying degrees of workload?
- Can these differences be described in terms of explicitness vs. implicitness of coordination and in terms of heedful vs. heedless interrelating?
- Based on the answers to these three questions, we hope to also find first answers to a forth question:
- Which types of rules in what combination, and derived from that, which specific forms of standardization support successful coordination?

In Figure 1 the overall design of the study is presented. The effects of varying workload and varying degrees and types of standardization on coordinating behaviours and indirectly on team performance are to be analysed in the two settings cockpit and emergency room.

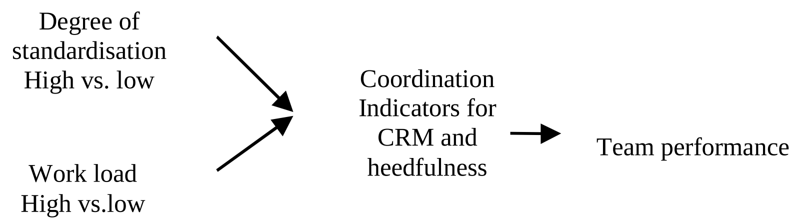


Figure 1 - Study design
(ER= Emergency room; CRM = Crew Resource Management)

Data

While the data from the medical setting are still in the process of being collected, flight data (video tapes from simulator sessions) are available already from 80 simulator training sessions, which were taped as part of another project (cf. Naef, Klampfer & Häusler, 2001) within the umbrella project "Group Interaction in High Risk Environments" to which our study belongs as well.
We analyse one scenario, during which an approach and landing has to be performed without flaps and slats. This so-called clean approach entails high landing speed, unusual visual information due to unusual attitude of airplane and the requirement of very good manipulative skills by the pilot flying.
Workload is operationalized by means of the NASA Task Load Index (Hart & Staveland, 1988). An external expert also rates task load for the overall situation based on the NASA-Index.
Standardisation was broadly operationalized in terms of the two setting studied, i.e. low standardization in the emergency room and high standardization in the cockpit. A more fine-grained analysis of the types of rules relevant in the two settings will be performed by means of document analysis and expert interviews based on the categories developed by Hale and Swuste (1998).
Team performance is rated by the team members themselves and by an external expert (same as for external workload rating) according to technical and social performance.
Coordinating behaviours are analysed based on observational categories. Videotapes of cockpit simulator training sessions and emergency room operations are rated based on these categories, using the ATLAS-ti (Muhr, 1997) program. These categories as can be seen below are based on relatively broad characteristics of communication. Non-verbal actions are not included in the categories. This method allow us to obtain a general impression about the whole co-ordination process.

Observational categories

We developed four main groups of categories driven by both theoretical and practical considerations. We used two behavioural marker systems for the evaluation of crew resource management, LOSA (Line Oriented Safety Audit, Helmreich et al., 1999) and NOTECHS (NOn-TECHnical Skill proficiency, Avermate & Kruijsen, 1998), as references. While LOSA and NOTECHS have been developed to obtain overall ratings of individual and/or team performance in cockpit crews on a number of quite general characteristics (leadership, decision making, planning, situation awareness etc.), our categories are intended to allow coding of all utterances in the two settings cockpit and emergency room. Therefore we could use the LOSA/NOTECHS-categories as an orientation regarding relevant areas of communication in the cockpit, but had to develop more specific categories within each topic and also develop categories that are applicable to both the flight and medical situation.

The first set of categories concerns the information flow on a general level without much reference to the content of the information. The aim was to create mutually exclusive categories and to be able to code all utterances made during the observed situation. Also, it was attempted to differentiate elements of explicit and implicit coordination as described in the previous sections of this article.

The category type *Information flow - explicit coordination* contains the following categories:

- Provide information
- Request information
- Provide information upon request
- Information containing a summary of a state of affairs or a process
- Reassurance (e.g. feedback about comprehension of a communication)
- Giving order
- Asking for help
- Communication with Air Traffic Control (specific for aviation)
- Discussion
- Standard communication (specific for aviation)

The category type *Information flow - implicit coordination* contains the following categories:

- Provide unsolicited information
- Offer assistance
- Silence
- Chatting

In the case of silence and chatting it is important to look at the whole situation to decide whether these categories indicate implicit coordination or absence of coordination. An important point regarding the other two categories is the anticipation effect, namely that a team member realizes the needs of other team members and provides the needed information or support without being explicitly requested to do so.

The information flow categories also provide a general quantitative account of the observed communication, concerning e.g. speaker dominance and proportion of standard versus non-standard information exchange.

The other two groups of categories are not fully exclusive and do not cover all utterances made. The second group of categories is connected to leadership, which was chosen as a focus due to the strong relationship between type of leadership and coordination and the effects of standardization on this relationship. In general, standards can be regarded as a form of depersonalised leadership, with personal leadership being made redundant in some respects and obtaining a different function as complementing or overriding standards.

The category type *leadership* contains the following categories:

- Making plans
- Assigning task
- Giving order
- Making decision
- Initiate an action
- Accepting decision or initiated action
- Questioning decision
- Ignoring initiated action
- Autocratic behaviour

The third group of categories contains elements of heedful interrelating:

- Considering others
- Considering the future
- Considering external conditions
- Initiate an action
- Questioning decision
- Providing unsolicited information
- Offering assistance
- Correcting the behaviour of others
- Teaching others
- Giving feedback about performance

These categories for heedful interrelating have then to be integrated into broader qualitative evaluations of the indicators listed in Table 1. In this process it has to be considered that some of the indicators are more directly linked to observable behaviours, while others are cognitive states which can only be inferred via the other indicators.

Results

The following results are based on 20 video records, which were coded according to the observational categories described before.

The difference in communication frequency performed by the captain and by the first officer: The captain communicates more frequently (1539) than the first officer (1155). The result is highly significant for the whole data set according to the binomial test. ($p = 0.000$) On the one side it is not very surprising since the captain is the dominant person who has to take responsibility but on the other side according to research outcomes on self-managing teamwork (Druskat, Prescosolido, 2002) it is central for those teams that team members develop a psychological ownership which means that both persons should remain active during the flight and the communication is certainly part of this activity. One could suppose that a close to equally distributed communication is a sign of good team work.

We looked as well which variables show the captain dominated communication and found the following categories: assigning task; giving order; making decision and plan; summarizing, reassurance and teaching. In one category, namely: providing unsolicited information, it is the first officer showing this behaviour more frequently. This is an important observational category for implicit coordination, where the anticipation of the needs of the team mates is a decisive point. (All those differences between first officer and captain are significant according to the binomial test.)

Flight sequences: In the following we will present some results based on different parts of the flight during the simulator session. The mean duration of a training is 23 minutes. The exercise can be divided into three flight sequences. The first one is the take off, which is highly standardized and nothing unexpected happens. It certainly requires a lot of attention but it is something which is regularly carried out therefore is the task load rather low. (Please note, that we do not refer to the subjectively felt work load here, which can vary among individuals, but to the task conditions which determine the grade of difficulty.) This part of the scenario takes approximately 3-4 minutes.

In the second part the failure is detected and the pilots have to go through a manual describing the special landing. The specialty of this sequence is that they have to find a solution for a problem, i.e. they have to check under which circumstances the landing is manageable. There are only a few rules about how this process should be done. This is the longest sequence the mean duration is approximately 9 minutes. Since there is no time pressure and not a real uncertainty about the solution we suppose that the task load is rather low. The third sequence is a very unusual, quite difficult landing, for that reason we suppose high task load. (Our assumptions about task load are supported by experts.) The process here is highly standardized.

	Take off	Problem solving	Landing
Duration (mean)	3- 4 minutes	9 minutes	3- 4 minutes
Standardization	High	Low	High
Task load	Low	Low	High

Table 3 – Overview of the three flight sequences during the training session

As we can see that setting provides a very nice opportunity to investigate our research questions within a work team and we can make some statements about the effect of task load and standardization although the data from the medical setting are not yet available.

Standard communication during the flight sequences: We coded standard communication when the pilots used prescribed wordings to communicate with each other. Complementing the detailed description of the rules based on content analysis which we plan to undertake in the next research step, this can also be taken as a sign to define how standardized the flight sequences are. Table 2 shows us how many of the coded interactions are standardised and non standardised during the 3 different flight sequences.

	Take off		Problem solving		Landing	
Interaction	Frequency	Std. residual	Frequency	Std. residual	Frequency	Std. residual
Standardised	268	10.1	163	-9.3	227	3.5
Non standard.	335	-5.7	1214	5.2	518	-2.0

Table 2 – Standard communication

This significant result (SPSS: Crosstabulation; Chi-Square Test: p = 0.000) supports our assumption about the flight sequences, where the 1st and the 3rd are highly standardized but not the 2nd. (We have to keep in mind that the 2nd sequence is a longer sequence than the other two. We tested this question with the cross tabulation which considers those differences, see the standard residuals).
Communication-Silence proportion: In table 4 is presented how often (frequency) and how long (mean duration in seconds) both pilots communicate and how often (frequency) and how long they stay silent.

	Take off			Problem solving			Landing		
	Freq.	Std. Residual	Duration seconds	Freq	Std. Resid.	Duration seconds	Freq	Std. Resid.	Duration seconds
Verbal communication	430	73 -1.6	3.6	1130	47 2.1	5.7	543	02 -1.3	3.84
Silence	173	3.0	8.64	247	-3.8	7.92	202	2.4	9.12

Table 4 - Verbal communication – silence proportion during the different flight sequences

What we can observe here is that the different tasks the pilots have to carry out determine the verbal communication quite strongly. They interact more frequently than they keep quiet, but the duration of the interaction is shorter, than the duration of silence. The differences in frequency between the flight sequences were tested with the Pearson Chi-Square test (SPSS Crosstabulation) which was very significant (p = 0.000). The differences in duration between communication and silence are also significant for all three sequences. (One Sample T-test, p=0.000 for all three sequences)
The duration differences between the flight sequences are significant in the case of verbal communication (One-Way ANOVA, p=0.000) but not significant in the case of silence (One-Way ANOVA, p=0.326).

Explicit - Implicit coordination: We analysed first the differences according to the frequencies

	Take off		Problem solving		Landing	
	Frequency	Standard Residual	Frequency	Std. residual	Frequency	Standard Residual
Explicit	421	-0.8	1105 1105	3.3 3.3	454	-3.8
Implicit	182	1.3	272	-5.4	291	6.1

Table 5 – Explicit and implicit coordination

The problem solving task requires the most explicit coordination and during landing there is more explicit coordination than during take off, although as we have seen before they are very short interactions during landing. (SPSS Crosstabulation; Chi-Square Test: p= 0.000)
The highest frequency of implicit coordination is in the 3rd sequence, which is the high task load phase during the simulator session. As we know from the work of Orasanu (1993) this kind of coordination plays a very important role during high work load situations where the cognitive capacity of a pilot can be absorbed by different information which has to be processed in parallel. The best way not to overwhelm the team mates with even more information is to anticipate what they really need and provide only this.
We were curious to see how this analysis looks if we focus on the duration of those two coordination forms. We created an index for every team and every flight sequence, where the summarized duration of explicit coordination was divided by the summarized duration of implicit coordination. If this value is bigger than 1 it means that more explicit than implicit coordination was performed during that flight sequence. This index shows a highly significant difference between the flight sequences.

	Take off	Problem solving	Landing
Mean / Std. Dev	1.01 / 0.86	3.63 / 2.1	0.79 / 0.47
(T test) p	0.000	0.000	0.000

Table 7 – Explicit-Implicit coordination index

During the 1st sequence the index is near to 1, which means that more or less the same amount of time was spent for those coordination types. In the second phase the value of this index is strongly increased, we can state that the dominant coordination strategy is explicit coordination. In the 3rd phase it looks again different and the implicit coordination is going to be a dominant form of coordination.
Leadership: Very few observations falling into this category during the 1st and 3rd flight sequences, as we can see in table 8. There is no need for this kind of coordination, since decisions and the work division are made in the 2nd sequence. Generally speaking we can state that such a low number of observation in this category is a sign for a shallow hierarchy within those teams. (SPSS Crosstab.; Chi-Square Test p= 0.000)

	Take off		Problem solving		Landing	
Interaction	Frequency	Std. residual	Frequency	Std. residual	Frequency	Std. residual
Leadership	17	-3.5	139	5.2	21	-3.9
Non leadersh.	586	0.9	1238	-1.4	724	1.0

Table 8 - Leadership

Heedful interrelating: This kind of coordination could be important in the second phase, where the crew has to agree upon the steps to undertake during the landing and in the 3rd phase as well, where the support from the pilot non flying is essential. In the 1st phase they have to execute a well trained process (take off), which is highly standardized, where the heedfulness can be replaced by the frequently employed standards. Our data supports these assumptions. (SPSS Crosstabulation; Chi-Square Test: p= 0.000)

	Take off		Problem solving		Landing	
Interaction	Frequency	Std. residual	Frequency	Std. residual	Frequency	Std. residual
Heedful	22	-6.0	183	1.3	125	3.7
Non heedful	73	2.2	1194	-0.5	620	-1.4

Table 9 – Heedful interrelating

Description of rules

We analysed the “Flight Procedures” chapter of the “General Basics; Flight Crew” (of a commercial airline) and the relevant parts of the “Aircraft Operations Manual” specific for Airbus 319/320/321. The categories we used to classify the rules are the following: The **content** of the rule can prescribe a goal, a process or a concrete action (Hale, Swuste,1998). The **strictness** of a rule is ordered into two categories: order versus advice. In some rules **exceptions** are mentioned and some others are without exceptions. Some rules include a reasoning why someone should follow it. Some rules have more **scope** than others, allowing to perform more autonomously. It is also distinguished whether the rule holds for **normal, abnormal** or **emergency** situation and who is the **person** addressed by the rule. The results are summarized in table 10:

	General Basics	AOM
Content: Goal / Process/ Action	17 -112 - 368	0 – 13 - 118
Strictness: Advice / Order	45-454	23 – 112
With exception / Without exception	33 - 464	9 - 126
Rule with / without reason	180 – 318	26 – 109
With scope / Without scope	285 - 215	54 - 81
Normal situation / all / emergency	12-192-294	1 – 87 - 47
First officer / Pilot in Command	1 - 2	2 - 8
Pilot flying / Pilot non flying	74 - 6	19 - 13
Both / None	54 - 362	17 - 76

Table 10 – Summary of the rules according to the content analysis

What we can see at first sight is that there are many rules mainly describing concrete actions and quiet some in the General Basics describing processes.

A major part of the rules is strict, they are commands not allowing any free choice for the actor.

Exceptions are mentioned rarely, but the rules are occasionally presented with the reasoning and they leave frequently some scope for the actor. Most of the rules are for emergency situations. The person addressed mostly is the pilot flying or even more frequently none of the pilots. This is a very general picture about the rules and the next research task is to specify the rules relevant for the flight sequences and to define their effects on the coordination processes. And of course this analysis will be a basis for the comparison of standardization with the medical field.

Conclusion

In this article a part of an ongoing research project has been outlined. Regarding the theoretical benefits of the research, its main contribution is seen in filling the void concerning detailed knowledge on the effects of different forms of standardization on team behaviour.

The most important statement we can derive from the present analysis is that standardization and task load effect the coordination processes. In highly standardized situation where nothing unexpected happens the frequency of explicit and implicit coordination is low and almost no observations occur in the category of

leadership and heedful interrelating. If the standardization is not so strong or the task load is increasing there is clear change in the coordination pattern. Low standardisation increases the amount of explicit coordination and heedful interrelating. The high task load increases the amount of implicit coordination and heedfulness. The main question is still to answer: which degree of standardization based on which types of rules can support teams in developing and maintaining flexible and situation-adaptive patterns of coordination? The answer hopefully will follow from the completed analysis in both professional fields.
On the most general level, it is hoped that the research outlined here will support a shift from minimizing uncertainties to competently handling uncertainties even in high-risk organizations, achieving a more balanced approach which will avoid both overreliance on rules as well as inappropriate deviation from rules.

Acknowledgements
We gratefully acknowledge the financial support for this project by the Daimler-Benz-Stiftung as well as the stimulating discussions in the umbrella project "Group interaction in high risk environments (GIHRE)" to which our project belongs. Also, we are very grateful for the enormous support by the GIHRE aviation project through making their raw data available to us.

References

Amalberti, R. (1999). Risk management by regulation. Paper presented at the 19th Myron B. Laver International Postgraduate Course "Risk Management", Dept. of Anaesthesia, University of Basel, Switzerland, March 26-27, 1999.

Avermate van, J. A. G. and E. A. C. Kruijsen, Eds. (1998). NOTECHS The evaluation of non-technical skills of multipilot aircrew in relation to the JAR-FCL requirements

Brannick, M., T. and C. Prince (1997). An Overview of Team Performance Measurement. Team Performance Assesment and Measurement. Theory, Methods and Applications. M. Brannick, T., E. Salas and C. Prince. Mahwah, New Jersey, London: 3-16.

Cannon-Bowers, J. A. and E. Salas (2001). "Reflections on shared cognition." Journal of Organizational Psychology 22: 195-202.

Druskat V.A., A.T. Pescosolido (2002). The content of effective teamwork mental models in self-managing teams: Ownership, learning and heedful interrelating. Human Relation 55(3): 283-314.

Grote, G. (1997) Autonomie und Kontrolle - Zur Gestaltung automatisierter und risikoreicher Systeme (Automomy and Control - On the Design of Automated and High-Risk Systems). Zürich: vdf Hochschulverlag.

Hackman R., J. and G. Morris; C. (1975). Group tasks, group interaction process, and and group performance effectiveness: A review and proposed integration. Advances in experimental social psychology. L. Berkowitz. New York, Academic Press. 8: 45-99.

Hale, A.R., and Swuste, P. (1998). Safety rules: procedural freedom or action constraint? Safety Science, 29: 163-177.

Hart, S. G. and L. E. Staveland (1988). Development of NASA-TLX (Task Load IndeX): Results of empirical and theoretical research. Human Mental Workload. P. A. Hancock and N. Meshkati. Amsterdam, The Netherlands: Elsevier.

Helmreich, R. L., Wilhelm, J. A., Klinect, J.R., and Merritt, A.C. (1999). Culture, Error and Crew Resource Management. In E. Salas, C. A. Bowers, and E. Edens (eds), Applying resource management in organizations: A guide for professionals. Hillsdale, NJ: Erlbaum.

Kanki, B., G. and M. Palmer, T. (1993). Communication and crew resource management. Cockpit Resource Management. E. L. Wiener, B. Kanki and R. L. Helmreich. San Diego, Academic Press.

Kieser, A., and Kubicek, H. (1992) Organisation. Berlin: de Gruyter.

Leplat, J. (1998) About implementation of safety rules. Safety Science, 29: 189-204.

Muhr, Thomas. Scientific Software Development, Berlin.

Naef, W., Klampfer, B., and Häusler, R. (2001). Aviation. In: Group Interaction in High Risk Environments (GIHRE), Project extension proposal submitted to the Gottlieb Daimler- and Kalr Benz-Foundation, Ladenburg.

Orasanu, J. M. (1993). Decision-making in the Cockpit. Cockpit Resource Management. E. L. Wiener, B. Kanki and R. L. Helmreich. San Diego, Academic Press: 137-172.

Perrow, C. (1967) A framework for the comparative analysis of organizations. American Sociological Review, 32: 194-208.

Perrow, C. (1984) Normal accidents. Living with high-risk technologies. New York: Basic Books.

Rasmussen, J.(1997) Risk management in a dynamic society: A modelling problem. Safety Science, 27: 183-213.

Suchman, L.A. (1987). Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge: Cambridge University Press.

Susman, G.I. (1976). Autonomy at Work. A Sociotechnical Analysis of Participative Management. New York: Praeger.

Tesluk, P., J. Mathieu, E., et al. (1997). Task and Aggregation Issues in the Analysis and Assessment of Team Performance. Team Performance Assessment and Measurement. Theory, Methods, and Applications. M. Brannick, T., E. Salas and C. Prince. Mahwah, New jersey, LEA.

Thompson, J.D. (1967). Organizations in Action. New York: McGraw-Hill.

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

Assessing Negative and Positive Dimensions of Safety: A Case Study of a New Air Traffic Controller-Pilot Task Allocation

Laurence Rognin*, Isabelle Grimaud**, Eric Hoffman, Karim Zeghal

Eurocontrol Experimental Centre, BP 15, 91222 Bretigny, France
* Pacte Novation, 2 rue du Dr Lombard, 92441 Issy les Moulineaux, France
** CRNA Sud Est, rue V. Auriol, 13617 Aix en Provence, France
{laurence.rognin; isabelle.grimaud; eric.hoffman; karim.zeghal}@eurocontrol.int

Abstract: The work reported in this paper is a preliminary investigation of the safety of a new controllers-flight crews task distribution. It proposes a vision of safety assessment as a dual process of ensuring that a situation is simultaneously "not error prone" and "error free". It suggests that in addition to measuring the risk of error occurrence, it is essential to assess how the new situation provides means for error avoidance. Results issued from small-scale real time simulations conducted with air traffic controllers illustrate this approach. Typically, the new task distribution improves controllers availability, maintain them in an anticipative position and introduces redundancies that could contribute to the system dependability. In terms of risks identified, it points out risks not only in terms of facts (e.g. loss of separation, violation of applicability conditions), but also in terms of process. It suggests to go beyond the counting of abnormal events and to perform microscopic analyses of situations aiming to combine various indicators, such as aircraft parameters and control instructions in order to detect "soon to be unsafe" situations. Last of all, in addition to automated data analysis, the paper stresses the need for replay tools enabling operational experts to make sense of controllers activity.

Keywords: Air-traffic control, delegation, human-in-the-loop experiments, task distribution, safety indicators.

Introduction

Safety is generically defined as the “Freedom from unacceptable risk”. (ISO/IEC Guide 2, 1996). In air traffic control, it is transposed into:

> “While providing an expeditious service, the principal safety objective is to minimise (…) the risk of an aircraft accident as far as reasonably practicable” (SMS Policy, 2000).

Safety assurance lies on the combination of four main means, which are error prevention, prediction, avoidance and tolerance (Laprie et al., 1993). This approach relies on the acceptance of errors as unavoidable events. Even though analyses provide solutions to eliminate as much errors as possible, they also define back-up solutions supporting the tolerance of errors through preventing their propagation. The implementation of such means requires preliminary identification of risks, including the understanding of errors and of their context of occurrence. As illustrated by Shorrock & Kirwan, retrospective, predictive and real-time based applications are complementary to conduct such investigation of safety. However, most of the existing approaches show three main limits.

First, a failure is fortunately quite a rare event. Even though its analysis enables to identify the succession and combination of unsafe events, those are by nature accidental and might not, hopefully happen twice in the same exact conditions. Consequently, the "micro-incident" approach (Bressolle et al., 1996; Rognin et al., 2000) is interesting. Microscopic analysis of what looks like nominal situations in air-traffic control, highlight numerous mistakes and abnormal events detected and recovered before leading to failures. The studies show first how the system organisation (including redundant monitoring supports) enables implicit interactions and unexpected loops of control to emerge and second how these emerging mechanisms contribute to the early recovery of errors. What is suggested is to go beyond an overall safe-looking system, and question underlying and not so apparent tolerance means rather than failures. Such an approach provides insight on context of error occurrence ("unsafety" indicators) and on tolerance means introduced in the system, often under the form of recovery actions performed by the controllers themselves (safety indicators).

Second, a failure is a resulting event, which often materialises the combination of erroneous information processing, decision making and execution of actions. Error prevention and tolerance require risks of errors to be previously identified and understood so that safety means can be introduced in the system. It rests more on understanding the context of production, than on qualifying the production itself (number, frequency and criticality of the event for example). Consequently, from safety perspective, understanding the context of production of initial errors is more essential than counting their materialisation under the form of failure[4].

Third, even though the objective of safety is the absence of failure (typically, in air-traffic control, the absence of loss of separation between aircraft), the main indicator used in safety assessment analysis is the probable number of failures. Safety is actually tackled through the assessment of risks of un-safety. Safety assessment methods often focus on the negative side of safety. Their main objective is to ensure that errors will be absent from the system. Typically, the three steps of the safety assessment analysis (FHA, PSSA and SSA) proposed by EUROCONTROL (Eurocontrol, 2001) focus on the identification and mitigation of potential risks. Functional Hazard Assessment (FHA) aims to identify potential hazards, evaluate their consequences and specify the maximum probability of occurrence of such hazards. Preliminary System Safety Assessment (PSSA) questions possible degradation of safety. System Safety Assessment (SSA) provides assurance that each system element (and ultimately the whole system) as implemented meets its safety requirements. Even more promising method, such as CRIA (Marti et al., 2001) focusing on interactions between components, investigates criticality of interactions. These methods tend to "prove the completed design is safe rather than construct[ing] a design that eliminates or mitigates hazards" (Leveson et al., 2001). Safety is more than the absence of error or failure. It should also include the definition of the context in which errors should not occur or could be tolerated. In air-traffic control, this could be understood as the guarantee that despite the occurrence of local errors (which must be accepted as unavoidable events), the overall traffic handling (including safe separation between aircraft) is ensured. It relies on providing a "not unsafe" working environment. Even though the nuance might sound subtle, there is a gap between avoiding errors and avoiding error-prone conditions. Typically, it requires ensuring that actors know the objectives of their tasks, are in a position to make the appropriate decisions and execute correctly the appropriate actions. This means that people are continuously aware of the situation, in terms of perceiving the information, making sense of it and anticipating correctly how it will evolve (Endsley, 1994). In air traffic control, it requires for example usable tools, acceptable workload and time pressure enabling actors to anticipate events (rather than react to them). In addition, interactions between the various components, information sharing, mutual understanding and control loops that have proven to contribute to systems safety (Rognin et al., op. cited) also need to be secured.

This vision of safety as the provision of an environment that is both safe (i.e. error free) and not unsafe (i.e. prone to error prevention and tolerance) will be illustrated in the present paper, through a case study in air traffic control domain. Human in the loop experiments have been conducted in order to assess the impact of a new task distribution on controllers and flight crews activity. After a brief description of the context, the new task distribution, known as "delegation of spacing tasks from the controller to the flight crew" is introduced. The experimental method set up to assess the concept of delegation is explained in the second section. In the third section, benefits and expected risks induced by delegation are presented. In the last section, initial indicators and measures of safety and unsafety issued from the case study are discussed.

Delegation of spacing tasks

Spacing tasks in approach: Air-traffic control is the service provided to airlines, ensuring that the separation standards between aircraft are maintained. Closely related to the temporal organisation of flights (air-traffic management), it is restricted to actions on aircraft aiming to avoid collisions and manage the daily traffic. Air traffic control is usually decomposed into three main activities: guidance (tower), flows integration (approach) and crossing (en-route). In approach control, controllers' objective is a sequence of horizontally spaced aircraft exiting the sector. Controllers' strategies are usually based on two main options: either act on the speed or on the heading parameter. The controller's task consists in first identifying in which order to build a sequence (based on each aircraft speed, type, level, current position and heading), second choosing the appropriate strategy enabling space to be created and third ensuring that the obtained space is maintained down to the sector exit.

Rethinking the function allocation: Today's challenge in the domain of air traffic control is the foreseen increase of traffic load. While the capacity of the system is expected to double within the next 10 years, its level of safety has to be maintained if not improved. One of the options aiming at supporting controllers to cope with increasing traffic is to envisage a new distribution of tasks. This was widely investigated between controllers and systems through the development of assistance tools or automation (e.g. for conflict detection or resolution). New distributions can also be envisaged between controllers and flight crews. In this context, most of the studies rely on the "free flight" paradigm in which the whole separation assurance lies to flight crews (e.g. Johnson et al, 1999). In terms of task distribution, this induces a swapping of roles: flight crew becomes the primary actor for separation assurance whereas the controller is supposed to

[4] For an extended discussion about the distinction between error and failure, see Laprie, 1993. In the paper, we consider error as an initial local event (e.g. wrong instruction given) which might lead, under certain circumstances to a failure (loss of separation) that is visible at the system level.

supervise the traffic and intervene as a last resort in case of failure. The controller is thus in a position of acting by exception, which raises the key issue of human performance (Wickens, 1997; Billings, 1997; Corker, 1999; Dekker & Woods, 1999). On the opposite side of the spectrum, some studies have proposed a more conservative approach (Zeitlin et al., 1998; Pritchett et al., 1999; Casaux & Hasquenoph, 1997). Rather than following technical driven approaches, exploring how safety-critical tasks could be delegated to automated systems, we consider how function re-allocation (or tasks re-distribution) among the existing "components" could improve the current system.

Redefining the controller - flight crew task distribution: In the scope of redefining a task distribution between controllers and flight crews, from the onset of the project, two key constraints were identified and adopted. The first one is related to human aspects and can be summarised by "take into account current roles and working methods of controllers and flight crews". The second one is related to technology and can be expressed by "keep it as simple as possible" (Grimaud et al., 2000). Actually built around controllers and flight crews existing roles and activities, it is based upon the following elements: delegation remains upon controller's initiative, who delegates only "low level" tasks (e.g. implementation and monitoring) as opposed to "high level" tasks (e.g. definition of strategy).

For illustration purposes, let us consider the following example: two aircraft (DLH456 and AFR123) are flying along merging trajectories in descent with compatible speeds. In order to sequence DLH456 behind AFR123, the initial spacing (4 nautical miles) needs to be increased to 8 and then maintained until the waypoint (WPT), despite speed reductions imposed to the leading aircraft in descent phase. The strategy consists in giving first a heading in order to increase spacing, and then adjusting speed to maintain the acquired spacing (Figure 3).

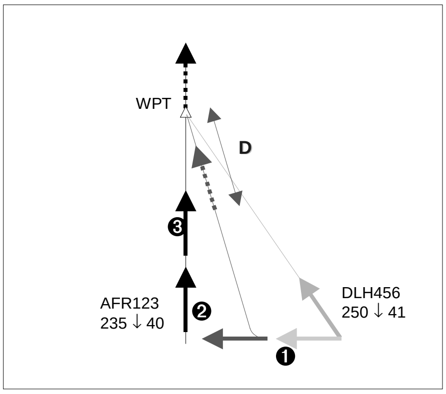


Figure 3 - "Heading then merge behind" scenario.

Without delegation, the controller gives a first heading instruction, which is executed by the flight crew. The controller monitors the aircraft until enough space is obtained. Then, the controller gives a new heading instruction, followed by heading change in the cockpit. The controller monitors the resulting spacing in order to assess if additional instruction is required to maintain it. If so, the controller gives a speed instruction, which leads to speed adjustment in the cockpit. With delegation, the controller still gives a heading instruction to initiate the spacing, and instructs the flight crew to "merge behind the target". This instruction encompasses: ❶ report when the predicted spacing at the merging point reaches the desired spacing; ❷ resume own navigation to the merging point, and ❸ adjust speed to maintain the desired spacing. The flight crew implements the task which now includes the understanding of the context, the identification of the resume point, the monitoring of the situation, the execution of the resume and the potential speed adjustments. The new task distribution is expected to contribute to safety in reducing controller workload, improving situation awareness on ground and in the air and enabling anticipation of actions in the flight deck. In addition, delegation should provide redundant monitoring and control, contributing to error detection and recovery.

Expected impacts of delegation

Human-in-the-loop experiments involving controllers and flight crews have been conducted in order to assess the possible benefits, limits and impact of delegation. An initial part task real time simulation, involving five controllers and two airline pilots, was conducted in 1999 in order to get initial feedback on the operational feasibility and potential interest of the concept (Grimaud et al., op. cited). Beyond, the objective was to identify user requirements, other possible applications and evolutions, as well as indexes of evaluation for future experiments. The results were qualitative indications gathered through questionnaires and debriefings, with an inherent subjective component in controller and pilot responses. The overall feeling was "promising" with a "great potential", and "could reduce workload". Later (in 2000 and 2001), three small-scale real time experiments over a total of seven weeks involved eighteen

controllers from different European countries. The main objective was to validate the concept in a more realistic ground environment and get an initial quantitative evaluation of its impact on controllers' activity (Grimaud et al., 2001). During the latest experiment, in addition to issues previously identified, new indicators of quality of control, safety (typically transfer conditions), controllers activity (typically monitoring tasks) and initial impact on airborne side were included. Controllers' activity was decomposed between active control (derived from instruction given) and monitoring tasks. System recordings provided information about types of instructions, time of occurrence, duration and number of communications.
From the onset of the project, hypothesis related to the impact of delegation on safety emerged. Both advantages and limits of delegation were identified, in terms of impact on the following aspects, relevant for both controllers and pilots. However, only controllers' perspectives are addressed in this paper.

Availability: With delegation, controllers should focus their attention on the building task (first part of the sector entry) in order to identify what can be delegated. Once spacing task is delegated, controllers are expected to be relieved of the speed instructions previously associated with maintaining spacing. However, in terms of monitoring, the situation should remain unchanged: controllers responsibility require them to "keep an eye" on the traffic, even when delegated. It is expected that the availability gained in terms of active control could be used for the monitoring task, and typically through readiness and ability to detect and recover drifting situations. Regarding the flight deck perspectives, despite a limited realism, we considered the number of instructions per aircraft as a possible indicator of overload in the cockpit.
Task distribution: Delegation induces a new task distribution, where the general task of "sequencing" is split between the sequence elaboration and the spacing achievement (acquisition and maintaining). We expect a better organisation of task, where actors' expertise is appropriately used. Typically, controllers are in the best position to define strategy (thanks to their overall picture of the traffic and their knowledge of how to sequence traffic) while flight crews are more appropriate to identify the accurate action enabling the spacing to be obtained. However, roles might potentially become confused. First, controllers might forget their responsibility and reduce their monitoring, expecting too much from flight crews' ability to maintain the spacing despite the non-respect of applicability conditions. Second, because traffic is displayed in the cockpit, flight crews might be in a position to question controllers strategies.
Situation awareness (mental picture): With delegation, controllers should anticipate the traffic situation, in order to prepare delegations that are initially and remain ultimately feasible, despite flight variations and later flows integration. This requires controllers to build an overall mental picture (as opposed to a more local one used to define the strategy). In addition, the availability provided by the removal of speed instructions represent extra time possibly used by controllers to update their situation awareness. However, delegation induces strong dependencies between aircraft. Maintaining the global picture of chained aircraft and anticipating the propagation of constraints along the chain might become complex and cognitively demanding.
Error management: In addition to reducing controllers peaks of workload (and consequently risks of errors), flight crews are put in a position to contribute directly to error detection and recovery. Typically, when detecting an infeasible delegation or a violation of applicability conditions, flight crews might detect an erroneous strategy chosen by controllers, or an omission of updating the situation. Delegation provides redundant monitoring and loops of control. However, with the introduction of new tasks (target selection, monitoring), delegation also induces new types of errors that need to be identified and managed.

Measuring the positive impact of delegation
The results presented in this section correspond to the second week of measurements, in condition of high traffic load (31 arrivals per hour), without and with delegation.
During the experiment, neither the rate nor the duration of delegation was constant: it evolved from 35% of aircraft delegated during 25% of their flight time to 68% delegated over 60% of their flight time. Influencing factors were the level of traffic load (medium or high), the sector configuration (requiring more or less anticipation in integrating the flows) and also the controllers confidence in the new instructions. The progressive use and the adaptation to the situation constraints highlight an understanding of the concept both in its benefits and limitations. In addition, the type of instruction delegated (maintain spacing versus obtain and then maintain) differs according to the situations: 82% versus 18%. According to us, it reflects that providing different levels of delegation enables controllers to handle predictability issues. Typically, controllers pay attention before delegating the most complex type (obtain then maintain) in which they lose part of their control on the situation. Indeed, the resume phase (and consequently the trajectory up to the resume) is under flight crews decision.
Availability: With delegation, the number of instructions (including delegation ones) was reduced by 28%. Typically, speed instructions were reduced by 70% and heading instructions by 47%. In addition, we considered the geographical distribution of instructions and eye fixations (Figure 4). It appears clearly that

with delegation sequences are built earlier, and the tasks of maintaining the spacing is no longer performed by controllers. This confirms that delegation provides availability, and enables controllers to focus on more complex problems. Because delegation relieves controllers from time-consuming and more importantly time-critical tasks (e. g. resume an aircraft at the appropriate moment), it should reduce the time pressure and consequently associated risks of errors. In terms of monitoring tasks, with delegation controllers are still focusing on the building area, whereas their monitoring is over the latest area without delegation. Without delegation, it looks like controllers are no longer anticipating the sequence building, but rather reacting and taking "last minute" decisions. Regarding pilots availability, we analysed the number of instructions given to each aircraft. We observed that in high traffic, with delegation, the maximum number of instructions per aircraft was between 1 and 6 (only 3 aircraft received 7 or 8 instructions), while it went up to 11 without delegation (2 aircraft even receiving 12 or 13 instructions per flight). Focusing on speed instructions, we see that with delegation more than twice the number of aircraft get no instructions, and only 2 aircraft received 2 speed instructions per flight (against 19 without delegation). It invalidated the hypothesis that the use of delegation could be detrimental to some aircraft. Even though delegation reduces the number of instructions given to pilots, the ground experiment does not inform us about the number of speed adjustment performed in the cockpit. Such results are expected from the experiment run specifically on the airborne side.

Task distribution: The geographical distribution of instructions (Figure 2) shows that the controller no longer performs the task of maintaining the spacing (corresponding to the usage of speed instructions between 0 and 100 Nm from the IAF). In the reported experiment, the limited realism of the airborne side does not inform on the performance of these tasks by the flight crews[5]. The monitoring curves show that with delegation controllers spend most of their monitoring time focusing on building areas, supposedly getting aware of the situation and possibly defining the appropriate strategy. Whereas a large percentage of fixations is noted over the building areas, there is still some monitoring all over the sector. We assume that it reflects regular even if less frequent monitoring, assessing if everything works as expected. However, this raises an issue in terms of unsafety: some monitoring is still performed all along the sector, but what does the reduction of the monitoring curve once the aircraft are delegated mean? Are delegated aircraft less monitored? Are there risks that controllers over-trust flight crew and implicitly forget that they remain responsible for the safety of the traffic? In order to answer these questions, we performed a microscopic analysis of fixations per aircraft: first we analysed the number of fixations per aircraft and then the interval between two fixations on a same aircraft. The results provide three answers: no aircraft were forgotten, neither without nor with delegation. The frequency of monitoring was similar for aircraft delegated and not delegated.

Even though their investigation will be necessary, we did not investigate the task distribution between executive and planning controller at a same position. However, we did question possible impact of delegation on interaction between sectors, because the quality of transfer has an impact on the next sector activity (typically if recovery or correction of actions is necessary). We considered the distribution of aircraft as a function of their spacing value. Without delegation, 17% of the aircraft are transferred with a spacing between 7 and 9 Nm, whereas delegation enabled 52% of the aircraft to be transferred in similar conditions. From the perspective of the receiving controller, the traffic is more homogeneously spaced. Regarding the closure rate (speed difference between delegated and target aircraft), the impact of delegation is less impressive: with delegation60% of the aircraft were sent in stable conditions (+/– 10 knots difference between the aircraft), against only 45% without delegation.

Situation awareness: even though the eye-tracker analysis show that something happened in terms of visual information acquisition, we have no objective evaluation of the possible impact on situation awareness. It is envisaged in the future to analyse controllers' ability to detect abnormal events, either announced or not by flight crews (e.g. technical problem on board or erroneous action in the cockpit). The use of degraded scenarios and consequently the management of failures can be interesting in terms of duration of failure, in the sense that they could inform about controllers' situation awareness. Typically, indicators such as time to detect a problem, time to recover it and quality of the solution implemented (e.g. number of manoeuvre per aircraft, number of aircraft impacted) will be analysed. In terms of unsafety, these observations stress risks that could be defined in terms of controllers tendency to delegate too much, expecting flight crews to do more than what is really delegated (including maintaining spacing even outside feasible conditions). A complementary indicator to consider here could be the transfer moment. It appeared that with delegation, some aircraft were transferred later. The underlying question here is: were the aircraft forgotten and possibly removed from controllers' mental picture?

[5] However, initial observations of the latest airborne experiment show that not only pilots did perform the speed adjustments, but they also succeeded in remaining within less than 1 Nm from the required spacing.

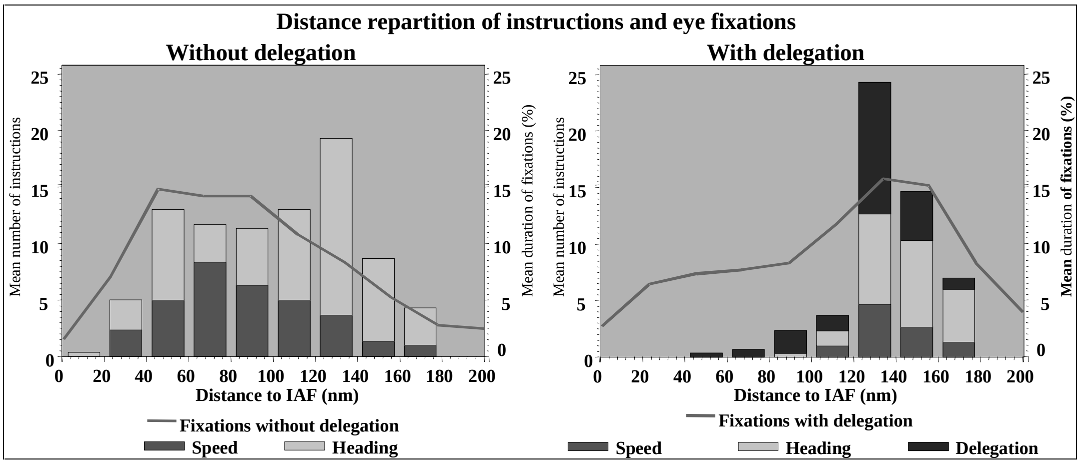


Figure 4. Geographical distribution of instructions and eye fixations without delegation (left) and with delegation (right).

Measuring risks induced by delegation

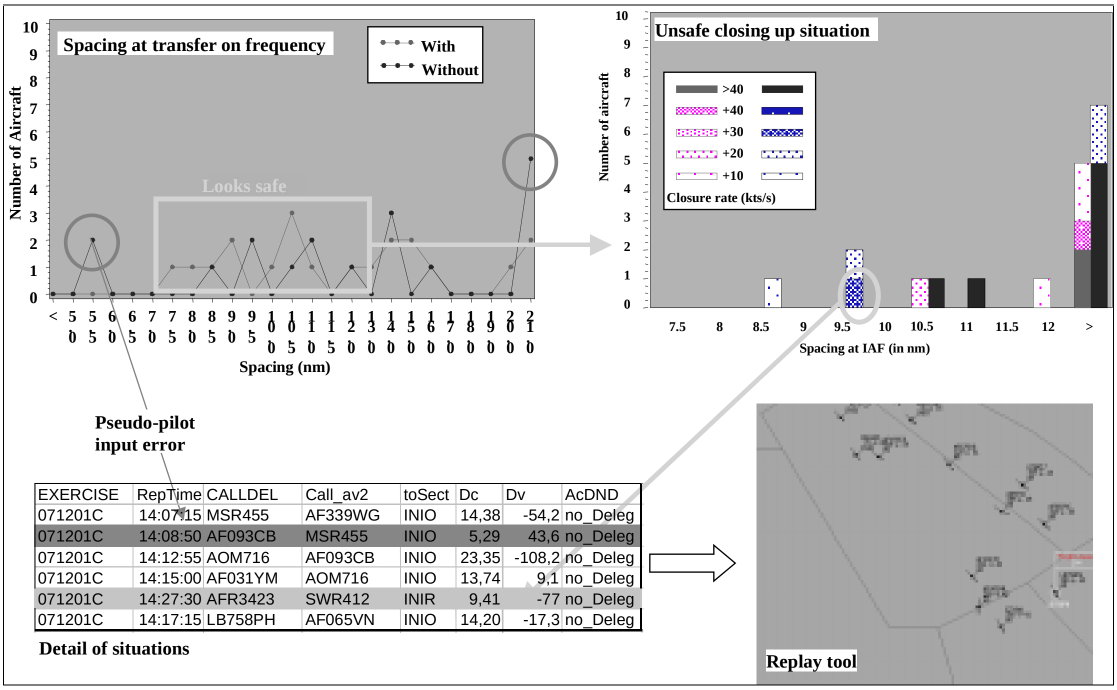
Delegation-related errors: An initial typology of delegation-related errors described their potential contexts of occurrence, causes (e.g. cognitive tunnel vision, slips, incomplete/incorrect knowledge) and possible means for their avoidance or tolerance (Rognin et al., 2001). In order to automatically detect some of them, we defined types and conditions of occurrence. The resulting structure is composed of four categories: non-respect of initial applicability conditions, non-maintaining of these applicability conditions, misuse of delegation (e.g. giving instructions not compatible with delegation) and use of superfluous instructions. Once these events were detected and documented (exercise and aircraft concerned), an operational expert analysed the conditions in order to understand motives behind the error (lack of training, misuse of delegation, simulation pilot error).

Losses of separation and spacing at exit point: Compared to standard task distribution, delegation did not induce more failures (i.e. losses of separation). With high traffic, 4 losses of separation were observed in both conditions (without and with delegation). Initial discussion about the limitations of such an indicator must be completed here, in the specific context of the spacing tasks in approach. The objectives of the controller (and actually of the flight crew with delegation) is to obtain and maintain a given spacing[6] between aircraft until their exit point. Therefore, in addition to losses of separation we looked at losses of spacing at the exit point (i.e. difference between requested and obtained spacing). We considered the number of aircraft transferred with less than 7 Nm (8 was the requested one). It showed that with delegation more aircraft had not acquired the desired spacing when transferred. This could be explained: with delegation, the desired spacing is supposed to be acquired over the exit point, and not when transferred to the next frequency (which occur about 40Nm before the exit point). The reason why we did not consider the exact spacing when the aircraft are over the exit point is related to the experimental set-up. Indeed, once transferred, the aircraft might be given new instructions by the next sector. Therefore, when geographically over the next sector, they might no longer reflect the results of initial instructions.

Conditions of transfer and impact on next sector: Following the idea that even behind seemingly nominal situations (orange rectangle on Figure 5), abnormal events or process could be taking place, we considered cautiously the conditions of transfer. In addition to building sequences and maintaining safe spacing between aircraft, it is expected that controllers transfer safe and stable situations. Whereas the spacing at transfer reflects a discrete event, it does not inform about dynamic aspects. We did observe without and with delegation, situations where correctly spaced aircraft were transferred in a catching up situation. As evoked previously, spacing is not a discrete event, but rather an evolving situation. Due to aircraft respective speed variations, a "soon to be" unsafe situation was transferred to the next sector. In order to measure unsafe conditions of transfer, we combined spacing indicators with closure rate indicators (basically speed differences between successive aircraft), in order to detect acceptable spacing possibly between delegated aircraft faster than its target (orange circle on Figure 3). In addition to analysing the detail of identified problem (exercise, time, aircraft involved, speed difference, current spacing, delegation status, as listed in the table on Figure 3), we used replay tools to go back in time and investigate the whole process that led to the loss of spacing. Indeed, analysing in details context of losses, including their recovery, was essential. It enabled the understanding of the initial conditions, the reasons for losing separation (controllers or flight crews' error, slip or mistake) the status of aircraft concerned (delegated or

6 Separation and spacing do have a very different meaning in air traffic control. Separation refers to safety minima defined by air traffic management regulations, whereas spacing refers to a distance fixed by controllers in order to organise their flows of traffic.

not), the time needed to detect and recover the problem, the complexity and the quality of the solution implemented (including the number of aircraft impacted). Beyond unsafety per se, unstable transfers may lead to an unacceptable workload increase in the receiving sector. In the current working practices, the receiving sector is in charge of integrating various flows of aircraft (typically Paris Orly approach is composed of two main and one marginal flows). Spacing and stabilising incoming aircraft might then represent an additional task. In addition to assess the possibility of integrating an aircraft within an existing flow, approach controllers are then in charge of ensuring spacing within this existing flow.



EXERCISE	RepTime	CALLDEL	Call_av2	toSect	Dc	Dv	AcDND
071201C	14:07:15	MSR455	AF339WG	INIO	14,38	-54,2	no_Deleg
071201C	14:08:50	AF093CB	MSR455	INIO	5,29	43,6	no_Deleg
071201C	14:12:55	AOM716	AF093CB	INIO	23,35	-108,2	no_Deleg
071201C	14:15:00	AF031YM	AOM716	INIO	13,74	9,1	no_Deleg
071201C	14:27:30	AFR3423	SWR412	INIR	9,41	-77	no_Deleg
071201C	14:17:15	LB758PH	AF065VN	INIO	14,20	-17,3	no_Deleg

Figure 5. Microscopic analysis of transfer conditions.

Respect initial applicability conditions, then maintain them: Specific applicability conditions need to be respected in order to benefit from delegation. Ensuring the feasibility of delegated tasks is part of the controllers' tasks. In addition to this initial assessment, controllers are in charge of maintaining the applicability conditions during the flight. The main items defined in the applicability conditions are compatible speeds (e.g. ensure a slow aircraft is not asked to catch up on a much faster one), compatible trajectories (e.g. ensure an aircraft is not asked to merge behind an aircraft following a diverging path) and compatible flight levels. One of the difficulties induced by delegation is the mutual dependencies between delegated aircraft, and consequently the cognitive cost of maintaining appropriate situation awareness in case of long chains. Whereas it is quite easy for a controller to understand that an aircraft is reducing speed because its target is descending, the task becomes harder when the speed reduction is suddenly observed for an aircraft ending a long chain (e.g. number 6) and actually reacting to the descent of the aircraft number 1. The resulting speed reduction might be amplified all along the chain and therefore result in inability to respect the delegation any longer. In order to investigate systematically the conditions of delegation, we defined what were the applicability conditions in most of the expected situations: e.g. stable situation, descending target. Then, basic indicators, such as relative trajectories, relative speed, relative altitude were associated (and sometimes combined) to each cases. The third step consisted in analysing applicability conditions for each delegation, from its start until its end. The results show that the most frequent errors were related to the initial assessment of applicability conditions (non compatible speeds and target not direct to a waypoint). These results are now questioned in terms of causes and potential means for their prevention or/and tolerance. Typically, whereas some errors were knowledge-based (lack of expertise), other were simulation-related and should have been detected by the flight crews in a real environment. In addition to detecting errors, the analysis of applicability conditions will guide the definition of algorithms for an error detection function (possibly introduced as a ground or an airborne support). Last of all, analysing monitoring patterns could help assessing the cost of checking and respecting applicability conditions. Typically, whereas we expect delegation to induce a change in the data monitored (aircraft flight level instead of respective positions), it is necessary to evaluate first if the modified pattern is efficient and second if it is more or less complex. This is directly related to the issue of predictability. How do controllers anticipate aircraft behaviour, are they continuously aware of the current situation and do appropriately envisage how the situation should evolve. For the time being, monitoring efficiency could be investigated using degraded scenarios might indicate controllers ability to detect drifting situations. However, it will be necessary to distinguish two underlying assumptions: soon detection would reflect an appropriate monitoring, based on correct expectations and focused on relevant information. For the time being, comparing the complexity of monitoring is only envisaged through controllers subjective feedback.

Conclusion and next steps

The work reported in this paper is a preliminary investigation of the safety of a new controllers-flight crews task distribution. It proposes a vision of safety assessment as a dual process of ensuring that a situation is simultaneously not error prone and error free. It suggests that in addition to measuring the risk of error occurrence, it is essential to assess how the new situation provides means for error avoidance. Results issued from small-scale real time simulations conducted with air traffic controllers illustrate this approach. Typically, the new task distribution improves controllers availability, maintain them in an anticipate position and introduces redundancies that could contribute to the system dependability. In terms of risks identified, it points out risks not only in terms of facts (e.g. loss of separation, violation of applicability conditions), but also in terms of process. It suggests to go beyond the counting of abnormal events. Typically it proposes a microscopic analysis of the situation, in which various indicators, such as aircraft speed, distance, altitude are combined in order to detect "soon to be unsafe" situations. Last of all, in addition to automated data analysis, the paper stresses the need for replay tools enabling operational experts to make sense of controllers activity.

In terms of safety assessment, a further step could consist of investigating controllers and flight crews ability to detect and recover incidents. However, the main limitations of the current experimental setting is the limited realism of the simulation environment, and more specifically the airborne component. Typically, flight crew contribution to error detection (e.g. non respect of applicability conditions) could not be simulated. This limit should be overcome in the context of the next ground simulation, where the system will provide simulation pilots with information enabling them to assess the feasibility of delegation. A human in the loop experiment, focusing on the airborne side was also conducted in May 2002 with five flight crews. The data collected are currently analysed, using similar indicators. A new airborne experiment is planned next winter. For the time being, no fully integrated simulation (combining controllers and flight crews) is envisaged: indeed, such an experiment requires the concept to be at a more advanced stage.

Acknowledgements

The authors would like to thank controllers, airline pilots and simulation pilots for their participation in the experimentation and the technical team for their support. The eye tracker analysis has been performed with the support of Luc Rodet and Anne Pellegrin from Novadis, Grenoble, France.

References

Billings, C. E. (1997). Aviation automation. The search for a Human-Centered Approach. L. Erlbaum Publishers. NJ.

Bressolle, M.-C., Decortis, F., Pavard, B. & Salembier, P. (1996). Traitement cognitif et organisationnel des micro-incidents dans le domaine du contrôle du trafic aérien: Analyse des boucles de régultaions formelles et informelles. In G. De Terssac & E. Friedberg (Eds.) Coopération et conception. Toulouse, Octarès Editions.

Casaux, F. & Hasquenoph, B. (1997) Operational use of ASAS, USA/Europe ATM R&D Seminar, Saclay, France.

Corker, K., Flemming, K. & Lane, J. (1999). Measuring controller reactions to free flight in a complex transition sector. Journal of ATC, Oct.–Dec.

Dekker, S. & Woods, D. (1999). To intervene or not to intervene: the dilemma of management by exception. Cognition, Technology & Work, 1, 86-96.

Endsley, M.R. (1994). Situation Awareness in dynamic human decision-making: Theory. In R.D. Gilson, D.J. Garland & J.M. Koonce (Eds), Situational awareness in complex systems (pp.27-58). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

EUROCONTROL (2001). Guidelines for the safety assessment of EATMP Programmes. SAF.ET1.ST03.1000.MAN-02-00. Edition V3.1, 08.05.2001.

Grimaud, I., Hoffman, E. & Zeghal, K. (2000). Evaluation of Delegation of Sequencing Operations to the Flight Crew from a Controller Perspective - Preliminary results. SAE/AIAA World Aviation Congress, San Diego.

Grimaud, I., Hoffman, E., Rognin, L. & Zeghal, K. (2001). Insight into Controller Activity through a Geographical-Based Analysis of Instructions.
Digital Avionics Systems Conference, Florida, USA.

ISO/IEC Guide 2 (1996). Standardization and related activities - General vocabulary.

Johnson, W., Battiste, V. & Holland, S. (1999). A Cockpit display designed to enable limited flight deck separation responsibility. SAE/AIAA World Aviation Congress, San Francisco.

Laprie J-C. Dependability: from concepts to limits. Symposium on Safety of Computer Control Systems (SAFECOMP'93), Poznan, Poland. Springer: Berlin, 1993. p. 157-68.

Leveson, N., de Villepin, M., Srinivasan, J., Daouk, M., Neogi, N., Bachelder E., Bellingham, J., Pilon, N. & Flynn, G. (2201). A safety and human-centered approach to developing new air traffic management tools. 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, USA.

Marti, P., Lanzi, P. & Pucci, F. (2001). Evaluating safety and usability of ATM systems. 4th USA/Europe Air Traffic Management R&D Seminar, Santa Fe, USA.

Pritchett, A.R., Yankosky, L.J. & Haserueck, D. (1999). Simultaneous design of cockpit display of traffic information and air traffic control procedures. In 10th International Symposium on Aviation Psychology (pp.88-94), Columbus, USA.

Rognin, L., Grimaud, I., Hoffman, E. & Zeghal, K. (2001). Implementing Changes in Controller-Pilot Tasks Distribution: the Introduction of Limited Delegation of Separation Assurance. International Workshop on Human Error, Safety and Systems Development (HESSD), Linköping, Sweden.

Rognin, L., Salembier, P., & Zouinar, M. (2000). Cooperation, reliability of socio-technical systems and allocation of function. International Journal of Human-Computer Studies, 52 (2), pp.357-379.

Shorrock, S. T. & Kirwan, B. (2002). Development and application of a human error identification tool for air traffic control. Applied Ergonomics, 33, 319-336. Elsevier.

SMS Policy. (2000). EATMP Safety Policy. SAF.ET1.ST01.1000-POL-01-00, Edition 2.0.

Wickens, C. D., Mavor, A. S. & McGee J. P. (1997). Flight to the Future: Human Factors in Air Traffic Control. National Academy Press.

Zeitlin, D., Hammer, J., Cieplak, J. & Olmos, B.O. (1998). Achieving early CDTI capability with ADS-B, USA/Europe ATM R&D Seminar, Orlando.

Head-Mounted Video Cued Recall: A Methodology for Detecting, Understanding, and Minimising Error in the Control of Complex Systems

Mary Omodei, Jim McLennan (2), Alexander Wearing (3)

(1) School of Psychological Science, Latrobe University, Melbourne, 3086, Australia.
Email: m.omodei@latrobe.edu.au
(2) Student Services, Swinburne University of Technology, Hawthorn, 3122, Australia.
Email: jpmcl@ozemail.com.au
(3) Department of Psychology, University of Melbourne, Parkville 3052, Australia.
a.wearing@psych.unimelb.edu.au

Abstract: In this paper we introduce head-mounted video camera footage as a sensitive and reliable method for enhancing post-incident recall of mental events associated with decision error. We first present a model of human decision making control of complex, uncertain, multi-person dynamic systems, which allows for a potentially useful conceptualisation and classification of error. The head mounted video cued recall procedure is then described, an own-point-of-view psychological perspective being emphasised as the defining characteristic of this procedure. It is argued that this perspective accounts for the methodology's ability to generate data not obtainable using other methods. As such, it is particularly suited to obtaining data on underlying psychological processes, especially those associated with error. We illustrate and evaluate the application of the methodology in several contexts that have hitherto proved to be particularly difficult to study satisfactorily, including firefighting and forest navigation.

Keywords: Head-Mounted Video, Cued-Recall, Decision Making, Error, Command and Control

Introduction

Identifying, understanding, and minimising error behaviour, and/or associated error tendencies, constitutes one of the main themes in the literature on human decision making processes in the control of complex socio-technical systems. Such a focus on error is a productive research activity, not only because of its direct contribution to the implementation of safer systems, but also because the study of error provides a "window" on psychological processes underlying human decision making generally.

A Theoretical Framework for the Decision Making Control of Complex Systems

We have found it particularly useful to integrate the theoretical insights of two research traditions: laboratory-based investigation of dynamic decision making (Brehmer, 1992; Dörner & Wearing, 1995) and field-based observations of naturalistic decision making (Zsambok & Klein, 1997). In doing so we have developed a conceptual model of the key psychological processes involved in the decision making control of complex, multi-person, dynamic systems (Figure 1). The central (vertical) axis of this model comprises the stages through which each individual can be understood to cycle over time in attempting to control a problematic situation. These stages encompass three major classes of cognitive activity: (a) situation assessment, (b) intention generation, and (c) action selection. Naturalistic accounts of decision making in complex settings suggest that a relatively large proportion of cognitive activity is allocated to the first of these phases, the development and maintenance of situation awareness (Klein, 1989). Following Endsley (1995), situation assessment can, in turn, be described as comprising three stages through which an individual cycles over time: (a) perception of salient elements of the environment, (b) comprehension of the relationships among these elements, and (c) projection of the future state of the environment.

On the basis of findings from our own earlier work on decision making by individuals in dynamic environments (Omodei & Wearing, 1995) we argue that any such model needs to incorporate specifically "self-evaluative control processes" (the left hand component in Figure 1). Skilled decision makers actively seek to maximise the continuing efficacy of their decisions by engaging in a constant process of self-monitoring and self-regulation. Furthermore, such self-evaluative control processes are directed not only at cognitive and attentional resources (e.g., mental workload), but also at affective and motivational states. As suggested by Brehmer (2000) many decisions are made not with a view to directly controlling a current aspect of the problematic environment, but with a view to controlling the rate at which one must make further decisions. This notion of decision 'pacing' has a long tradition, commencing with Bruner, Goodnow, and Austin's (1956) observation that purposive cognitive activity is a compromise between meeting the immediate demands of the task environment and needing to conserve one's cognitive resources to meet ongoing exigencies.

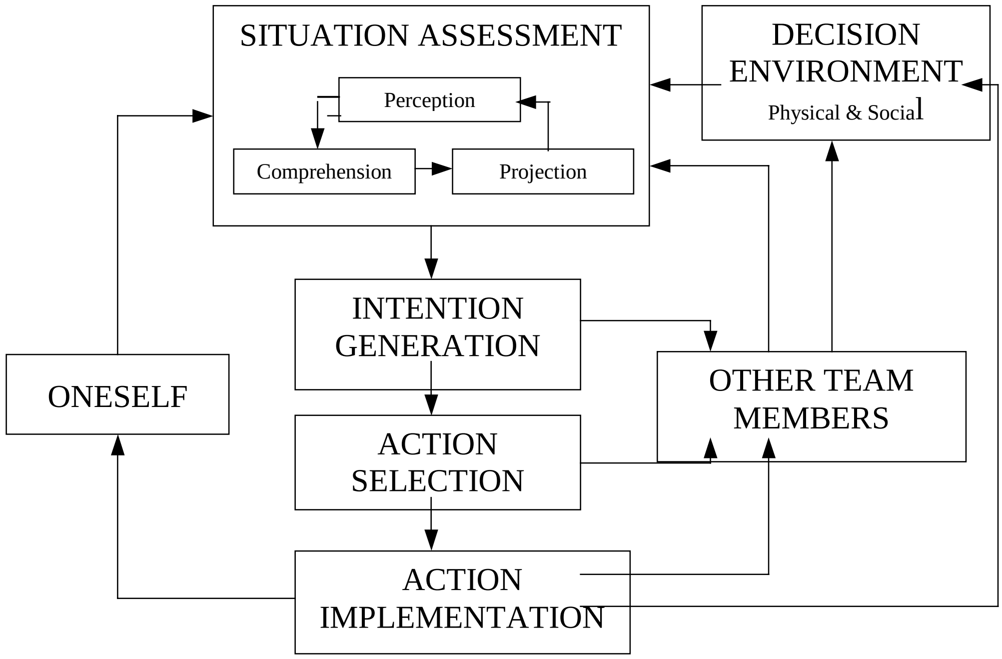


Figure 1 - Adaptive Control Model of Decision Making in Multi-Person Dynamic Systems

In extending the model to encompass the distribution of decision making responsibility in multi-person dynamic contexts, any one decision maker must take into account other persons in the decision making team. Such "social evaluative control processes" (the right hand component of Figure 1) involve assessing whether any particular individual is sufficiently competent and sufficiently motivated to follow particular instructions. Having made such a determination, the individual in charge can then issue commands in such a manner as to align a subordinate's psychological state to be maximally effective.

Note that this Adaptive Control Model allows distinctions among (a) micro-management (achieving one's aims by direct action), (b) direct orders (achieving one's aims by ordering others to implement selected actions), and (c) the communication of intent (achieving one's aims by communicating the intention(s) one has generated). The model also makes explicit (a) the continuous and iterative nature of situation assessment and intention generation/action, and (b) the fact that any action taken in the environment can serve dual purposes: (i) eliciting further information concerning the current state of the environment and/or (ii) achieving a desired change in an already identified problematic aspect of the environment.

With respect to more basic psychological processes underlying the components of the Adaptive Control Model, we have found the following to be particularly useful: (a) cognitive concepts pertaining to memory structure and functioning and to human information processing systems, (b) metacognitive concepts (how one thinks about ones' thinking and one's memory), and (c) social cognitive concepts (how one understands interactions with and among other decision makers). Although the scope of the present paper does not allow for an extensive discussion of these more basic concepts, it may be helpful for us to point out that we appeal to these same basic psychological processes in arguing for the efficacy of head mounted video cued recall as a procedure for studying decision making.

Conceptualising and Classifying Error

As proposed by Dörner (1987), humans do not appear to be well adapted to coping with the sorts of dynamic uncertain environments which characterise modern technological society. Dörner and Schaub (1994) reviewed evidence to suggest that the best way to develop a person's ability to control a complex dynamic reality is to confront the person with his or her error tendencies. Therefore to guide research into such error tendencies, what is needed is a conceptualisation and associated taxonomy of errors which allows (a) identification of errors, (b) understanding of the psychological processes generating such errors, and thereby (c) identification of potential methods for reducing such errors (both at the systemic level by redesign of the decision environment and at the human level by the implementation of specific training strategies).

In dynamic environments, the concept of error is itself complex. There is no consensus on a clear behavioural operationalisation of error, something that is more readily achieved in static well-specified behavioral domains (Shafir & LeBoeuf, 2002). For example it is not apparent how much an outcome needs to deviate from some optimal outcome before the underlying decision making is to be labeled as error. In many complex, time-pressured situations, it is almost inevitable that some decisions prove inadequate, making the term "error" perhaps misleading. Furthermore, what might constitute an error in a static environment may not do so in a dynamic environment in which there are opportunities for error recovery, the outcomes of earlier errors providing useful information on the nature of the current problem.

In order to formulate a potentially comprehensive taxonomy of error we therefore considered the writings of those decision researchers who have sought to identify error patterns in dynamic environments. A detailed account of the range of conception or error proposed, together with a full discussion of specific error types can be found in recent writings of Brehmer (1992), Dörner (1990)), Funke (1995), Green (1990), Jansson (1994), Lipshitz (1997), Norman (1998), Omodei et al (in press), Rasmussen (1990), and Reason (1990). Although there is some overlap in the specific error types listed by these various authors, this overlap is not sufficient to constitute a consensus on how errors in general should be conceptualized and categorized, or on what constitutes the more common forms of error.

In order to provide a conceptualisation and categorization of error which more comprehensively integrates the many error types identified in the literature cited above, we suggest that error can usefully be regarded as being generated by shortcomings in the execution of the various elements of our Adaptive Control Model (cf. Figure 1). The types of errors identified in this literature can be categorized according to the elements of the Adaptive Control Model (see Table 1). We suggest that such a classification of error is useful because it allows for the ready identification of specific errors, while also indicating the possible origin of these errors in underlying cognitive, metacognitive, and social processes.

Head Mounted Video Cued Recall as a Methodology for Identifying and Minimising Error

Challenges for Researchers Studying Error: The notionally direct study of error tendencies by the study of actual accidents and/or near misses (i.e., accident/near-miss autopsies) poses serious methodological challenges involving: self-justification, blame, incomplete recollection, and remoteness in time. Moreover, the effects of these factors on accuracy in error investigation are likely to be further compounded by the operation of a general hindsight bias. The operation of this bias (in which knowledge of outcome biases one's judgment about the processes that led up to that outcome) is argued by Woods and Cook (1999) to render useless many current applications of post incident accident reporting. Other less direct, but potentially more sensitive, methods for studying error comprise (a) concurrent (on-task) reporting procedures and (b) retrospective interview procedures. These approaches also, however, create problems concerning the validity of any information obtained.

The first major problem for the researcher concerns the possible "reactivity" of such methods. One has to seriously consider the extent to which concurrent reporting actually alters decision behaviour in the situation being investigated. There is ample evidence, for example, that traditional "think aloud" techniques not only distract the decision maker from the task at hand but also alters the actual phenomena under investigation (Bartl & Doerner, 1998; Dickson, McLennan, & Omodei, 2000). Such method reactivity, therefore, poses serious concerns not only for the validity of any findings, but also for the safety of all involved.

A second major problem concerns the adequacy of these methods for obtaining useful data on behaviours and experiences specifically related to error. What is needed are methods that allow for reliable, representative, and comprehensive assessment of those behaviours and experiences which underlie error tendencies. We suggest that typical strategies for obtaining self-reports (either concurrently or retrospectively) create situations in which some types of psychological processes are much more likely to be recalled than others, leading to a distorted view of decision making processes in general and of errors in particular. What are least likely to be recalled are those perceptual, affective and motivational states that are essentially pre-verbal or at least not verbalised during the flow of the decision incident. The natural tendency in providing self-reports is to present an image of the self which is not only self-enhancing, but also self-consistent, leading to distortion and censoring of material (Swann, Griffin, Predmore, & Gaines, 1987). What is of particular concern for validity in the investigation of error is that those experiences *least* likely to be recalled are often those *most* likely to be associated with error.

Errors in situation assessment
Premature foreclosure in situation assessment

Perception
Failure to monitor effects of prior actions
Channelling of information collection (according to current assessment of the situation)
Giving undue weight to the more perceptually salient cues
Restriction/reduction in information gathering
Overinclusiveness with respect to level of detail

Comprehension
Misunderstanding probabilistic information
Interpreting relationships in terms of causal series rather than causal nets
Undue weight to the present situation at the expense of its developmental history
Uncritical acceptance of others assessments
Misunderstanding of others assessments
Failure to accept one's own contribution to problematic developments (e.g.,blaming others)
Premature foreclosure in hypothesis generation (too few hypothesis generated)
Oversimplification in hypothesis generation (too global, under weight to one central variable)
Confirmation bias in hypothesis testing
Hindsight bias in hypothesis testing

Prediction
Underestimating effects of non linear relationships (e.g., exponential, circular, inverted, & lagged effects)
Inappropriate weighting of the present problem at the expense of likely future developments
Premature foreclosure in hypothesis testing

Errors in intention formation
Too few plans
Restricted plans
Less considered plans
Less organised plans
Less integrated plans (e.g., thematic vagabonding)
Stereotypic plans (e.g., encystment)
More concrete/less abstract plans
Preference for plans with the most salient outcomes
Postponement of decisions
Uncritical acceptance of others' decisions
Abrogation of decision-making responsibility
Failure to appropriately delegate (e.g., micromanagement)
Inappropriate intrusion on other decision makers

Errors in action selection
Failure to implement prior intentions
Risky actions
Rule violations

Errors in action implementation
Failure to implement selected actions
Unmonitored actions (i.e., ballistic actions)
Skill deficits
Slips
Lapses
Inaccurate content of communication
Inappropriate form of communication

Errors in self-monitoring
Overconfidence
Inadequate or insufficient self appraisal
Defensive self justification

Errors in self-regulation
Overutilisation of attentional resources
Overutilisation of memory resources
Irrelevant negative self talk
Lapses in attention control

Table 1 - A Taxonomy of Decision Errors in the Control of Complex Dynamic Systems

This raises the question as to what would constitute a suitably non-reactive but sensitive research methodology for investigating decision making in complex, dynamic contexts? What is needed is a methodology which does not distort the inherent complexity and dynamics of the underlying decision making processes, while at the same time allowing such processes to be comprehensively assessed. In the remainder of this paper we describe our use of head-mounted video camera footage, and associated cued-recall techniques, as a methodology for research (and training) in complex dynamic decision making settings. We propose on theoretical grounds that such a methodology is particularly suited for studying theoretically-relevant variables in time-critical, complex, error-prone environments (Omodei, Wearing, & McLennan, 1997).

Overview of the Head Mounted Video Cued Recall Procedure: The essence of our methodology is to capture from the head of the participant a record of the central portion of their visual field during a decision making incident. This recording, with accompanying audio, is then replayed to the subject as soon as is practical after the incident. During this replay, a relatively non-directive (i.e., discovery-oriented) procedure is adopted in which the subject is encouraged to re-immerse themselves psychologically in the original incident and to verbalise all their recollections, uncensored for relevance or appropriateness. Such recollections can include any perceptual, affective, or cognitive experiences which occurred during the initial incident. The replay tape is paused as necessary to allow the subject time to articulate their recollections, including recollections of any pre-verbal or fleeting experiences.

This uncensored recall procedure essentially takes an "insider" perspective in which all evaluation and post-incident analysis is suspended: the goal is not to probe beyond the initial experience but to simply help the subject to "re-experience" exactly what went on for them during the course of the incident. Evaluative and analytical judgments by either the subject or interviewer are put on hold, at least until after the full incident has been replayed. Typically participants often make an evaluative or critical comment at the commencement of their first encounter with head camera footage (presumably out of habit or expectation). However once the interviewer/facillitator indicates that such comments can be deferred to later in the process, participants are able to readily comply, evaluative and critical comments being noticeably lacking in the remainder of the participants' spoken recollections. We do not deny that allowing a participant to take a more evaluative/critical perspective affords an opportunity to identify underlying errors and error tendencies. What we suggest, however, is that such evaluations are more likely to be accurate (and less censored) if the participant has first re-experienced the whole of the decision episode non-evaluatively. Once the replay has reached the end of the incident, only then do we explicitly incorporate probes to encourage such an evaluative/critical perspective. These probes typically include: (a) "Now that have finished watching the video, what stands out for you most about how you handled the incident"? (b) "If you could magically go back in time and do the whole thing again, what, if anything, would you do differently, and why"? and (c)"Suppose it had not been you in charge but someone else rather less experienced, what is the most likely way he might have mis-handled the incident so it went badly wrong"? These three probes provide the participant with an opportunity to identify what are the most likely errors that could have been made. By positioning these probes at the end of the cued recall procedure, we avoid, or at least substantially reduce, potential problems of vulnerability to the operation of self justification, demand characteristics, post-incident inferences. These post-cued probes typically elicit a large amount of relevant and sometimes surprising information. In addition to allowing the participant to identify what for him or her were the most salient aspects of the decision making processes, the participant is encouraged to engage in a process of error identification in a manner which is minimally ego-threatening.

The Own-Point-Of-View Psychological Perspective: The key element of both the capture of the head-mounted video footage and the subsequent cued-recall debriefing procedure is an own-point-of-view psychological perspective. As discussed in greater detail in the following section of this paper, this combination of an own-point-of-view video image with an own-point-of-view (insider) recall perspective constitutes a powerful procedure for stimulating the recall of the maximum amount of relatively uncontaminated information on prior experiences. The own-point-of-view video perspective provides as close as match as is possible between the initial experience and the replayed image. This close correspondence is a powerful cue to the concurrent recall of other images and experiences not directly represented in the image captured by the relatively narrow camera angle. For example, while watching their replay subjects often appear unable to distinguish that which is visible on-screen (captured by the relatively narrow camera angle) from that which as off to one side – it is as if they really do see this extra information on-screen (Omodei, McLennan, & Whitford, 1998).

The own-point-of-view video perspective takes on an added dimension when one considers the effect of head movement on the image. Theoretical accounts of the link between motion and perception suggest that the experience of watching a video image taken from one's own head as one moves through an environment is fundamentally different, not only to that of watching a video taken of oneself, but also to watching a video taken from someone else's head (Kipper, 1986). We suggest that head mounted video images evoke the same psychological reactions that underlie the immersive experience of 3D virtual reality simulations, in which as a person moves their head the image displayed in the VR Goggles changes accordingly. For example, the possibly quite jerky movement in the image resulting from one's head movement actually becomes an advantage as one watches one's own replay. While increased psychological immersion is achieved by such image movement, the actual movements are quickly adapted to and become unnoticed. This was found to be the case even with the rapid movement associated with running in forest terrain (Omodei et al., 1998; Omodei & McLennan, 1994). It would appear, therefore, that the same perceptual stabilisation processes occur during image watching as occur in naturally-occurring movement. This impact of the own-point-of-view visual perspective is further enhanced by the replay of the concurrent recorded audio information. In addition to recorded verbal interactions, other sounds, ranging from one's footfall and breathing to sounds of doors closing, alarms sounding, and radio traffic, all augment the impact of the video images.

Advantages of the Own-Point-Of-View Perspective: In addition to the fact that video recording from an external perspective is often impossible or impractical (e.g., attempting to follow a fire commander into a burning or collapsed building), even when such an external video recording is possible, the own-point-of-view perspective provided by a head-mounted camera seems a far superior option for stimulating the recall of psychologically-relevant data with miminal self-conscious distortion. There are three main advantages of the own-point-of-view perspective with respect to the video recording. First, as indicated above (and the main reason for its use), this perspective provides the most accurate representation of an individual's perceptual field that it is possible to achieve, and as such generates a high level of experiential immersion. Experiential immersion is a broader concept than perceptual immersion, including not only perceptual embeddedness, but also thoughts, feelings, plans and actions (Csikszentmihalyi, 1975). As such, this own-point-of-view visual perspective is a maximally-powerful stimulus to the recollection of all types of mental events which occurred while the recording was being made. The high level of experiential immersion in the replayed incident generates a representative and comprehensive set of recollections, with minimal self-defensive justification. That is, it cues the recollection not only of specific cognitions, but also of perceptual, affective, and motivational states that were essentially pre-verbal or at least not verbalised during the original decision incident.

Secondly, the own-point-of-view perspective overcomes several of the observed disadvantages of an external video perspective that includes the subject in the field of view. Most notably, the increase in objective self-awareness (self consciousness) induced by images of oneself in action has been shown to lead to self-protective attributions and selective reporting of those experiences participants feel it appropriate or expected by the enquirer (Duval & Wicklund, 1972). In a current project studying anaesthetists in hospital operating rooms, one anaesthetist on being asked to compare the procedure with the replay of conventional 'external' video footage gave a comment that was typical of participants in the various studies we have conducted: "I don't like to see myself on camera. I find it distracting such that I often don't watch. With this [cued recall]… I don't have the image of me up there on the screen. I can see what I could see and that helped me to put myself back into what I was feeling at the time".

Thirdly, the actual wearing of a head mounted camera does not appear to generate the same level of self-consciousness as having an external video camera focused on oneself. Because it is out of sight, it is out of mind. A related concern for us initially was the extent to which the behaviour of others in the subject's environment would be altered by their awareness of the camera on the head of the subject. This would be particularly problematic should it result in changes in the way such persons interact with, or respond to, the subject. We have been quite surprised at the extent to which others habituate to, and thus become unaware of, the camera. We suggest that because the subject wearing the camera acts "naturally", likewise his or her interactions with others do not evoke ongoing self-consciousness in these others.

Addition of a second-stage debriefing process: In several of our studies we have used the material generated during the cued-recall (insider perspective) debriefing procedure described above as stimulus material for a subsequent (second) debriefing stage (McLennan, Omodei, Rich, & Wearing, 1997; Omodei et al., 1998). This second debriefing procedure is as follows: as the subject verbalises their recollections during the first stage (as above), these are added (overdubbed) to a copy of the video and audio of the original incident. This video and combined audio (of the original incident and the uncensored recollections) is then replayed to the subject. In contrast to during the first stage, during this second stage the interviewer encourages the subject to take an "outsider" perspective in order to tap into other high-level psychological processes which generated the initial recollections, including possible contextual sources of error and inappropriate decision strategies. This second stage affords both the decision maker and the interviewer the opportunity to take a more analytic and evaluative perspective on the incident, without having to be concerned that such a potentially critical perspective might inhibit or distort participant's primary recollections (because these have already been externalized during the first stage).

The addition of such a second stage debriefing is most likely to be of value in those research contexts in which the interest is not only in understanding decision making but in improving such decision making. Improvements in decision making can be achieved by creating a context in which the participant is appropriately receptive to self-confrontation with error (Dörner & Schaub, 1994), such receptivity being achieved during the first stage recollection process. Data obtained in this process of training can also be examined for theoretical insight into decision processes in general and error in particular.

Studies Using Head Mounted Video Cued Recall

Contexts in which we have employed the methodology include operational firefighting (McLennan, Omodei, & Wearing, 2001), firefighter training (McLennan et al., 1997; McLennan, Pavlou, & Omodei, in press), and the sporting activities of competitive forest navigation (i.e., orienteering) (Omodei et al., 1998; Omodei & McLennan, 1994) and football umpiring (McLennan & Omodei, 1996). Other contexts which are being investigated using this methodology include intensive care nursing, anaesthesia hospital OR emergencies, and the piloting of modern aircraft.

By way of illustration, the study of operational firefighting (McLennan et al., 2001) involved five male Senior Station Officers (SSOs) with the Melbourne Fire and Emergency Services Board in command of a shift team of approximately 12 firefighters manning three firefighting appliances. When a turnout call was received the researcher on duty ran to the firefighting appliance, switched on the video equipment and assisted the SSO to don his camera-fitted safety helmet and tunic as the vehicle left the station. Ten incidents (occurring in 30 x 12-hour shifts) were judged by the SSO concerned to require the exercise of significant incident command skill. On completion of the replay cued procedure, the officer participants were able to identify ways in which they thought they could have handled the incident better. The types of errors reported included (a) premature foreclosure in situation assessment, (b) giving undue weight to the more perceptually-salient cues, and (c) premature foreclosure in hypothesis generation. More details on these errors can be found in McLennan et al (2001). The detailed information obtained on these errors indicate the sensitivity of the methodology for revealing subtle error tendencies.

What is particularly encouraging for the validity of the own-point-of-view replay cued recollection procedure is that in all of these studies the recollections included a large number of verbalisations associated with primarily pre-verbal perceptual, affective and motivational states, and negative self-reflections such as self doubt and lapses in confidence

With respect to the robustness of the head mounted video methodology, in most of these studies the wearer of the camera moved quickly through environments characterised by noise, clutter, and potential dangers. Because a useful video image was obtained under such relatively severe conditions, an adequate recorded image should be readily obtained in most other naturally occurring decision making situations.

An Appraisal of Head Mounted Video Cued Recall for the Study of Error

In describing the rationale for the use of head mounted video recall as an own-point of view psychological perspective which induces high levels of experiential immersion, we have grounded the procedure in cognitive concepts pertaining to memory structure and functioning and human information processing systems. Furthermore, the findings of the studies referred to in the previous section illustrate the successful use of the methodology in a diverse set of dynamic decision making environments.

The findings of these studies can also be examined for evidence of the extent to which the methodology meets the two main challenges for studying error identified in an early section of this paper. The first concern is “reactivity”: the possibility that the methods adopted to detect and study error might actually alter behaviour in the situation being investigated. In all the studies conducted so far, participants reported that they soon adapted to wearing the head mounted camera and, once involved in a decision incident, actually forgot they were wearing the equipment. A typical response when approached by a researcher who wished to retrieve the camera was to express surprise. The second concern was the adequacy of the methods to afford a reliable, representative, and comprehensive assessment of those behaviours and experiences which underlie error tendencies, avoiding self-protective distortion and censoring of material. As predicted, participants in the various studies reported a high degree of experiential immersion (as a putting oneself back in time and feeling the same things again). Wherever appropriate, participants in our studies (and their trainers) have been given the opportunity to compare the head mounted video cued recall procedure with their organization’ current approach to performing post-incident reviews. In the studies referred to above, participants have commented that they not only remember more, but that they are able to (a) detect errors they were unaware of previously and (b) suggest strategies they might adopt in the future to avoid such errors (McLennan et al., 1997; McLennan et al., in press; Omodei et al., 1998; Omodei & McLennan, 1994). Furthermore, in those studies in which the participant’s trainers provided comments, they enthusiastically endorsed the procedure as providing valuable insights into trainee error patterns.

Potential limitations: The major limitation of head mounted video cued recall is that the methodology is inappropriate in those task environments in which the decision phenomena of interest are not available for conscious self report. As such it is not likely to be as useful in those tasks characterized by highly automated, as distinct from controlled, processing (Shiffrin & Schneider, 1977). It should be noted, however, that there seems to be no alternative potentially-superior method which allows preconscious cognitive activity in complex dynamic decision contexts to be assessed.

Because of the high-level of cognitive processing involved in the dynamic decision tasks of interest, it is assumed that the relevant mental events, if not at a high level of conscious awareness initially, can be brought into conscious awareness if appropriately cued. Evidence in support of such an assumption can be found in the fact that focused self-reflection improves decision performance in complex dynamic systems (Dörner & Schaub, 1994). What remains to be addressed is the possibility that in attempting to bring such material into conscious awareness, participants feel compelled or constrained to *infer* mental events (rather than merely recall such events) and otherwise 'construct' consistent explanations for observed behaviour (Nisbett & Wilson, 1977). Because head mounted video is a particularly non-directive method for cueing recollections which also maximizes the amount of material that can be brought into conscious awareness, it is most likely to minimize distorted inferential accounts.

Concluding Remarks

The head mounted video cued recall methodology constitutes a valuable methodology for stimulating the recall of psychologically-relevant data in general and for the investigation of error behaviour in particular. Specifically with respect to error detection, the procedure has the dual advantages of (a) directly confronting a participant with his or her errors and error tendencies, while (b) minimizing the self-protective tendencies to deny or distort such personal error. It has been demonstrated to be a robust methodology suitable for use in quite demanding operational, as well as training, contexts. We are currently trialling extension of the methodology to include (a) the integration of eye tracking information with the initial video recording and (b) the real time transmission of head mounted video images by various wireless communication methods, including microwave radio and mobile phone networks. These developments can be expected to further extend the range of contexts in which the use of head mounted video can be applied, and to increase the sensitivity of the methodology for identifying and minimising error.

References

Bartl, C., & Doerner, D. (1998). Sprachlos beim denken--zum einfluss von sprache auf die problemloese- und gedaechtnisleistung bei der bearbeitung eines nicht-sprachlichen problems. / Speechless when thinking--about the role of speech for problem solving.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*(3), 211-241.

Brehmer, B. (2000). Dynamic decision making in command and control. In C. McCann & R. Pigeau (Eds.), *The human in command: Exploring the modern military experience* (pp. 233-248). New York: Plenum.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: John Wiley and Sons.

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass.

Dickson, J., McLennan, J., & Omodei, M. M. (2000). Effects of concurrent verbalisation on a time-critical dynamic decision-making. *Journal of General Psychology*, *127*(2), 217-228.

Dörner, D. (1987). On the difficulties people have in dealing with complexity. In K. Rasmussen & J. Leplat (Eds.), *New technology and human error* (pp. 97-109). Chichester, England: John Wiley & Sons.

Dörner, D. (1990). The logic of failure. *Philosophical transactions of the Royal Society of London, 327*, 463-473.

Dörner, D., & Schaub, H. (1994). Errors in planning and decision-making and the nature of human information processing. *Applied Psychology: An International Review, 43*(4), 433-453.

Dörner, D., & Wearing, A. J. (1995). Complex problem solving: Toward a (computer simulated) theory. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 65-99). Hillsdale, NJ: Lawrence Erlbaum Associates.

Duval, S., & Wicklund, R. A. (1972). *A theory of objective self awareness.* NY: Academic Press.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Special Issue: Situation Awareness. *Human Factors, 37*(1), 32-64.

Funke, J. (1995). Some pathologies in the study of pathologies. *Sprache und Kognition, 14*, 91-95.

Green, R. (1990). Human error on the flight deck. *Philosophical Transactions of the Royal Society of London, 327*, 503-512.

Jansson, A. (1994). Pathologies in dynamic decision making: Consequences or precursors of failure? *Sprache und Kognition, 13*, 160-173.

Kipper, P. (1986). Television camera movement as a source of perceptual information. *Journal of Broadcasting and Electronic Media, 30*, 295-307.

Klein, G. A. (1989). Recognition-primed decisions. In W. Rouse, B. (Ed.), *Advances in Man-Machine Systems Research* (Vol. 5, pp. 47-92). Greenwich, Connecticut: JAI Press.

Lipshitz, R. (1997). Naturalistic decision making perspective on errors. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 151-160). Malwah, NJ: Lawrence Erlbaum.

McLennan, J., Omodei, M. M., Rich, D., & Wearing, A. J. (1997). Helmet-mounted video: Applications for fire officer training and operations. *Journal of the Fire Service College*, *3*, 63-74.

McLennan, J., Omodei, M. M., & Wearing, A. J. (2001). Cognitive processes of first-on-scene fire officers in command at emergency incidents as an analogue of small-unit command in peace support operations. In P. Essens ,A. Vogelaar, E. Tanercan, & D. Winslow (Eds.), *The Human in Command: Peace support operations* (pp. 312-329). Amsterdam: KMA Royal Netherlands Military Academy, Breda.

McLennan, J., Pavlou, O., & Omodei, M. M. (in press). Cognitive control processes discriminate between better versus poorer performance by fire ground commanders. In R. Montgomery, R. Lipshitz, & B. Brehmer (Eds.), *How professionals make decisions*.

McLennan, J. P., & Omodei, M. M. (1996). The role of prepriming in recognition-primed decision making. *Perceptual and Motor Skills, 82*, 1059-1069.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know. Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Norman, D. A. (1998). *The design of everyday things*. London, England: The MIT Press.

Omodei, M. M., McLennan, J., & Whitford, P. (1998). Improving decision making in complex natural settings: Using a head-mounted video camera to improve performance of competitive orienteers. *International Journal of Sport Psychology, 29*, 115-131.

Omodei, M. M., & McLennan, J. P. (1994). Complex decision making in natural settings: Using a head-mounted video camera to study competitive orienteering. *Perceptual and Motor Skills, 79*, 1411-1425.

Omodei, M. M., & Wearing, A. J. (1995). Decision making in complex dynamic settings: A theoretical model incorporating motivation, intention, affect, and cognitive performance. *Sprache and Kognition, 14*(2), 75-90.

Omodei, M. M., Wearing, A. J., & McLennan, J. (1997). Head-mounted video recording: a methodology for studying naturalistic decision making. In R. Flin, M. Strub, E. Salas, & L. Martin (Eds.), *Decision making under stress: emerging themes and applications* (pp. 137-146). Aldershot: Ashgate.

Omodei, M. M., Wearing, A. J., McLennan, J., Elliott, G. C., & Clancy, J. M. (in press). More is Better? Problems of self regulation in naturalistic decision making settings. In B. Brehmer, R. Lipshitz, & H. Montgomery (Eds.), *How professionals make decisions*.

Rasmussen, J. (1990). The role of error in organizing behaviour. *Ergonomics, 33*, 1185-1190.

Reason, J. (1990). *Human error*. New York, NY, USA: Cambridge University Press.

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology, 53*, 491-517.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic information processing. II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Swann, W. B., Griffin, J. J., Predmore, S. C., & Gaines, B. (1987). The cognitive-affective crossfire: When self-consistency confronts self-enhancement.

Woods, D. D., & Cook, R. I. (1999). Perspectives on human error: Hindsight bias and local rationality. In F. Durso (Ed.), *Handbook of applied cognitive psychology* (pp. 141-171). New Yord: Wiley.

Zsambok, C. E., & Klein, G. (Eds.). (1997). *Naturalistic decision making*. Mahwah, NJ, USA: Lawrence Erlbaum Associates Inc.

Tool Support for Scenario-based Functional Allocation

Alistair Sutcliffe, Jae-Eun Shin, Andreas Gregoriades

Centre for HCI Design, Department of Computation
University of Manchester Institute of Science and Technology (UMIST)
P.O.Box 88, Manchester, M60 1QD, UK
a.g.sutcliffe@co.umist.ac.uk

Abstract: Tool support is described for analyzing requirements and creating conceptual models from scenarios. A schema of scenario-based knowledge is proposed that extends the **i*** ontology with concepts to represent the system environment and natural language semantics to categorize arguments. Modelling tools are introduced to support the process of transforming scenarios into models and requirements. We illustrate use of the tools by analysis of the London Ambulance case study. The advisor guides the analyst with contextual appropriate advice on functional allocation of agent roles and generic requirements to avoid errors and system failure. The advice is based on research in human reliability engineering.

Keywords: scenarios, functional allocation, socio-technical systems, human error.

Introduction

Scenarios have received considerable attention as a means of eliciting and validating requirements (Carroll, 1995; Potts, Takahashi & Anton, 1994; Rolland et al., 1998). In requirements elicitation, models are derived from scenarios by a process of generalization, while in requirements validation, scenarios can be used as examples to test a model or requirements specification. However, there are few methods or tools that help the transformation of scenarios to models or support the use of scenarios in requirements validation.

One use of scenarios is to capture information about the system environment (e.g. Kyng, 1995) which is often ignored in conceptual models. Yu and Mylopoulos (Yu, 1997) emphasize the need to model the system environment, since lack of domain knowledge frequently leads to inadequate requirements and hence system failures Curtis, Krasner & Iscoe, 1988). The ***i**** framework (Yu, 1997) was developed for modelling and reasoning about the impact of organizational environments on information systems, and ***i**** does provide reasoning mechanisms for validating relationships between agents, tasks and goals; however, we argue that requirements analysis tools should go further and provide advice on issues such as functional allocation and socio-technical system design. In previous work we investigated taxonomies of influencing factors and proposed scenario-based techniques for diagnosing problems in communication and functional allocation in socio-technical systems (Sutcliffe, 2000; Sutcliffe et al., 1998). To assist this modelling, we introduce tools that support the process of transforming scenarios into models and requirements specifications. These tools are based on schema of scenario-based knowledge, explained in the following section. The tools are illustrated by analysis of the London Ambulance case study.

Knowledge Representation Schema for Scenarios

Scenarios have many definitions and even more diverse content (Carroll, 2000, 1995; Cocchiarella, 1995), so a general purpose ontology of knowledge (Hovy, 2001; Sowa, 2000) might seem to be an appropriate choice. However, we wish to build upon existing conceptual modelling languages (e.g. UML) and ***i**** in particular because this is established in RE. Our schema, therefore, contains concepts that are familiar in many modelling languages (i.e. agents, objects, tasks, goals), but it adds new constructs for modelling the system environment and, more radically, for argument and communication. We propose a unified schema that represents arguments expressed in natural language and the domain of discourse (i.e. the modelled world) to which those arguments pertain. The motivation for this is simple. Scenarios frequently report opinions and causal arguments that explain aspects of the modelled world. Capturing and analyzing such arguments is often critical to discovering accurate requirements.

A schema of scenario components was derived from the review of relevant literature (Carroll, 1995; Carroll et al., 1994; Daren, Harrison & Wright, 2000; Mylopoulos, 1998; Sutcliffe et al., 1998, Mylopoulos, 1998), ontologies and knowledge representation (Chung & Nixon, 1995; Guarino, 1997; Sowa, 2000; Van Heijst, Schreiber & Wielinga, 1997; Waterson & Preese, 1999). The schema categorizes scenario narratives into five areas (*Actors & Structures, Intentions, Tasks, Environment*, and *Communication*) and three levels (*Strategic, Tactical,* and *Operational*).

Semantics to express structures and properties of the system environments and argumentation were drawn from functional theories of language (Mann & Thompson, 1988) to augment the semantics in the ***i**** model (Mylopoulos, 1998). Concepts are first order modelling primitives which have properties. The concepts and relationships in each of the five areas are as follows:

- Actors & structures: agent, attribute, group, organization, physical structure, role; *properties* of agents: motivation, capability, dependability, power, reputation, responsibility, trust
- Intentions: goal, objective, policy, strategy; *properties* of goals: importance, quality
- Activity-related: action, event, object, procedure, resource, state, task; *properties* of tasks: predictability, complexity, criticality
- Environmental: social, economic and physical environments including location; *properties* of environment: predictability, interruptions, weather state, stress, climate, noise; *properties* of social environment: management culture, time pressure, stress, inclusiveness
- Communication: argument, attitude, background context, causation, consequence, decision, elaboration, evidence, issue, interpretation, hypothesis, justification, motivation, position, viewpoint.

The schema concepts and relationships are shown in Figure 1. The actors, intentions, task and environment components all represent the modelled world and are connected to communication components by user-defined relationships. The communication area is not explicitly coupled to the modelled world because it represents arguments about the domain of discourse. In many cases, segments in scenario narratives may refer to argument and to properties of concepts in the modelled domain; for instance, a narrative that argues for system users being well trained can be described by properties of agents (their motivation) as well as given a motivating argument for their training.

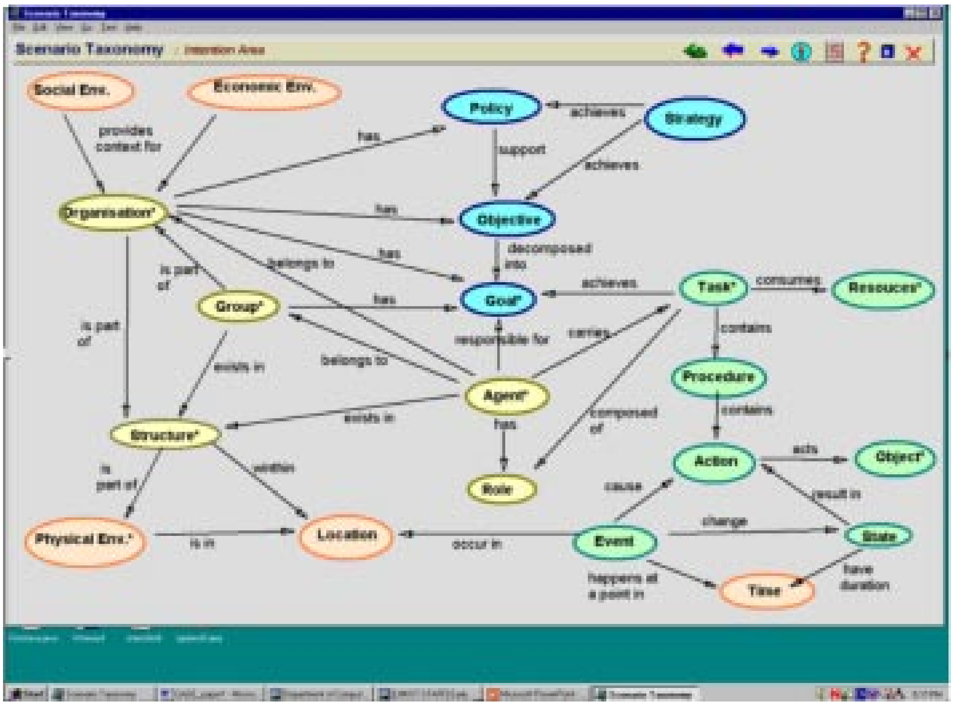


Figure 1 - Hypertext tool showing schema map interface with domain model components and relationships at the tactical and strategic level.

Hypertext Tool for Scenario Management: An object-oriented authoring system (Asymetrix's Tool-Book Instructor II), a hypertext tool, was used to construct a concept map interface of the scenario schema, as illustrated in Figure 1. The user could access definitions as a pop up "tool tip" text to explain each component with examples and synonyms to help understanding.

Scenarios are marked up using the annotator editor which also functions as a model editor so that model components can be linked to scenario narrative segments. We illustrate use of the scenario modelling tools with the computer-aided dispatch (CAD) system in the London Ambulance Service (LAS). The aim is to show how a scenario modelling tool can be used in requirements analysis to identify key design issues by a retrospective analysis of the LAS CAD system starting with a narrative that presents a scenario-like summary of the system failure. (Finkelstein & Dowell, 1996)

The annotation editor provides a direct manipulation interface so the user can highlight a segment of scenario text and then point to the schema component that describes it. This causes markup tags to be placed in the selected text, *e.g. <agent> text editor </agent>*. Text can be selected several times so part of a scenario narrative can be linked to communication as well as domain model components, *e.g. <justify> to support the user's task it was necessary to create a tool which was the <agent> text editor </agent></justify>*.

The annotation editor could also be used to create models by a simple pick and place dialogue. This allows the user to create a model of the domain with agent, task, object instances and then link the model components to their corresponding origin in one or more scenarios. The LAS system model derived from this analysis is shown in Figure 2. Links in the annotation editor provide traceability so cause-consequence arguments in the scenario can be traced to the relevant system model components. For example, the frustration experienced by the ambulance crews which led them to poor reporting of call status relates to the relationship between the Crew agent and the Reporting goal/task.

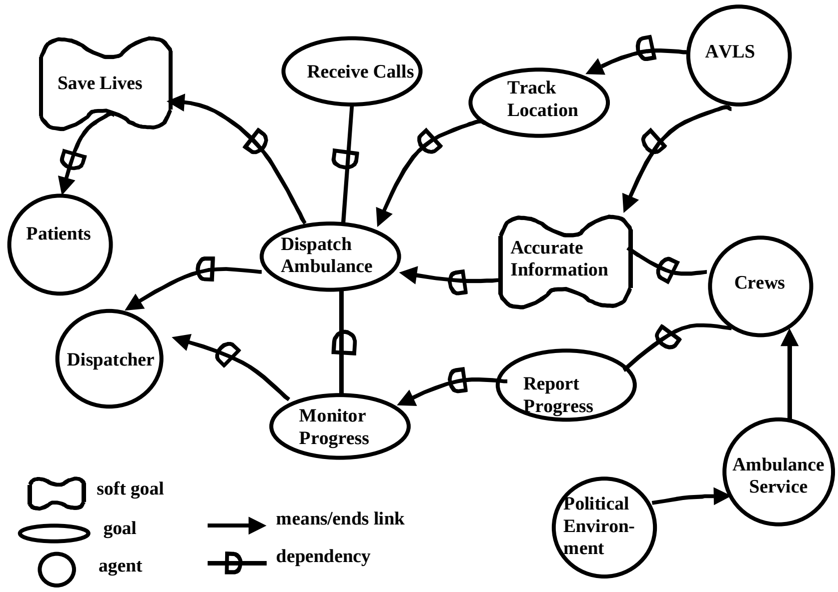


Figure 2 - Model of the LAS system in adapted ***i**** notation. Only goals and agents are shown for simplicity, with additions of organization and environment components taken from our scenario schema. Causal influences on the crews are modelled as means-ends links.

Scenario Analysis Advisor: The scenario analysis advisor uses rules and relationships between schema components to produce advice on human factor problems and generic requirements that indicate solutions to those problems. Three types of advice are available:

Functional allocation issues: this advice concerns the trade-off decisions about which functional requirements should be fully automated, partially automated or left as manual tasks. This advice is accessed when querying goal/task or agent components in models or scenarios. The knowledge is drawn from the HCI literature (Bailey, 1982; Wright, Dearden & Fields, 2000) and contains warnings about flexibility of processes, the predictability of events, workload estimation techniques and social issues such as human reactions to changes in responsibility and authority brought about by automation.

System reliability and human error: this advice pertains particularly to task/goal-agent relationships but it may also be accessed via organization-agent relationships. The knowledge for this is drawn from the human reliability engineering literature (Hollnagel, 1998; Leveson, 1995; Reason, 2000) and covers typical errors that may occur in certain task-agent combinations with generic requirements to prevent or contain such problems.

General analysis information about socio-technical system problems attached to relations between agents, tasks, goals, organizations, and the social/political environment. This knowledge is taken from the RE and studies of socio-technical systems.

The advice is accessed via browsing on the schema map and selecting components. Alternatively, nodes in marked-up scenarios can be selected to access advice dialogues. To illustrate the process, assuming the user has selected four nodes: agent, task, structure and physical environment; first the system requests further information about the properties of the selected nodes as shown in Figure 3. For instance the task properties are entered as high or low estimates for the level of complexity, the level of training and familiarity with the task, the agent types (e.g. human or machine agent), and the environmental properties (e.g. interruptions, weather conditions). This information is used by the system to narrow its search for appropriate advice.

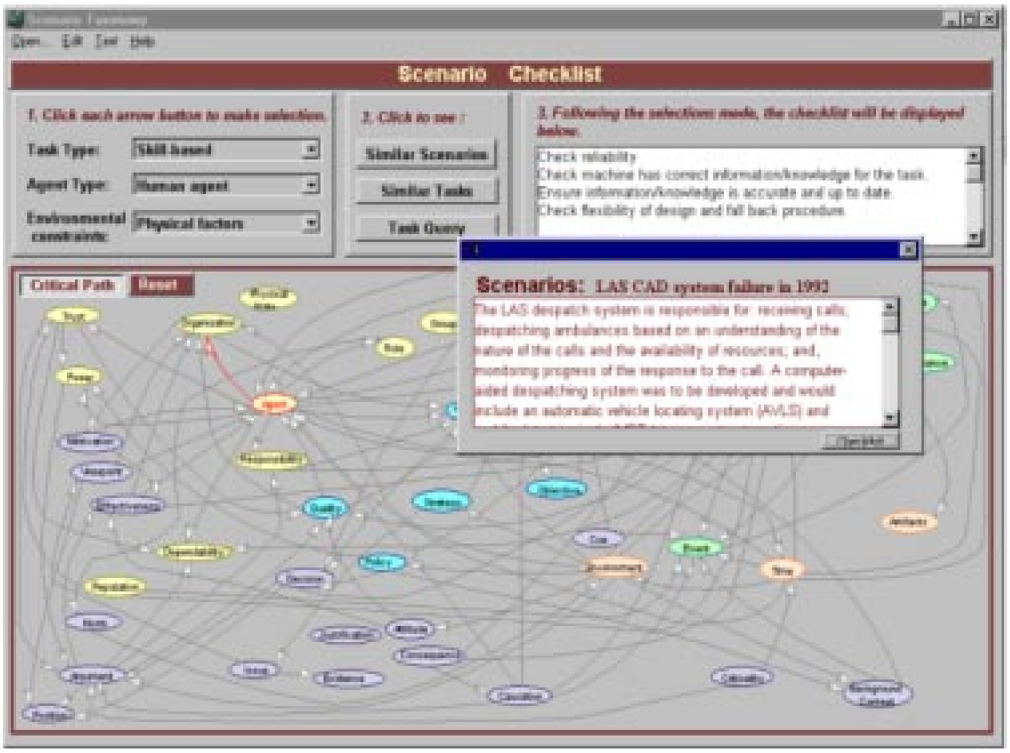


Figure 3 - Scenario analysis advisor, showing input of properties for the selected relationship (1. top left) and advice (3. top right hand message box).

If functional allocation advice is chosen with agent and task nodes selected then the system provides the advice illustrated in Table 1 which shows the setting of the properties of the schema components, and the advice and its justification with respect to those settings.
Rules indicate potential constraints on agent-task combination such as the dangers of allocating complex tasks to poorly motivated agents. If the organization node that the agents belong to and properties of management culture are set to poor, this will set the agents' motivation low, so the influence of property settings is propagated along schema relationships. Table 2 shows error prevention advice for the task-agent relationship.
Advice is created by rules that follow the schema links from task and agent to structures and organization (the immediate system environment) and then to physical and social environment nodes. The error advice database is organized using the schema to enable access to the appropriate information. The advice is generated by rules that link the preconditions to types of error. A sample of the rules is given below, with the general format followed by an example:

- Functional allocation rules:
  If task <property=H/L> *Then* allocate to <Machine or Human or Collaborative (machine support)>: e.g. *If* task <complexity=L> *Then* allocate to <Machine agent>
  If agent <property = H/L> *Then* allocate to <Machine or Human (training advice)>: e.g. *If* agent <capability = H> *Then* allocate to < Human agent>

- Error reliability rules:*If* organisation <property = H/L> *Then* agent <property = H/L>: e.g. *If* organisation <incentive = L> *Then* agent <motivation = L>
  If agent <property = H/L> *Then* errors <(slips/mistakes) probable/not probable>: e.g. *If* agent <motivation = L> *Then* slips are likely
  If physical environment <property = H/L> *Then* errors <(slips/mistakes) probable/not probable>: e.g. *If* physical environment <time pressure = H> *Then* slips are likely.

The number of rules that trigger error predictions is counted to increase a confidence rating, e.g. slips are likely (influencing factors = 3/8).

Table 1 - Functional allocation advice for Tasks and Agents according to their property settings.

Component	Properties	if High, then Implications	if Low, then Implications
Task	Complexity	Capable and well trained operators; allocate to humans	Little training, suitable for automation
	Predictability	Automate, if not too complex	Allocate to humans
	Importance/criticality	Motivate operators, back-up, recovery and fail safe design	Less training and error prevention needed
Agent	Motivation	Allocate demanding tasks to humans	Manual allocation for non-critical simpler tasks
	Capability	Check time for complex tasks Human operation	Automate for simple, predictable tasks
	Task knowledge	Skilled tasks Human operation	Decision support; training necessary
	Dependability	Allocate critical tasks to humans, or automate	Automate; Humans for simpler, non-critical tasks

Table 2 - System and human reliability advice for task-agent relationships.

Component	Properties	if High, then Implications	if Low, then Implications
Task	Complexity	Mistakes unless operators are well trained	Slips when operators become bored
	Predictability	Slips in routine operation; beware time pressure	Mistakes in reacting to usual events; training and simulation help
	Importance/criticality	Time pressure, fatigue and stress cause errors	Slips unless well motivated
Agent	Motivation	Mistakes less likely, slips still occur	Prone to mistakes and slips
	Capability	Fewer errors if well trained	Errors unless trained and given simple tasks
	Dependability	Errors less likely unless time pressure, tired or stressed	Prone to mistakes, lapses and slips

The error advice points to high likelihood of mistakes which are errors in intention and planning, while slips and lapses are failures of attention and concentration. Slips and lapses occur in skilled operation when the user agent is familiar with the task, whereas mistakes occur more frequently in tasks that require more judgment and decision making. Time pressure, fatigue and stress increase error rates, even when agents are well motivated, capable and reliable (Reason, 1990).

Case Study: London Ambulance Service

The LAS scenario was investigated using the analysis advisor. First, a general checklist is provided followed by more specific advice on critical design issues through understanding relationships between them. The task type for “Dispatch Ambulances” is set to *Critical = H*, *Complexity = H* and *Predictability = M*. This task requires judgment and knowledge about identifying calls and ambulance resource availability.

Since full automation of the dispatch task was planned, the agent type is set to machine. For the environment properties, time pressure and interruptions were set to high while the other environmental factors like the weather could also be poor, so the following advice checklist was displayed:

- Function allocation advice is:
  Check reliability.
  Check machine has correct information/knowledge for the task.
  Ensure information/knowledge is accurate and up to date.
  Check flexibility of automated design and fallback procedures.
  With implications for the human role:
- Check change in responsibility is acceptable to the user.
  Investigate the assignment of authority of agents for tasks/goals.
  Question the impact on users' motivation and morale of changes in responsibility and authority.
  Investigate users' trust in technology; if it is poor, then operation may be ineffective.
- And implications of physical environment for task effectiveness:
  Investigate time for decision making.
  Check for interruptions and flexibility in handling unexpected events.
  Investigate weather impact on task operation.

Clearly such advice was not followed during the development of the CAD system, so it is possible that provision of this knowledge might have prevented some of the design mistakes. The CAD system did not have the correct information in the gazetteer of London streets and the location of the ambulances was inaccurate because radio blackspots prevented the system from tracking the vehicles in some locations. Furthermore, the information on call progress was inaccurate because of the crews' failure to report calls via mobile data terminals (MDTs). Implementation of the system changed the responsibility and authority of the dispatcher controllers because the system made the choices for them, with little opportunity to override decisions. This created an inflexible system. The motivation and morale of the dispatchers was probably impacted before the system went live, but rapidly became worse when the unreliability of the system became obvious.

In the second example, the report task and ambulance crew node is selected with advice on potential system errors and human reliability. This delivers the following guidance organized in four areas where errors may occur: the human agent (i.e. the ambulance crew users), design of the computer system (in this case the mobile data terminal) and the environment in which the system is used. The task in this case involved the crews entering progress reports into the MDTs. The properties settings of the task and environment were:

Task: Critical = High, Complexity = L, Skill = H and Predictability = M
Environment: Time pressure = H, Stress = H, Predictability = L

The task is critical because accuracy of the CAD system databases depends on it. Although the task of reporting in itself is not complex, its predictability can vary as some calls do not go to plan and the time pressure is created by crews having to attend to a higher priority task first, such as giving first aid, and getting the patient to hospital. The task is assumed to be a trained skill. The analysis advice is accompanied by generic requirements as follows:

- Human Error: slips and lapses are likely with skilled tasks. Check that training is adequate, and that the skill has been practised recently.
  generic requirements: to prevent/remedy lapses, use timeouts to check progress, provide reminders, status indicators, keep routines short; to prevent/remedy slips, trap slips with validation routines, minimize distractions, make task operations clear, ensure objects acted upon cannot be confused with others, provide reminders, undo facilities, editing facilities to correct errors.
- User Interface Design
  generic requirements: predictable and consistent user interface, same layout of screens, consistent commands, simple actions, clear and unambiguous feedback, ergonomic requirements of visibility, audibility of output.
- Environment influences: slips and lapses
  generic requirements: minimize distractions and interruptions, e.g. non-essential noise, extraneous visual stimuli, non-essential communication. Ensure user has sufficient time to complete task without being pressured. Minimize fatigue and stress, which adversely affect user concentration. Investigate user motivation.
- Social/political environment
  generic requirements: management culture should motivate users to complete tasks by encouragement and incentives. Goals and standards of performance should be clearly communicated and addressed by training. Users should feel that they own the system and are involved with its success. Avoid authoritarian management styles if possible.

In this case the advice was pertinent to the LAS system at both levels. There were several user interface design defects in the MDT terminals which made them difficult to use, such as the order of entering call progress and poor visibility of the displays. The crews didn't use their MDTs effectively because of motivational problems caused by a poor managerial culture which did not involve crews in the design of the system. Furthermore, no incentives were given for effective use, and training was inadequate. Finally, when failures began to accumulate in the system, crews became stressed and tired which led to more slips and errors. Careful design was necessary because the crews' working environment was prone to interruptions and time pressures so error prevention should have been built into the user interface; for instance, by making the reporting sequence clear. This last point is not made explicit in the LAS report; however, it can be inferred as another contributing factor from the above checklist.

Lessons Learned

The scenario annotator/advisor is currently a prototype/concept demonstrator which we created to gain feedback on the suitability of developing this approach. The feedback obtained so far from demonstrations of the system to industrial users has been reasonably encouraging; however, several problems have emerged. Firstly, the advice often requires human factors knowledge to interpret it. Our reaction to this problem is twofold. Firstly, we intended the system to be used by software engineers who have received at least some HCI training, and secondly to make the advice easier to understand, although this will make it more verbose. The second problem was anticipated from the outset: that marking up scenarios is a labour-intensive task which leads to the question about whether the annotation and traceability between scenarios and models will provide sufficient added value for the effort. As yet we have no answer to this point; however, to persuade industrial users to try out the system, and hence allow us to capture effectiveness data, the next step is to add an information extraction tool to partially automate the markup. Information extraction tools work by being trained to recognize text segments using rules that combine domain specific vocabularies with discourses marker phrases; e.g. *because, as a result*, etc., point to *cause-consequence* components. Another direction is to restrict markup by only documenting parts of a scenario narrative that relate to important requirements. Other problems are concerning the readability of the complex schema graphs and understanding their semantics, although these problems were alleviated by follow-up explanation, so an explanation facility is another extension. The concept of integrating communication/argumentation and the modelled domain was considered worthwhile as documentation on scenarios design discussion and models tended to be kept separately, making traceability difficult. Reaction to the system advice was most favourable overall, although the users pointed out that this could be driven directly from the schema graph without the scenarios.

Conclusions

The contribution that the scenario advisor tool has made so far is to explore the feasibility of tool support for eliciting conceptual models by generalization from scenarios and delivering advice on human factors issues in system development. The focus on human error was motivated by our previous work modelling the causes of system failure and human error using Bayesian Belief Networks (Galliers, Sutcliffe & Minocha, 1999; Sutcliffe, 1993). BBN models enable error probabilities to be predicted for system operators and software components by running system models against scenarios describing the environment, agents and task. However, BBN models hide the knowledge that motivated their construction, so in validation studies users requested more explicit representation of that knowledge. We have reverse engineered the knowledge out of the BBN to make it available as a checklist. One future test of the advisor prototype is to try it in combination with the BBN tool. The advice contained in the current systems is preliminary and will be improved by tailoring it with more domain-specific evidence; however, our initial intention was to evaluate the feasibility of tool-based assistance for functional allocation.

The second contribution of this work is to propose a role for model-driven advice in computer aided systems engineering. Our source of advice in the safety critical systems and human factors literature (Bailey, 1982; Hollnagel, 1998; Reason, 2000) needs to be imported into mainstream system development since many requirements failures, which the LAS system illustrates, could be prevented by more systematic analysis of functional allocation and potential causes of error. Furthermore, we believe that embedding such advice in model editors allows it to be delivered in the appropriate context during modelling activity. Further validation tests of our existing preliminary prototype are the next step to assess the utility and effectiveness of a scenario annotator/advisor tool.

Acknowledgements

This research was supported by EPSRC Systems Integration Programme SIMP project (Systems Integration for Major Projects). Special thanks to David Corrall in BAE Systems for his valuable comments and help with this research.

References

Bailey, R.W. (1982). *Human Performance Engineering: A Guide for System Designers.* Prentice Hall, Englewood Cliffs NJ.

Carroll, J.M.(1995). *Scenario-based Design: Envisioning Work and Technology in System Development.* John Wiley, New York.

Carroll, J.M. (2000). *Making Use: Scenario-based Design of Human-computer Interactions*. MIT Press, Cambridge, MA.

Carroll, J.M., Mack, R.L., Robertson, S.P, and Rosson, M.B. (1994). Binding Objects to Scenarios of Use. *International Journal of Human-Computer Studies* 41:243-276.

Chung, L., and Nixon, B.A. (1995). Dealing with Non-Functional Requirements: Three Experimental Studies of a Process-Oriented Approach. In *Proceedings of the 17th International Conference on Systems Engineering*. IEEE Computer Society Press, Los Alamitos, CA. 25-37.

Cocchiarella, N.B. (1995). Knowledge Representation in Conceptual Realism. *International Journal of Human-Computer Studies* 43:697-721.

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.

Daren, A, Harrison, M. and Wright, R. (2000). Allocation of Function: Scenarios, Context and the Economics of Effort. *International Journal of Human-Computer Studies*,. 52: 289-318.

Finkelstein, A., and Dowell, J. (1996). A Comedy of Errors: the London Ambulance Service Case Study. In *Proceedings of the 8th International Workshop on Software Specification & Design IWSSD-8*, IEEE Computer Society Press, Los Alamitos, CA. 2-4.

Galliers, J.R; Sutcliffe, A.G; and Minocha, S. (1999). An Impact Analysis Method for Safety-critical User Interface Design. *ACM Transactions on Computer-Human Interaction* 6:341-369.

Guarino, N. (1997). Understanding, Building and Using Ontologies. *International Journal of Human-Computer Studies* 6**:**293-310.

Hollnagel, E. (1998). *Cognitive Reliability and Error Analysis Method: CREAM.* Elsevier, Amsterdam.

Hovy, E.H. (2001). Comparing Sets of Semantic Relations in Ontologies. In R. Green and S.H. Myaeng (Eds) *Semantics of Relationships*.

Kyng, M. (1995). Creating context for design. In J.M. Carroll (Ed.) *Scenario-based Design*. Wiley, New York. 85-108.

Leveson, N.G. (1995). *Safeware: System Safety and Computers.* Addison Wesley, Reading, MA.

Mann, W., and Thompson, S. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization *Text* 8:243-281

Mylopoulos, J. (1998). Information Modelling in the Time of the Revolution. *Information Systems* 23:127-155.

Potts, C., Takahashi, K, and Anton, A.I. (1994). Inquiry-based Requirements Analysis. *IEEE Software* 11:21-32.

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

Reason, J. (2000). *Managing the Risks of Organizational Accidents*. Ashgate, London.

Rolland, C. Arhur, B.C., Cauvel, C., Ralyte, J., Sutcliffe, A.G., Maiden, N., Jarke, M., Haumer, P., Pohl, K., Dubois, E., and Heymans, P. (1998). A Proposal for a Scenario Classification Framework. *Requirements Engineering* 3:23-47.

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

Sutcliffe, A.G. (2000). Requirements Analysis for Socio-technical System Design. *Information Systems* 23:213-233.

Sutcliffe, A.G., Maiden, N.A.M., Minocha, S., and Manuel, D. (1998). Supporting Scenario-based Requirements Engineering. *IEEE Transactions on Software Engineering* 24:1072-1088.

Sutcliffe, A.G. (1993). *Modelling Business Goals…: report HCID/93/10.* City University, London.

Van Heijst, G., Schreiber, A.T.., and Wielinga, B.J. (1997). Using Explicit Ontologies in KBS Development. *International Journal of Human-Computer Studies*. 45:183-292.

Waterson, A., and Preece, A. (1999). Verifying Ontological Commitment in Knowledge-based Systems. *Knowledge-based Systems* 12:45-54.

Wright, P.C., Dearden, A.M., and Fields, R. (1997). Function Allocation: A Perspective from Studies of Work Practice. *International Journal of Human-Computer Studies* 52:335-356.

Yu, E. (1997). Towards Modelling and Reasoning Support for Early-Phase Requirements Engineering. In *Proceedings of the 3rd IEEE Int. Symposium on Requirements Engineering*. IEEE Computer Society Press, Los Alamitos, CA. 226-235.

Time-Related Trade-Offs in Dynamic Function Scheduling

Michael Hildebrandt and Michael Harrison

Department of Computer Science
University of York, York YO10 5DD, UK
Tel: +44-1904-433376, Fax: +44-1904-432767
{Michael.Hildebrandt, Michael.Harrison}@cs.york.ac.uk

Abstract: A possible route to managing workload peaks facilitated by advances in technology is to use Dynamic Function Allocation, in other words to design work so that it is possible to switch adaptively between levels of automation. In the main, current approaches to Dynamic Function Allocation assume that functions are to be serviced as soon as possible, and in order of arrival. These methods utilise online allocation decisions along the human-automation resource dimension. Dynamic Function Scheduling takes a different approach and considers the organisation of functions along a joint human-automation timeline using scheduling mechanisms developed for real-time embedded systems. This paper highlights the limitations of Dynamic Function Allocation as currently considered and argues for the introduction of a temporal dimension to work design. Time-related trade-offs faced by the system designer (e.g. flexibility vs. simplicity) and the operator (e.g. value-based scheduling decisions) are discussed.

Keywords: Work design, automation, time.

Introduction

Time is an ubiquitous and often inconspicuous property of physical and psychological processes. At almost every level of granularity, temporal structures can be identified. "Time is nature's way of keeping everything from happening at once", as Woody Allen put it. However, although processes necessarily unfold in time, this is not of itself a property of primary scientific interest. Indeed, many disciplines adopt a Newtonian view and treat time as a background variable that "flows equably, without relation to anything external." This is true in both psychology and computer science, though notable exceptions can be found. Computer science, despite the strong influence of non-temporal logics and computational theory, is also concerned with designing systems that can adapt reliably to the temporal contingencies and requirements of the environment. This focus has resulted in useful models for scheduling concurrent tasks under conditions of scarce processing resources. In human factors engineering, queuing models (e.g. Walden and Rouse, 1978) have been used to address similar problems. In psychology, time perception was an issue for many early researchers such as Wilhelm Wundt and William James. Interest in time subsided when psychology adopted the information processing paradigm and state transition models from Artificial Intelligence, where temporality is reduced to pure sequence. Recent years have seen a revival in the psychology of time, and research is now going beyond the traditional interest in the psychophysics of time to cognitive models of temporal memory, temporal perspective and time as information. Few attempts have been made at unifying the diverse notions of time across different disciplines. Exceptions are Fraser's (1978) model of 'temporalities' and, with a more socio-psychological focus, Doob's (1971) 'taxonomy of time'.

Time and work: Psychological aspects of time in human factors are often reduced to problems of reaction times and the duration of elementary actions and cognitive operations. While time is fairly well understood and modelled at this fine-grained level of behaviour (e.g. the 'Keystroke-Level Model', Card, Moran and Newell, 1980), many temporal phenomena on a wider temporal horizon are still elusive. Advances in the cognitive psychology of time (see for instance Block, 1990; Friedman, 1990; Macar, Pouthas and Friedman, 1992; Michon and Jackson, 1985; Prabhu, Drury and Sharit, 1997; Roeckelein, 2000) have triggered a new interest in temporal issues in human factors (for instance, Decortis, De Keyser, Cacciabue and Volta, 1991; De Keyser, 1995; De Keyser, Ydevalle and Vandierendonck, 1998; Grosjean and Terrier, 1999; Hollnagel, 1991, 2001; Svenson and Maule, 1993). These studies are concerned with temporal awareness and anticipation, temporal planning and control, temporal errors, and decision making under time stress. Despite this progress in human factors, the work design and automation literature has so far given little consideration to temporal organisation.

It is important to emphasise that this line of research is not following a Taylorist agenda – we are not proposing a return to time-and-motion studies. On the contrary, where Taylorism sees the operator as a mainly reactive, event-driven agent who has to adapt to the rhythm of the system, our interest is in the operator's active shaping of the joint human-automation timeline. Instead of breaking work down into elementary, disconnected units, this approach aims at understanding behavioural integration on the operator's temporal horizon.

Structure of the paper: The next section introduces Dynamic Function Allocation, a work design concept, and discusses some unresolved issues and limitations of this approach, relating both to system design and operation. To provide a temporal perspective on work design, an outline of the Dynamic Function Scheduling approach (Hildebrandt and Harrison, 2002) is presented. Time-related trade-offs in system design and operations are discussed.

Dynamic Function Allocation

One of the defining features of modern work is its dynamism. Processes unfold rapidly and sometimes in unexpected ways, resource constraints have to be accommodated online, actions have to be synchronised and coordinated, information needs to be updated and distributed, plans have to be revised and adapted. Automation, introduced to help the human operator handle this complexity, can produce new problems by removing the operator from the control loop and leaving him/her unaware of the state of the system in case of a failure. To address the problems of all-or-nothing automation and static Function Allocation methods, where a level of automation is selected at the design stage, Dynamic Function Allocation (sometimes also called 'Adaptive Automation') provides systems with multiple levels of automation, and decision rules to switch between them at runtime (see Scerbo, 1996, for an overview). Empirical evaluations, mostly based on microworld simulations of production line tasks, air traffic control or aviation scenarios, suggest significant improvements in situation awareness, handling of faults and workload peaks, and overall productivity (e.g. Endsley and Kaber, 1999; Moray, Inagaki and Itoh, 2000; Parasuraman, 1993; Walden and Rouse, 1978; Rencken and Durrant-Whyte, 1993; Tattersall and Morgan, 1997).

The Dynamic Function Allocation literature is diverse. Studies differ in the problems they address (mainly workload and situation awareness), the control over level-of-automation switches (human-initiated, automation-initiated, or comparisons of distinct blocks of trials under different automation levels), the levels of automation provided (full automation vs. full human control or automation scale), and the decision rule used to switch between them (human-initiated, critical event logics, workload- or model-based logics). Despite the multitude of empirical basic research, these approaches have not yet been translated into a unified, mature design method (see Hancock and Scallen, 1998, for some recommendations). As few Adaptive Automation systems are available outside the aviation domain (e.g. Morrison, 1993), the long-term benefits and problems of this approach are as yet difficult to assess. The following two sub-sections discuss a number of unresolved issues relating both to the design and operations of Adaptive Automation systems.

Design considerations: To be more adaptive than all-or-nothing automation approaches, Dynamic Function Allocation provides a number of different levels of automation for a given system. For instance, Sheridan's (1981) widely cited automation scale, which applies most readily to information processing and problem solving purposes, comprises 10 distinct levels (for a more recent scale, see Endsley & Kaber, 1999). Implementing this diversity is likely to be a major challenge. Not only must the designer develop and test a variety of different solutions for the same function, but also provide a sensitive and reliable decision logic, which might involve workload and context measures. The costs and benefits of this development effort are not currently discussed, and it is unclear how easily the current scales can be adapted to a variety of application domains.

Current research in this area tends to assess the effects of Adaptive Automation for single, isolated functions. In these studies, the relevant aspect of the automation decision is the effect on workload and situation awareness, and not the potential, more specific implications for the servicing of other functions. Even when multi-task paradigms are used, the functions are often not strongly causally related. However, as functions in modern socio-technical systems are usually highly inter-connected, the effects of a mode change in one function might have significant implications for a whole network of other functions. The requirements of the specific problem or problem-solving strategy might be a much stronger constraint on the automation decision than workload reduction and maintaining situation awareness (see next section). Before Dynamic Function Allocation can develop into a mature work design method, it has to be able to take account of the inter-dependencies of functions and the contexts in which they might occur (see Harrison, Johnson and Wright, 2002, for an example of such an approach in static Function Allocation).

The only option for workload balancing in Dynamic Function Allocation is automation – 'Dynamic' here refers to a decision on the resource axis, not on the timeline. In so far as the decision is based on performance data or critical events, the method has a temporal element, but it often takes into account only a narrow, retrospective temporal window around the decision point. As the specific effects of the allocation decision for the future timeline are not usually considered, this approach can be characterised as 'snapshot allocation'. However, understanding workload in practice will need to allow considerations of the operator's pro-active, future oriented behaviour.

Operator considerations: Among the primary concerns for Dynamic Function Allocation methods is the loss of situation awareness. As this phenomenon can occur under long periods of automation, Parasuraman (1993) suggested that automation levels should switch periodically, even without being triggered by critical workloads, to keep the operator in the control loop. However, this approach could be problematic if the manual operation cycles of various different functions are not well synchronised, creating the risk of task interference. Confusion can also be caused if automation levels switch too quickly and frequently as a result of insufficient inertia in the decision logic, or if the decisions are intransparent to the operator.
Another critical issue in Dynamic Function Allocation is complacency or over-reliance on automation (Parasuraman, Molloy and Singh, 1993). Especially if the automation is fairly reliable (but still not perfect), operators could be lulled into a false sense of security and thereby neglect their supervisory duties. The authors suggest that vigilance could be encouraged by "simulat[ing] a variable-reliability system by including (at variable intervals) artificial failures that would require an operator response". On the other hand, some studies (Harris, Hancock and Arthur, 1993; Tattersall and Morgan, 1997) have documented human failure to engage automation even when available. More specifically, Harris et al. report that fatigued participants failed to use automation, even though they might benefit most from automatic support. Unfortunately, as with most studies in this field, these papers report summary results and not individual strategies, making it difficult to generate explanations for these results.
A problem of Dynamic Function Allocation, especially if allocation shifts are to be triggered by the human operator, is the added processing demand induced by the decision process (note that a similar problem occurs in real-time systems, where there is a trade-off between more sophisticated and effective scheduling algorithms and the processing time required to execute them). This problem becomes aggravated the more adaptivity a system provides, as more levels of automation have to be considered. Thus, a compromise has to be found between flexibility and simplicity. The computational complexity can be reduced when automation levels for different functions are not seen as independent of each other, but instead as bound up into automation configurations, with each configuration appropriate for a certain operation scenario. This perspective, seeing automation in the context of strategy choice, is not strongly developed in current approaches.
Most current Dynamic Function Allocation concepts assume or require that all available levels of automation provide equal quality of solution, so that re-allocation decision can be based purely on the required workload reduction. While this assumption is feasible for some isolated automation scenarios, under a more naturalistic perspective, function servicing strategies often involve satisficing decisions and trade-offs. For instance, in a medical context, expert systems could be used by more junior staff as part of a backup strategy if advice from senior staff is unavailable. Similarly, unavailability of automatic medical equipment such as blood gas monitors or ventilators might require higher manual involvement, even though the quality of this treatment may be lower. In fault analysis, different levels of data integration (e.g. high integration with decision support or access to raw data) will be chosen according to the cognitive strategy of the operator, not necessarily for the workload reduction they provide.

Dynamic Function Scheduling
Dynamic Function Scheduling (Hildebrandt and Harrison, 2002) brings a temporal perspective to workload-related problems in high-consequence systems, and also aims at understanding and designing a broader range of scheduling and satisficing phenomena in normal operations. It considers allocation along the joint human-automation timeline as a strategy in multi-task servicing (Fig. 1). In this sense it goes further than the automation option considered in Dynamic Function Allocation. In addition to asking *who* should perform a function, it asks *when* and *if* a function should be performed, taking into account the agents' current and predicted workload, available resources, service rates, and the configuration of other functions on the joint timeline. Scheduling options include postponing, swapping and dropping of functions. For instance, Hildebrandt and Harrison (2002) discuss a fault servicing scenario for an aviation hydraulics system and identify conditions where different scheduling strategies are appropriate (diagnose fault first, then fix it; switch to redundant circuit, then diagnose fault; drop function, i.e. ignore problem, if leak will not become critical before touch-down). Arguing that scheduling is an ubiquitous problem, the authors also discuss a supermarket checkout scenario, where both function allocation and scheduling can be observed: if a customer cannot pack the items quickly enough, the cashier will often switch from his/her primary function of scanning the items to assisting the customer in packing in order to optimise overall throughput. From an allocation perspective, part of the packing function has been re-distributed to the

cashier. From a scheduling perspective, the operator has postponed the primary function (scanning) to increase performance of the joint packing function (note that this decision could be context dependent: if the cashier is fatigued, the delay in packing could provide a welcome break). A combination of scheduling and allocation is characteristic of most multi-agent systems.

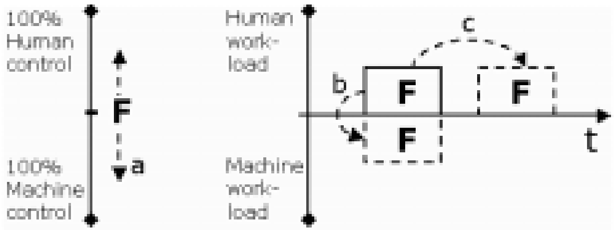


Figure 1. Conceptual differences: Dynamic Function Allocation (left) allocates on the resource dimension (a). Dynamic Function Scheduling (right) allocates on the resource (b) and/or the temporal dimension (c).

Value-based function scheduling / strategy selection: Dynamic Function Scheduling considers both temporal and quality-related aspects ('value') of a function, and considers the trade-offs involved in trying to accommodate concurrent functions in a given time frame. To address some of the limitations of current Dynamic Function Allocation, the approach distinguishes between functions and the strategies available for servicing a function. This results in two different notions of value: one is a measure of the contribution a function makes to the overall system objectives and is used in planning, i.e. to prioritise and order concurrent functions by comparing their values (for example, in aviation the highest priority is given to safety-related functions, followed by passenger comfort and economy). In the above example, the value of assisting in packing becomes greater than the value of continuing scanning when items pile up. The other notion of value is a measure of the quality of solution a particular strategy (possibly involving a certain level of automation) provides in servicing a certain function. It is used to select among the different strategies available for servicing a function. Seeing the hydraulics example as a case of strategy selection (though it also involves scheduling), the decision to 'diagnose first, fix second', 'fix first, diagnose later' or 'drop function' will depend on the utility of obtaining a closer diagnosis, the time required for the diagnosis, the time available for fixing the problem, current workload, and the stage of the mission. Though these computations can, in theory, become very complex, most expert operators will have developed efficient heuristics and decision rules to assess the dynamics of the problem and resolve speed-quality trade-offs in strategy selection (e.g. Amalberti and Deblon, 1992).

Both notions of value are closely related; a lower-value, but faster, strategy may have to be selected if there is insufficient time (or resources) for executing the higher-value, but slower, strategy by the function's deadline. The quality of the selected strategy will, in turn, affect the value of the function itself. To reason about such relations, it is useful to introduce the notion of urgency, which can be obtained by relating the time required and the time available for servicing a function or executing a strategy. The urgency approaches 1 as the function gets closer to its deadline. If the ratio exceeds 1, the function cannot be serviced in time, and might have to be dropped.

A further dimension is added by assuming that values change over time. The value of servicing a function may be lower when the function is far from its deadline than when it is very close to it. Similarly, a strategy that requires a shorter execution time than an alternative strategy will have a higher relative value when the deadline is close than when the deadline is still a long time away. When applied to actual work situations the concept of value will have to be extended to represent dynamic changes over time and to allow for linear or non-linear value functions. It will also be necessary to integrate the notions of value and utility in the psychological literature on judgment and decision making.

System design trade-offs: For the designer, the main challenge related to Dynamic Function Scheduling is in deciding on the sequential flexibility or rigidity of the functions in the system. The order in which functions should be serviced can be constrained by their physical and logical nature (e.g. lowering the landing gear and landing), or by requirements and limitations of the human operator (e.g. biases in temporal reasoning or tendency to omit actions and confuse sequence in high workload situations). In many high-consequence domains such as aviation and power plant control, there is a need to provide rigid sequentialisation in the form of checklist procedures to avoid omissions and to ensure correct ordering. In other situations, procedural diversity might be necessary to operate in a dynamic environment. Flexibility is also necessary if the operator has to find solutions to unforeseen failures. Thus a compromise has to be found between the risk and the diversity provided by flexible temporal organisation. In terms of the hydraulics example mentioned above, this would involve analysing the benefits of the different strategies (diagnose-fix, fix-diagnose, drop) in different scenarios, considering the operator's decision effort for matching a strategy to a situation, and possibly considering a redesign of the function (using automation) and the physical system.

The designer should also be aware of the overall temporal properties of the system. This includes the assessment of the expected function arrival rates, temporal properties of the functions (e.g. continuous, periodic, sporadic), service rates for human and automation, and the ability of the combined system to accommodate unexpected events on the timeline. This temporal inventory of the domain will be the basis for designing levels of redundancy and a function distribution policy that can achieve the required performance within acceptable workload levels.

Operator trade-offs: The operator's value-based function scheduling and strategy selection often involves online satisficing decisions and speed-quality trade-offs (see discussion above). This can take the form of more shallow processing (e.g. in problem solving and decision making, Payne and Bettman, 1988), use of an alternative processing strategy (Sperandio, 1978), or 'buying time' by slowing down the process itself (e.g. a production line). While these decision trade-offs strongly depend on the semantics of the specific function, temporal reasoning itself involves costs and benefits. Higher levels of temporal reasoning and awareness (Grosjean and Terrier, 1999) might support problem solving and situation awareness (provided that functions are sufficiently predictable), but will require close familiarity with the system and absorb attentional resources.

A similar trade-off exists between *control* (of the immediate system state) and *planning* (assembling a goal-directed action sequence or strategy). A more elaborate plan will simplify control decisions. With a rough or incomplete plan, control decisions will require more online reasoning. Either strategy might be appropriate depending on characteristics such as predictability, time pressure and operator capabilities. For instance, Amalberti and Deblon (1992) report that expert fighter pilots plan flight routes in more detail and consider more problem scenarios than less experienced pilots.

There is often a correlation between the quality of planning and control decisions and the operator's temporal horizon: if causes and effects are only assessed for the short term, or not at all, decisions tend to be erratic and based on arbitrary situational cues. Reasoning about a wider temporal window will take more potentially relevant factors into account. Hollnagel's (2000) Contextual Control Model captures these differences in the quality of control by the notion of control modes (scrambled, opportunistic, tactical, strategic). Hollnagel (2001) explicitly discusses the role of time in losing and regaining control.

Few studies have addressed temporal issues in planning directly. Smith, Hill, Long and Whitefield (1997) modelled planning and control of multiple task work in secretarial office administration and identified a number of control rules and planning heuristics for plan maintenance and revision, interruption handling, task switching and sharing, and prioritisation.

Conclusion

This paper introduced a temporal dimension to function allocation and discussed some of the trade-offs of temporal work organisation, both for the system designer and operator. To overcome the limitations of current Dynamic Function Allocation concepts, allocation along the joint human-automation timeline should be considered in addition to allocation on the human-automation resource dimension. If Dynamic Function Allocation is to be applied to a wider set of problems, automation decisions should be seen in the context of value-based strategy selection, allowing for speed-quality trade-offs. Dynamic Function Scheduling is a conceptual framework that has the potential to analyse a wide range of scheduling and planning behaviour and provide guidance for the designer in assessing the risks and benefits of temporal flexibility in a system. Future work, using both microworld experimentation and case studies, should address problems of temporal reasoning, awareness, and temporal planning and control.

Acknowledgements

This work was supported in part by the UK EPSRC DIRC project (www.dirc.org.uk), grant GR/N13999.

Referenecs

Amalberti, R. & Deblon, F. (1992). Cognitive modelling of fighter aircraft process control: a step towards an intelligent on-board assistance system. *International Journal of Man-Machine Studies, 36*, 639-671.

Block, R.A. (Ed.). (1990). *Cognitive Models of Psychological Time.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Card, S.K., Moran, T.P. & Newell, A. (1980). The keystroke-level model for user performance with interactive systems. *Communications of the ACM, 23*, 396-410.

Decortis, V., de Keyser, V., Cacciabue, P.C. & Volta, G. (1991). The temporal dimension of man-machine interaction. In G.R.S. Weir & J.L. Alty (Eds.), *Human-Computer Interaction and Complex Systems*. London: Academic Press.

De Keyser, V. (1995). Time in ergonomics research. *Ergonomics, 38*, 1639-1660.

De Keyser, V., Ydevalle, G. & Vandierendonck, A. (Eds.). (1998). *Time and the Dynamic Control of Behavior*. Göttingen: Hogrefe.

Doob, L.W. (1971). *Patterning of Time*. New Haven, CT: Yale University Press.

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

Fraser, J.T. (1978). *Time as Conflict*. Basel: Birkhäuser.

Friedman, W. (1990). *About Time: Inventing the Fourth Dimension*. Cambridge: MIT Press.

Grosjean, V. & Terrier, P. (1999). Temporal awareness: Pivotal in performance? *Ergonomics, 42*, 1443-1456.

Hancock, P.A. & Scallen, S.F. (1998). Allocating functions in human-machine systems. In R.R. Hoffman, M.F. Sherrick & J.S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William M. Dember* (pp. 509-539). Washington, DC: American Psychological Association.

Harris, W.C., Hancock, P.A. & Arthur, E.J. (1993). The effect of taskload projection on automation use, performance, and workload. *Proceeding of the 7th International Symposium on Aviation Psychology*.

Harrison, M.D., Johnson, P.D. & Wright, P.C. (2002). Automating functions in multi-agent control systems: supporting the decision process. *Proceedings of the Tenth Safety-critical Systems Symposium* (pp. 93-106). London: Springer.

Hildebrandt, M. & Harrison, M. (2002). The temporal dimension of Dynamic Function Allocation. Paper to be presented at *Eleventh European Conference on Cognitive Ergonomics.*

Hollnagel, E. (1991). The phenotype of erroneous actions: Implications for HCI design. In G.R.S. Weir & J.L. Alty (Eds.), *Human-Computer Interaction and Complex Systems*. London: Academic Press.

Hollnagel, E. (2000). Modeling the orderliness of human action. In N. Sarter & R. Amalberti (Eds.), *Cognitive engineering in the aviation domain*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hollnagel, E. (2001). Time and control in joint human-machine systems. *Proceedings of the 2nd IEE People In Control Conference* (pp. 246-253).

Macar, F., Pouthas, V. & Friedman, W.J. (Eds.). (1992). *Time, Action, Cognition*. Dordrecht, The Netherlands: Kluwer.

Michon, J.A. & Jackson, J.L. (Eds.). (1985). *Time, Mind and Behavior*. Berlin: Springer.

Moray, N., Inagaki, T. & Itoh, M. (2000). Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied, 6*, 44-58.

Morrison, J.G. (1993). *The adaptive function allocation for intelligent cockpits program: Interim research and guidelines for the application of adaptive automation* (Technical report). Warminster, PA: Naval Air Warfare Center, Aircraft Division.

Parasuraman, R. (1993). Effects of adaptive function allocation on human performance. *Proceeding of the FAA/NASA Conference on Artificial Intelligence and Human Factors in Air-Traffic Control and Aviation Maintenance*. Daytona Beach, FL: Embry-Riddle Aeronautical University.

Parasuraman, R., Molloy, R. & Singh, I.L. (1993). Performance consequences of automation-induced 'complacency'. *International Journal of Aviation Psychology, 3*, 1-23.

Payne, J.W. & Bettman, J.R. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 534-552.

Prabhu, P., Drury, C. & Sharit, J. (1997). Using temporal information in time-constrained tasks. *Proceedings of the Human Factors Society 41st Annual Meeting*.

Rencken, W.D. & Durrant-Whyte, H.F. (1993). A quantitative model for adaptive task allocation in human-computer interfaces. *IEEE Transactions on Systems, Man, and Cybernetics, 23*, 1072-1090.

Roeckelein, J.E. (2000). *The Concept of Time in Psychology: A Resource Book and Annotated Bibliography*. Westport, CT: Greenwood.

Scerbo, M.W. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M.

Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 38-63). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sheridan, T.B. (1981). Understanding human error and aiding diagnostic behavior in nuclear power plants. In J. Rasmussen & W.B. Rouse (Eds.), *Human detection and diagnosis of system failures* (pp. 19-36). New York: Plenum Press.

Smith, W., Hill, B., Long, J. & Whitefield, A. (1997). A design-oriented framework for modelling the planning and control of multiple task work in secretarial office administration. *Behaviour & Information Technology, 16*, 161-183.

Sperandio, J-C. (1978). The regulation of working methods as a function of work-load among air traffic controllers. *Ergonomics, 21*, 195-202.

Tattersall, A.J. & Morgan, C.A. (1991). The function and effectiveness of dynamic task allocation. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics, vol. 2* (pp. 247-255). Aldershot: Ashgate.

Walden, R.S. & Rouse, W.B. (1978). A queueing model of pilot decision making in a multitask flight management situation. *IEEE Transactions on Systems, Man, and Cybernetics, 8*, 867-875.

An Examination of Risk Manager's Perceptions of Medical Incidents

Michele Jeffcott and Chris Johnson,

Dept. of Computing Science, University of Glasgow, Glasgow, G12 9QQ, Scotland.
http://www.dcs.gla.ac.uk/<~shellyj><~johnson>

Abstract: Although much research has examined the risk perceptions of 'lay' public to a variety of environmental and public health hazards, little attention has been given to how subjective opinions and value judgements affect those who manage and assess risks as their profession. This paper outlines the results of a psychometric questionnaire administered to 'risk managers' who work as part of a nation-wide risk network, dedicated to improving the quality of Scottish health care. A number of medical incident scenarios were presented and the participants were asked to rate them according to nine pre-determined risk characteristics. The results allow a comparison of the risk perceptions that those who actually compose risk decisions and implement interventions have in regard to hazards resulting from technological, human-machine interaction (HMI) and 'human' error incidents. The analysis concludes that both technology alone and 'human' error incidents are rated much more positively than those involving HMI failures are.

Keywords: risk perception, medical incidents, human machine interaction.

Introduction

Before reviewing research on perceptions of risk, it is instructive to examine the very nature of the risk concept itself. It contains elements of subjectivity that provide insight into the complexities of public perceptions. The Oxford English dictionary defines risk as the chance of suffering harm or loss. In contrast, Vlek and Stallen (1980) believe risk to be comprised of: the *probability* of a potential loss (chances or likelihood), and some *magnitude* of that potential loss (severity of significance). Leveson (1995) relates risk intricately to the hazard concept: a set of conditions of a system that, together with other conditions in the environment of the system, will lead inevitably to an accident. The combination of severity and likelihood of occurrence is often called the hazard level. Risk is therefore the hazard level combined with (1) the likelihood of the hazard leading to an accident and (2) hazard exposure or duration.

Regardless of the definition, however, the probabilities and consequences of adverse events, and hence the "risks," are typically assumed to be objectively quantified by risk assessment. Much social science analysis rejects this notion, arguing instead that such objective characterisation of the distribution of possible outcomes is incomplete at best and misleading at worst. These approaches focus instead on the effects that risky outcome distributions have on the people who experience them. In this tradition, risk is seen as inherently subjective (Pidgeon et al.1992; Weber, 2001) and as including a considerable number of factors, many of them intangible (Slovic, 2001). Large uncertainties and gaps in knowledge still exist, and questions of completeness, the quantification of 'human' error, the wide use of judgement, and the influence of management and organisation, have led to doubt as to the relevance and even usefulness of quantitative risk analysis in risk management decision-making. Risk management is a social and political process and the impact of public risk perception and knowledge is paramount (NAS Report, 1996). This is particularly true of a patient-centred rather than profit-centred industry like healthcare.

A progressive rise in medical incident litigation, added to a genuine desire to improve quality of care to patients, has motivated dramatic advances in safety and the re-evaluation of risk management and communication strategies throughout the British National Health Service (NHS). However, the success of these initiatives relies on the wide integration and support of healthcare workers at all organisational levels. This cannot be achieved without a thorough understanding of the underlying attitudes that these professionals have to the risks and hazards of their daily work (Hale & Glendon, 1987).

This research examines the risk perceptions of those currently spearheading risk management in NHS trusts throughout Britain. To give them their generic name they are 'risk managers', however this group encompasses a number of different job titles, such as medical director, clinical governance manager and nursing and division heads. Despite this variation in titles, they are all appointed by their respective trusts to oversee the introduction of incident reporting schemes. As a result of this, they share a commitment to learning how to better educate staff on the risks and hazards they face with both the equipment they use and, the procedures they perform on patients.

This paper describes two studies. The first study involves the categorisation of primary causes of medical incidents by risk managers. This is in order to examine the influence that technological and non-technological factors are seen to play in the development of adverse events in hospitals. The second study presents some of the same risk managers with a number of incident scenarios, which they must rate against a number of pre-determined risk characteristics. The outcome provides a measure of their risk perceptions towards three different types of incidents: those caused by technology alone (TA), those caused by non-technology alone (NA) and those incidents which occurred as a result of human-machine interaction (HMI). Therefore the affect that technology, or lack thereof, has on risk perceptions is the main focus of this work.

Risk Perception and Communication: Just as the physical, chemical, and biological processes that contribute to risk can be studied scientifically, so can the processes affecting risk perceptions. The term ‘risk perception’ is used to describe attitudes and intuitive judgements about risk (Slovic, 1992); in a broader sense, however, risk perception often also includes more general evaluations of and reactions to risk (e.g. regarding the acceptance or mitigation of risk).
The importance of attempting to understand how those *affected* by risk decisions perceive risk is well established. Early psychometric studies (Starr, 1969; Fischhoff et al., 1978) on the ‘lay’ public found that different groups within society have differing perceptions of the risk from the same hazard. This was explained as a natural consequence of the host of variables, including an individual’s background, prior experience, etc., which affect their perception of risk (Slovic et al. 1979).
However, little attention has been given to how those *responsible* for making risk decisions and performing quantitative risk analyses actually perceive risk. If we accept that ‘lay’ public use subjective opinions and value judgements to assess risk then we must acknowledge that the same process may occur with risk manager’s in the NHS. And as their role in managing risk empowers them to control how risk is communicated throughout hospitals, it seems that their perceptions are of considerable importance. Johnson (1993) reported that information provided by risk managers’ could directly change public opinion. By controlling how risk information is presented, risk managers therefore have a large role in the formation of the risk judgements and attitudes that their staff hold towards certain hazards and equipment.

Risk Management in Scottish Healthcare: The context of this research is the Clinical Negligence and Other Risks Indemnity Scheme (CNORIS) - a risk management strategy which was introduced in the NHS in Scotland in June 2000. It was developed by the Scottish Executive Health Department (SEHD) in partnership with Willis Limited, the appointed scheme manager, and has two principal aims. Firstly, to provide cost-effective claims management and financial risk pooling arrangements for all of Scotland's NHS Trusts and Health Boards. And secondly, to encourage a rigorous and logical approach to risk management in both the clinical and non-clinical sectors of the NHS in Scotland (NHSiS). The CNORIS scheme provides incentives for organisations to manage their liabilities from negligent acts and omissions by their employees and from other risks (MEL, 2000; HDL, 2000).
The scheme revolves around ten standards, each with three levels and corresponding targets. Progress to date on Level One, involving the setting up of management systems to provide the necessary structure for an effective trust-wide risk initiative, has been encouraging. However, Levels Two and Three, which deal with more advanced requirements involving the wider integration of staff and other stakeholders, present a more difficult challenge. An example of this is at Level Two of the Clinical Risk Management Process Standard which requires that all relevant stakeholders are kept informed and, where appropriate, consulted on the management of significant clinical risks faced by the organisation. Responsibility for this more diffuse risk communication will naturally fall to risk managers. Therefore it seems both appropriate and timely to achieve a heightened appreciation of the underlying perceptions of risk managers towards the technological, human-machine interaction and ‘human’ error incidents that occur in hospitals.

Technological ‘Stigma’: There is increasing evidence that technology and the impact of technology on public safety is the most difficult of all subjects to communicate accurately and fairly (Garrick, 1998). Generally technical information is poorly presented and as a result, the impact of technological solutions misrepresented. This phenomenon is described as ‘technological stigma’ (Slovic et. al, 1994). Certain technologies are ‘marked’ or perceived as a threat or a risk to society. Gregory et al. (1995) make the important point that “technological stigmatisation is a powerful component of public opposition to many proposed new technologies, products, and facilities.”
This is particularly relevant to this research. Risk managers both compose and impose the majority of internal risk literature and protocols. It would be plausible that they could communicate stigma in regard to new and/or existing technologies to their staff. This may have a detrimental effect on attitudes and subsequent behaviour towards equipment and devices. Even more importantly though, is the effect that

risk managers' personal attitudes toward technology have on their comparative perceptions of incidents involving technical equipment/devices and those which involve 'human' error alone.

Study One: The role of technological and non-technological factors in medical incidents

Participants: In total, thirty-three risk professionals (20 male, 13 female) took part in an incident categorisation exercise. They were all members of the Scottish CNORIS risk management syndicate, which is split geographically into an East and West network. The categorisation task aimed to examine the role that risk managers attributed to technology versus 'human' error when deciding the primary cause of a number of medical incidents. The main participant group were split into 13 who attended the East network meeting, and 20 who attended the West network meeting. This enabled a comparison between the two groups to see if there was general agreement across Scotland of which risks are more salient in contributing to incidents in NHS hospitals. It also enabled selection of those scenarios which were most consistently selected as belonging to a particular category and therefore were most suitable to be taken forward for use in Study Two: The Risk Perception Questionnaire.

East Group Characteristics: The East group, consisting of 13 participants (8 male, 5 female), were first to carry out the categorisation task. Three of them were Nursing and Midwifery Managers, with an average of 3 years 6 months experience. Three were Clinical Governance Co-ordinators, with an average of 6 months experience. Two were Quality Managers with an average of 2 years experience and there were also three Risk Managers with 8 years 3 months experience. The remaining two participants decided to keep their details anonymous. Although the specific remits of these roles differ, all of these professional groups are responsible for the control and communication of Clinical Risks, within a framework of adherence to the CNORIS standards. Some perform this as part of their existing clinical work (e.g. nursing manager) and some as part of their wider responsibility in assuring the quality of patient care (e.g. quality managers). All are therefore well qualified to make judgements about medical incidents and risks in hospital environments.

Categorisation Exercise: The categorisation exercise required the participants to read 20 short incident scenarios. These were real-life hospital incidents, selected and summarised, from two government reports: An Organisation with a Memory (2000) and Building a Safer NHS (2001). After reading each, participants were asked to attribute the primary cause of the incident to one of four groups: Technology, Non Technology, Mixture and Don't Know. The definitions for each of the groups were given to the participants prior titles to the categorisation exercise. They were adapted from Hyman's work on errors in the use of medical equipment (Hyman, 1994) and are shown below in Table 1:

1.	A *Technology* incident is due to a malfunction of equipment and does not result from interaction by a member of staff or other person.
2.	A *Mixture* incident involves a Human-machine failure when the user causes or initiates a malfunction of equipment due to the complexity of the interface.
3.	A *Non Technology* incident involves 'Human' Error when there is no interaction with the technology and no technological failure occurred. It also includes those rare incidents where interaction with the technology occurred but was not due to a failure of the interface.
4.	The *Don't Know* category is an acceptable answer where you feel you are unable to make a clear judgement about incident causation.

Table 1 – Definitions of Incident Types

East Group Results: Out of the 20 original scenarios the East Group (13 participants) categorised 9 of the scenarios as Non Technology. The next highest categorisation was for the Both category, with 7 incidents being assigned. Only 4 incidents were assigned to the Technology category. This was a particularly surprising result as when the original 20 scenarios were selected, there was a conscious attempt to get a good mix of technological and non-technological incidents. These East group results reflect the risk managers' willingness to highlight the contribution of human fallibility to hospital incidents, as demonstrated by the high number of Non Technology category selections. At the same time they demonstrate a reluctance to 'blame' an incident on Technology alone, instead opting for the Both category in the majority of incidents involving technical failures. This is an interesting finding as it may reflect a tendency in NHS risk manager's to emphasize the human element as a contributory factor in incidents as they feel this is the most feasible area for them to affect changes and reduce risks within their respective trusts. Finally, no participants selected the Don't Know category for any of the incident scenarios, which is most likely a reflection on their experience.

Incident Scenarios: In terms of selecting the incidents that most strongly represented each category, and would be taken forward to form the basis of Study Two's risk perception questionnaire, the scenarios with more than 77% agreement (10 Ss out of 13) were chosen. This resulted in three groups of three scenarios being used (9 in total). The three incident groups were those caused by *Technology Alone* (TA), *Non Technology Alone* (NA) and *Human Machine Interaction* (HMI). The HMI group represents the 'Both' ategory where both technology failure and 'human' error were involved in incident causation. The nine tables below show these nine incident scenarios used in the risk perception study. In the questionnaire the incidents were obviously randomly distributed. However for illustration here in the paper they are divided into the three categories, with Technology Alone first (Tables 2a,b&c), then Non Technology Alone (Tables 2d,e&f) and finally the three Human Machine Interaction incidents (Tables 2g,h&i):

Table 2a "Button" Incident	A 23-year-old healthy mother had some difficulty in delivering the placenta, which looked ragged. The uterus had failed to contract and the woman began to bleed. After transfer from the labour ward, a junior doctor examined her. However, the examination procedure was hindered because the button on the bed jammed which prevented the bed being correctly positioned.
Table 2b "Probe" Incident	A vaginal probe used to promote continence via electrical muscle stimulation was to be used with the power level of the device set to minimum. However maximum muscle stimulation occurred as soon as it was switched on. Although no injury was caused, the extent of the stimulation was unexpected and distressing for the patient. Investigation showed that a breakdown in the manufacturer's quality system allowed the faulty device to be despatched after it failed inspection.
Table 2c "Infusion" Incident	An institution experienced an inadvertent delivery of a vasoactive drug via a computerised infusion device during cardiac anaesthesia. Due to prompt physician intervention, the misadministration had minimal consequences on the patient.
Table 2d "Dose" Incident	In a three-week period two young children received double the proper dose of medication in a hospital X-ray department, prior to having a scan. In both cases their weight was recorded in pounds, rather than kilograms. The children fortunately suffered minor ill effects
Table 2e "Clips" Incident	A number of women became pregnant following failure of earlier sterilisation's that had been carried out by laparoscopic surgery. The surgeon had attached the sterilisation clips to the wrong part of the fallopian tube.
Table 2f "Tablets" Incident	A hospital patient collapsed after a nurse gave her antibiotic tablets crushed in water via an intravenous drip. Only special fluids can be given via an intravenous drip. Similarly, antibiotics and other drugs can only be given in specially prepared solutions and not through the impromptu crushing of tablets. The patient was rushed to intensive care and subsequently recovered.
Table 2g "Alarm" Incident	A nurse adjusted the high and low alarm limits on a heart rate monitor for a patient with a tracheal tube under respiratory distress, in the absence of direct physician orders. Due to the design of the machine, the limit settings were not continuously displayed. Eventually when the selected 'dangerous' low heart rate alarm limit sounded, the patient's brain was irreversibly damaged. Secretions blocking the tracheal tube had resulted in decreased O2 and a long period of elevated heart rate. But this increase was not enough to trigger the high limit alarm set by the nurse. The subsequent decrease in heart rate due to O2 starvation sounded the low limit alarm, but far too late
Table 2h "Doppler" Incident	When Mrs X went into labour the FHR was monitored by external Doppler. This was then replaced by a scalp electrode, as the midwives were unable to monitor the FHR easily due to maternal size and distress. The trace showed that the FHR was normal up until the time of the scalp electrodes removal as the head was crowning at 12.14 but the delivery did not proceed. The Doppler was re-attached showing a reassuring FHR at 160-170 beats, which led the midwife not to seek assistance until 12.33. At 12.39, the compromised infant was delivered. The misleading CTG trace was a result of a coupling with the maternal heart rate.
Table 2i "Needles" Incident	Patients were injured when given incorrect doses of Lidocaine, for acute management of ventricular arrhythmias. All involved the erroneous use of two 20% preparations in place of a 2% preparation. The concentrates were either a 5-ml syringe containing 1000mg Lidocaine or a 10-ml syringe containing 2000mg Lidocaine. The 2% preparation was a 5-ml syringe containing 100 mg of Lidocaine. The errors occurred as the syringes were confused. The 5-ml syringes are identical in diameter and length. The 10-ml is the same length with a 40% larger diameter.

West Group Characteristics: The West Group (20 participants) then carried out the same categorisation task, but instead of using the original twenty scenarios, they selected from this new set of nine incidents. This was in order to ascertain whether the West group would agree with the East group's categorisations, and therefore whether the risk managers exhibited consistency in their opinions on the causation of different medical incidents.

In total, twenty risk managers (12 male, 8 female) completed this exercise as part of the West group. Seven of these were Nursing and Midwifery Managers, with an average of 2 years 8 months experience. Three were Medical Directors, with an average of 3 years experience, and just one participant had been a Health and Safety Manager for 16 months. The remaining nine participants were all Risk Managers or Co-ordinators, with an average of 2 years 4 months experience. Again although the job titles differ, all of these professionals are responsible for managing Clinical Risks, in accordance with CNORIS standards. The only exception in this group is the Health and Safety Manager whose emphasis is on Non-Clinical Risk.

West Group Results: Much consistency was found between the West Group results and the earlier East Group categorisations. For all three Non Technology Alone incidents, 82% agreement (49 Ss out of 60) with the East Group was achieved. A similarly high 80% agreement (48 Ss out of 60) was found for all three Human Machine Interaction incidents. In the case of Technology Alone incidents, only 38% agreement was found (23 Ss out of 60). However, the 55% majority left over (33 Ss out of 60) categorised that these involved a Human-Machine failure so it is clear that these participants recognised the contributory role of technology to the incident. This result is most noteworthy as it shows that West group risk manager's were as equally reluctant to categorise failures as being caused solely by technology as those belonging to the East Group. Again it appears that the human element is more often identified as the primary cause in medical incidents. It was decided that this theme required further investigation, which was achieved via the more exploratory risk perception study, outlined in the following sections.

Study Two: Risk manager's differing perceptions towards three categories of medical incidents
Participants: The same twenty West Network members who carried out the Study One categorisation exercise were used as participants in this postal risk perception study. The aim of this study was to discover whether risk managers had different risk perceptions of medical incidents depending on the role that technology played in the development of the adverse event. The participant's contact details were obtained via personal contacts within CNORIS. Questionnaires were sent out, with full instructions and a stamp-addressed envelope. Fifty questionnaires were originally sent and twenty received back within a two-week period. Although only a 40% response rate was recorded the sample group that participated were deemed representative and provided enough data for meaningful interpretation of the results.

Risk Characteristics: The Questionnaire presented the West Group participants with the same nine incident scenarios shown in Tables 2a to 2i. This time however they were required, after reading each scenario again, to rate the incidents on nine characteristics of risk similar to those found to be important in prior studies by Slovic, Fischhoff et al. (1985) and Kraus and Slovic (1988). Table 3 below shows these nine characteristics:

1. *Anticipatory knowledge* of risks by *risk managers*
2. *Anticipatory knowledge* by those involved in adverse event i.e. *health care workers*
3. *Severity* of the consequences (patient and/or staff present)
4. *Dread* of the entire range of potential consequences
5. *Confidence* in future use of the technology (or in performance of the activity)
6. The overall *Riskiness* of the technology or activity (to both patient and/or other staff)
7. Ability to *Control* the risks involved with the technology or activity
8. Ability to *Observe* the risks at the near miss stage prior to development of an incident
9. Future *Effort* needed for Risk Reduction

Table 3 – The Nine Risk Characteristics

The terms used for the characteristics were not explained explicitly to the participants, which goes against normal construct validity considerations. Construct validity (Child, 1954) refers to the degree to which a test measures the construct, or psychological concept or variable, at which it is aimed (e.g., intelligence, anxiety). In this case, the relevant constructs are the nine risk characteristics and the lack of explicit explanation of their respective meanings was by design. This is true to the nature of traditional psychometric risk studies where despite there being no freedom to choose the characteristics that are used for measurement, subjective interpretation of said characteristics is appropriate (Slovic, 2001). Each of the risk characteristics was rated on 10-point likert scales, with the less serious pole of each scale on the left-hand side and the more serious pole on the right. A partial illustration of the questionnaire is shown in Table 4. It shows a non technology scenario and the first three out of nine risk characteristic questions:

SCENARIO THREE. *A hospital patient collapsed after a nurse gave her antibiotic tablets crushed in water via an intravenous drip. Only special fluids can be given via an intravenous drip. Similarly, antibiotics and other drugs can only be given in specially-prepared solutions and not through the impromptu crushing of tablets. The patient was rushed to intensive care and subsequently recovered.*
i) To what degree should the risks involved in the activity or technology failure that led to the adverse event have been anticipated by risk managers?
Anticipated **1 2 3 4 5 6 7 8 9 10** Unanticipated
ii) To what degree should the risks involved in the activity or technology that led to the adverse event have been anticipated by those present during the event?
Anticipated ***1 2 3 4 5 6 7 8 9 10*** *Unanticipated*
iii) When the risk from this activity or the technology is realised in the form of an adverse event, how likely is it that the consequences will be fatal?
Not Fatal **1 2 3 4 5 6 7 8 9 10** Fatal

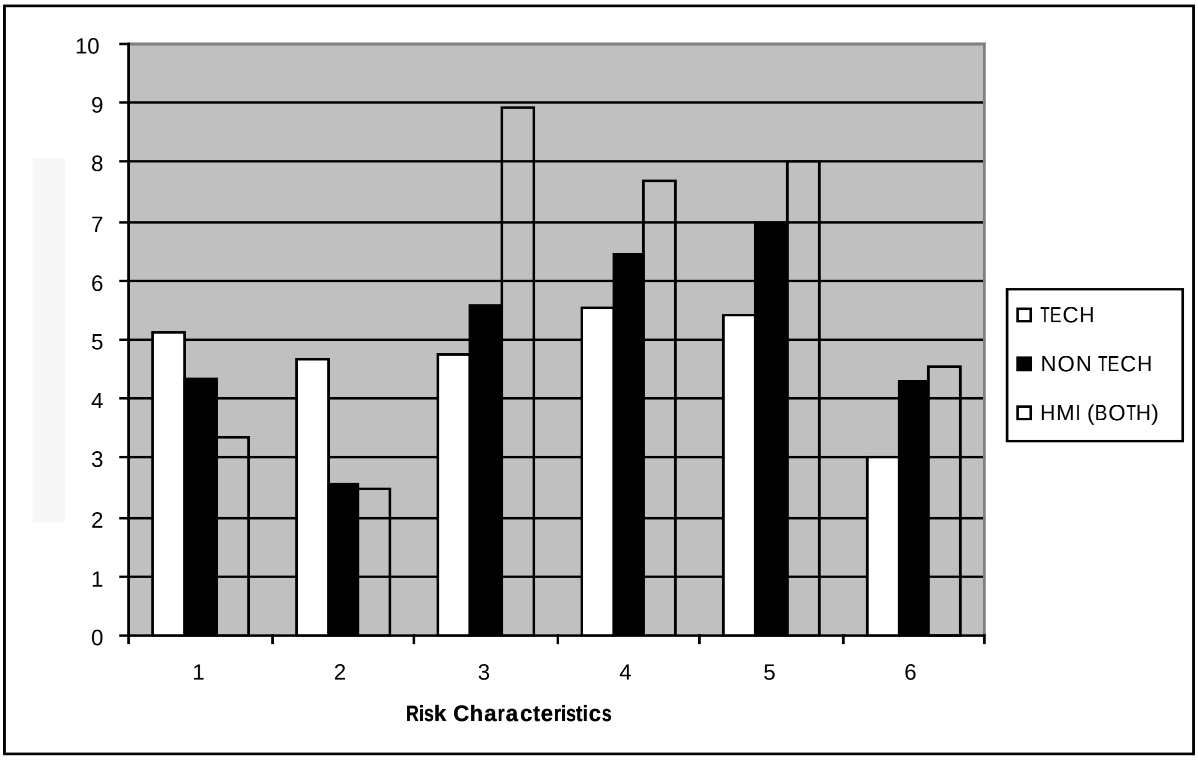
Table 4 – Partial Illustration of the Risk Perception Questionnaire

Results: The mean perceived risk of the three types of incident groups varied greatly, from 2.1 to 9.1 on the 10-point likert scales. The two incident scenarios judged to be most risky were those involving the Doppler Fetal Heart Rate (FHR) monitor and the Lidocaine Needles. Both these scenarios were judged by risk managers as involving Human Machine Interaction (HMI) failures. The two incident scenarios judged to be least risky were the Continence Probe and Bed Button scenarios, both from the Technology Alone (TA) category. Table 5 presents the incidents whose mean ratings were extreme on each of the nine judgment scales. Their corresponding categorisations are also included for each scenario. The three incidents involving Human Machine Interaction (HMI) failures are repeatedly the most negatively rated on all characteristics. The three Non Technology Alone (NA) incidents were consistently rated toward the less serious pole of each scale. This was true also for Technology Alone (TA) incidents, although to a smaller extent.

Risk Scale	**Highest Scenarios**	**Group**	**Lowest Scenarios**	**Group**
1 Knowledge (Risk Managers)	Probe 6.4 Button 5.8	TA TA	Dose 2.6 Needles 2.9	NA HMI
2 Knowledge (Health Workers)	Probe 6.7 Button 4.4	TA TA	Dose 2.3 Tablets 2.3	NA NA
3 Severity	Alarm 9.1 Needles 9.0	HMI HMI	Probe 2.7 Clips 3.4	TA NA
4 Dread	Doppler 7.9 Needles 7.8	HMI HMI	Probe 4.6 Clips 5.1	TA NA
*5 Confidence**	Doppler 5.9 Probe 6.0	HMI TA	Dose 4.2 Tablets 4.2	NA NA
6 Riskiness	Doppler 8.2 Needles 8.0	HMI HMI	Probe 4.2 Button 4.9	TA TA
*7 Controllability**	Probe 7.7 Clips 6.0	TA NA	Dose 2.6 Needles 2.6	NA HMI
8 Observability	Doppler 5.7 Alarm 5.5	HMI HMI	Button 2.2 Probe 2.6	TA TA
*9 Effort**	Doppler 5.7 Needles 5.5	HMI HMI	Tablets 3.5 Dose 4.2	NA NA

Table 5 (above) - Extreme Scenarios for the Nine Characteristics

Table 6 (below) – Comparing Risk Managers Mean Scores for the Three Incident Categories



Standard deviations across all mean risk scores were calculated. For most of the characteristics these ranged between 1.5 and 3 and so provided a large enough range for comparison between scenarios. Unfortunately though, *the standard deviations for the *Confidence*, *Control* and *Effort* characteristics were all below 1.0 (0.7, 0.6, 0.3 respectively). However, the uniformity with which twenty risk managers rated these characteristics shows us that differences in the areas of *Confidence*, *Control* and *Effort*, regarding technology and causation, appear to be negligible. The following analysis centres around the remaining six risk characteristics, which provide interesting insights about the variations in the risk manager's perceptions toward the three groups of medical incidents. Table 6 represents the mean risk perception scores for each of the three categories of incidents: Technology Alone, Non Technology Alone and Human Machine Interaction.

Referring to Table 6 above, Risk Characteristic number1 and 2 represent the *Knowledge* characteristic - firstly from the risk manager's own perspective and then from the level of risk knowledge that they hypothesise a healthcare worker directly involved in the incident might have. When comparing the mean perception scores calculated from the three different incident categories we see that risk manager's rated Technology Alone incidents consistently higher than the other two incident categories on both *Knowledge* risk characteristics. This means that Technology Alone incident scenarios were perceived as unanticipated to a greater extent than those involving Human Machine Interaction failure or Non Technology Alone. This suggests that risk manager's have an expectation of the reliability of technology and anticipate incidents only when humans interact with technological equipment or when they make mistakes alone. Although the *Confidence* characteristic result has been discounted, this *Knowledge* result does reflect a form of confidence that risk managers have in technology and its correct and safe functioning.

This finding is echoed on Risk Characteristics 3, 4 and 5, representing *Severity*, *Dread* and *Riskiness*. Technology scores the lowest, making it the most positively rated of all incident categories on these three characteristics. Non Technology is rated slightly higher, and therefore closer to the negative scales, but it is the Human Machine Interaction failures that score very highly and so are rated as both most severe, dreaded. and risky of all incidents. It appears therefore that technology is only perceived as risky when humans are involved, resulting in Human Machine Interaction (HMI) failures. When looking back at Table 5, the two most extreme negatively rated examples for *Riskiness* were the Doppler Fetal Heart Rate (FHR) Monitor and the Lidocaine Needles, with scores as high as 8.2 and 8.0 respectively. Both incidents involved unfortunate breakdowns in the complex interactions between equipment and user/s, with adverse effects to the patient/s.

When turning to the results for *Observability*, the last Risk Characteristic in Table 6, it becomes apparent that a lack of interface visibility may be the problem behind these breakdowns and the subsequent reason for such negative risk perceptions of Human Machine Interaction incidents with equipment and devices. Technology Alone incidents score lowest on *Observability*. This implies that risk managers' perceive incidents that involve devices malfunctioning in hospitals, without direct human input or error, as highly visible and therefore possibly more preventable if they were to occur in their hospital. Although, the difference between the Non Technology and the Human Machine Interaction mean ratings are negligible for *Observability*, it is once again the Human Machine Interactions which scored highest, and therefore are perceived as being the least visible of all incidents. Table 5 supports this as it shows two Human Machine Interaction incidents, the FHR monitor and the Alarm Limit Settings, as the highest scenarios at the negative pole. At the other extreme, the positive pole, are two Technology Alone scenarios – the Bed Button and the Continence Probe. From these *Observability* scores, it appears that risk managers perceive problems with technology as being the most visible. This may account for their low anticipation of the occurrence of technological incidents as problems are easily seen and fixed before an incident develops, that is, at the near-miss stage.

***Discussion*:** These results showed that those incidents viewed as being the result of Human Machine Interaction failure were generally judged to be highly *Dreaded* and *Risky*, and displaying poor *Observability*. Also, risk managers seem well aware of the potential risks that Human Machine Interaction incidents pose, as recorded by their high *Knowledge* results. Conversely, both Technology Alone and Non Technology Alone incidents scored low on *Dread*, *Severity*, and *Riskiness* and high on *Observability*. As such, anticipation (*Knowledge*) of Technology Alone and Non Technology Alone incidents were rated at the less serious pole of the scale as risk managers perceived these problems as occurring less frequently.

A possible limitation of this study is that in order not to tamper with the real-life incidents used, the outcomes of the scenarios were not standardised. Although this variation was only slight, it may have had a confounding effect on the risk managers' responses, that is, more *Dread* was perceived for an incident with a more damaging outcome. However, participants were instructed before beginning the questionnaire to consider the entire range of consequences for each incident and not be limited to what was reported in the scenario. Another drawback was the very short incident summaries, which meant that consideration of contextual factors contributing to risk, such as teamwork and communication issues, were beyond the scope of this study.

Conclusion

This paper has reported on two distinct studies. However, they share common ground and the conclusion of this work is supported by their complementary findings. Study One revealed that risk managers' across the Scottish NHS show consistency with their categorisations of different medical incidents. The most interesting aspect of this categorisation agreement was that they appeared to underestimate the contributions that technological factors have in causing incidents, emphasising non technological and 'human' error causes instead. The results from Study Two directly support this finding. The risk manager's ratings of the medical scenarios revealed that incidents involving Human Machine Interaction failures are perceived as posing the greatest risks, whilst those incidents involving Technology Alone are perceived as relatively innocuous, and on a par with Non technology Alone scenarios. This disproves the earlier hypothesis that the stigmatisation of technology in the NHS may result in the transfer of technophobic attitudes detrimental to the acceptance of both new and existing equipment and devices. On the contrary, the conclusion of this work is that risk managers are only weary of technology when there is human interaction and where errors involving complex interfaces and systems result in potentially dangerous situations for both patients and staff.

Therefore it can be concluded that the thirty- three risk managers that participated in our two studies perceive Human Machine Interaction failures as the biggest current challenge to the task of risk management in the Scottish NHS. This conclusion has a direct impact on the development of CNORIS risk management strategy, highlighting the need for more attention to be given to the relationship between healthcare professionals and the technologies that they use in their daily work. This may include stricter risk protocols for the use of equipment, incorporated within a more open reporting culture of common problems that people encounter using technology. Also a forum to share lessons learnt from near miss situations, that is how Human-machine Interaction (HMI) failures are recovered from before an incident develops, may also be of benefit to the development of the CNORIS Standards.

Finally and most importantly, more attention may be needed in terms of staff training for equipment use. Existing training schemes may be adequate for initial instruction, but refresher workshops and reminder notices given by risk managers may be necessary to help to limit the occurrence of Human Machine Interaction incidents in the future. Whatever course of action, it seems clear that value is added to the process of risk management by a better understanding of stakeholders risk perceptions towards different types of medical incidents. The natural progression of this work is to attempt to uncover some of the perceptual risk variations that exist within and between professional working groups at the frontline of healthcare.

Acknowledgements
Thanks are due to both Brian Kennedy and Eunice Muir at CNORIS for allowing this collaborative work and also to the twenty risk managers who kindly participated in the study. Also thanks go to my supervisor, Prof. Chris Johnson. This work is supported by the UK Engineering and Physical Sciences Research Council and is carried out as part of the Univ. of Glasgow Accident Analysis Group.

References
Child, I. L. (1954) Personality. Annu. Rev. Psychol. (5):149-171.
Department of Health. An Organisation with a Memory: Report of an expert group on learning from adverse events in the NHS chaired by the Chief Medical Officer. The Stationery Office. June 2000.
Department of Health. Building a safer NHS for patients: Report of an expert group on learning from adverse events in the NHS chaired by the Chief Medical Officer. April 2001.
Fischhoff B, Slovic P, Lichtenstein S, Read S & Combs B. (1978) How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. Policy Sciences (9):127-152.
Garrick B.J. (1998) Technological stigmatism, risk perception and truth. Reliability Engineering and System Safety (59):41-45.
Gregory R, Flynn J. & Slovic P. (1995) Technological stigma. American Scientist (83)3: 220-223.
Hale A. & Glendon I. (1987) Individual Behaviour in the Control of Danger. Elsevier, Amsterdam.
Hyman W.A. (1994) Errors in the Use of Medical Equipment. In M.S. Bogner (ed.) Human Error and Medicine, Lawrence Erlbaum. Hove, UK. (pp.327-348)
Johnson B.P. (1993) Advancing understanding of knowledge’s role in lay risk perception. Risk-Issues in Health and Safety (4):189-211.
Kraus N. & Slovic P. (1988) Taxonomic analysis of perceived risk. Risk Analysis (8)3:435-455.
Leveson N. G. (1995) Safeware: System Safety and Computers. Addison-Wesley, USA.
National Research Council (1996) Understanding Risk: Informing Decisions in a Democratic Society: National Academy Press, Washington DC.
NHS MEL(2000)18: Clinical Negligence and Other Risks Indemnity Scheme (CNORIS). Scottish Executive Corporate Services.
NHS HDL(2000)19: Clinical Negligence and Other Risks Indemnity Scheme (CNORIS): progress report. Scottish Executive Health Department
Pidgeon, N., Hood, C., Jones, D., Turner, B., & Gibson, R. (1992). Risk perception. In G. Royal Society Study (eds.), Risk: Analysis, Perception, and Management (pp. 89-134). London: The Royal Society.
Slovic, P., Fischhoff, B., & Lichtenstein, S. (1979). Rating the risks. Environment (21)3:14-20.
Slovic P, Fischhoff B, & Lichtenstein S. (1985) Charaterising perceived risk. In R.W. Kates et al. (eds.) Perilous progress: Technology as hazard, Westview, Boulder, CO. (pp. 91-123)
Slovic, P. (1992). Perception of risk: Reflections on the psychometric paradigm. In S. Krimsky & D. Golding (Eds.), Social theories of risk (pp. 117-152). New York: Praeger.
Slovic P, Flynn J. & Gregory R. (1994) Stigma happens – Social problems in the siting of nuclear waste facilities. Risk Analysis (14)5:773-777.
Slovic P. (2001) The Perception of Risk. Earthscan, London.
Starr C. (1969) Social benefit versus technological risk. Science (165):1232-1238.
Vlek C, & Stallen P.J. (1980) Rational and personal aspects of risk. Acta Psychologica (45):273-300.
Weber E.U. (2001b) Decision and choice: Risk, empirical studies. In N. J. Smelser & P.B. Baltes (eds.) International Encyclopedia of the Social and Behavioural Sciences (pp. 13347-13351). Oxford, UK: Elsevier.

User Adaptation of Medical Devices

Rebecca Randell and Chris Johnson,

Dept. of Computing Science, University of Glasgow, Glasgow, G12 9QQ, Scotland.
http://www.dcs.gla.ac.uk/~{rebecca}{johnson}.

Abstract: This paper looks at the adaptation of medical devices by nursing staff, based on an observational study carried out in three Scottish intensive care units. We use this as a starting point to consider the possibility of adaptive systems in the medical domain. For such an idea to be acceptable in the design of medical technologies requires accepting the situated nature of medical work. We compare this with the arguments for and against protocols in medicine. By studying adaptation in this context, we hope to encourage designers to develop medical devices that allow users to appropriate and adapt them in effective, productive and, most importantly, safe ways.

Keywords: medical technologies, adaptive systems, observational studies.

Introduction

Adaptation of computer systems by users is an accepted phenomenon (Mackay, 1990; Dourish, 1995a). This paper looks at the adaptation of medical devices by nursing staff in several intensive care units. We use this as a starting point to consider the possibility of adaptive systems in the medical domain.

The idea of adaptive systems is one that has received attention within the field of human-computer interaction (HCI). By adaptive systems, we mean those that allow users to adapt them to fit with their own working patterns. It follows on from work such as that by Suchman (1987) and the growing collection of workplace studies that emphasise the situated nature of computer use, pointing out the complexity of the relationship between the general pattern of use and the details of particular activities.

However, to consider the suitability of such an idea for medical equipment brings up new areas for consideration, with increased concern for safety issues. For such an idea to be acceptable in the design of medical technologies requires accepting the situated nature of medical work. It presents an interesting study of adaptation because there is a level of perceived necessity of such adaptations by those who carry them out that is missing from adaptation in more traditional office environments.

By studying adaptation in this context, we hope to encourage designers to develop medical devices that allow users to appropriate and adapt them in effective, productive and, most importantly, safe ways.

Structure of paper: We start by outlining the study and then describe several examples of user adaptations, taken from observations. We then consider the safety implications of such adaptations, using the arguments for and against protocols as a way of opening up the possibility for allowing such adaptations. We describe how accountability is maintained in the adaptation of medical devices.

The study

This paper looks at the adaptation of medical devices by nursing staff, based on an observational study carried out in three Scottish intensive care units. Eight weeks were spent observing in one unit, followed by two shorter studies of two weeks each. During this time, the hours of the nurses were worked (12 hours a day, including some night shifts). Observations included sitting by particular patients, sitting observing the whole of the ward, attending meetings of both nurses and doctors and training on medical devices. The decision to use observational methods meant that a more detailed understanding of adaptation in the setting could be obtained than would be available through the use of more traditional methods such as interviews.

Examples of user adaptation

To demonstrate what we mean by user adaptation, we start with a series of vignettes. Each vignette recounts an event either observed by or told to the fieldworker. We chose the vignettes so that there is one vignette from each intensive care unit where observations were carried out and also to show the variety that occurs in terms of adaptation of devices. However, various other similar events were observed by or described to the fieldworker in each of the intensive care units.

Vignette 1: Nine months before, a haemofiltration device was purchased by the intensive care unit. The device is designed to also manage the delivery of heparin, an anticoagulant given to assist the haemofiltration. However, the nurses put the heparin through a separate syringe driver rather than through the device because, if they put it through the device, they cannot change the rate of delivery once delivery is started. The decision to put the heparin through the syringe driver was made by the renal core group, a group of nurses responsible for the purchase of the device and also responsible for the subsequent training on how to use the device. After several months of using the device in this way, a critical incident occurred where there was a case of siphonage, meaning that all the heparin was given in one go, rather than being delivered gradually. There was no adverse effect on the patient. The nurses told the clinical physics technicians, who checked the syringe driver to check that it was okay. They also informed the distributor of the haemofiltration device, who passed on the information to the manufacturer. The manufacturer responded by saying that they should not be putting the heparin through the syringe driver but should be putting it through the device. There is a risk of siphonage with any equipment where there is negative pressure (pressure less than that of the ambient atmosphere). The nurses think that the syringe came out of its carriage. Therefore, they will now give the heparin through an infusion pump because it can stand greater pressure (still not putting it through the device). It requires more heparin, which is why they did not use the infusion pumps before. The infusion pumps are not saturated but it is felt that they "have to balance risks" and this is a better alternative to persisting with the syringe drivers.

Vignette 2: Portable monitors are used for transferring patients between wards. The intensive care unit has had the monitors for 4 months. When a patient was being transferred, the monitor being used switched itself off, despite the fact that the battery was charged and should last for two hours. When the nurse switched the monitor back on, a message appeared, saying "BATT COND". On returning to the unit, the nurse informed the ward manager. The ward manager looked up the error message in the user manual and found that the error message refers to the battery condition and means that the battery needs to be replaced. Whether or not the battery has been recharged, it must be replaced after the fiftieth time it is used. The ward manager says that it would be ideal to record the number of times that it is used so that they know when the fiftieth use is, but it is impractical because of a lack of time and more pressing tasks. Since this incident, they have found that if the same thing happens while transferring a patient, you can trick the monitor by taking the battery out and putting it back in, to "let it forget".

Vignette 3: The intensive care unit received a fluid heater, intended for warming blood. Nursing staff used the fluid heater for warming medication, with the device being used for long periods. The device 'packed in'. They informed the manufacturer and have subsequently received new fluid heaters. However, on these new fluid heaters, it is stipulated that they should not be used for more than twelve hours at a time, with a break of at least two hours between each use.

Discussion

User adaptation can take different forms. In Vignette 1, we see a direct change to how the device is used, in violation of the manufacturer's guidelines. Vignette 2 is an example of a workaround developed as a short term solution to the fact that it is difficult to keep an accurate record of how many times the device has been used. Vignette 3 is an example of a device being used for a purpose other than it was designed for.

Other adaptations are 'allowed' by the equipment, such as adjusting alarm limits. For example, one of the alarms on the monitor will always go when taking blood from the patient and the ventilator will alarm when clearing the patient's chest, so often the alarm will be silenced before carrying out the task.

Some adaptations do not affect how the device is used but are simple adaptations to ease use of the device. Frequently, post-it notes are attached to devices, detailing how to use them, and user manuals may be rewritten, adapting the language and removing unnecessary details to make them easier to understand. Information attached to equipment is usually for equipment that is not used often, where nurses may forget what they have to do. It is also a way of ensuring that everyone knows about changes to the way a device is to be used. Other adaptations include basic things such as an elastic band to keep part of a device in place.

Nurses have greater ability to appropriate the physical space (through workarounds such as the arrangement of devices, attachments to devices, and the creation of new user manuals) than to appropriate the devices themselves (the technological space), with the most obvious exception being alarms. When nurses do have the ability to appropriate technological space in this way, technological space does take on increased meaning for them. For example, it is possible to change the alarm limits on monitors and ventilators (or to silence the alarms for a short period). Wide alarm settings on monitors can be seen as a demonstration of confidence, while the colours on monitors have on occasion been changed to reflect the football team that the nurse supports.

Mackay (1990) talks of the 'perceived costs and benefits' of adaptation that determine whether or not a user will adapt a system. But in the intensive care unit, it is often the case that such adaptation is perceived as a necessity. It is also considered to be part of the job; nurses feel that they "almost have to be technicians". Therefore, while Mackay found that most people resisted spending much time customising because they are busy, in the intensive care unit, a lot of time may be given to such adaptations. In all of the intensive care units observed, there was a feeling that they were very much alone in managing the technology. Stories were told of reporting problems to a manufacturers or distributors and waiting a long time for a reply, if one ever came at all. So, nursing staff learn to adapt the devices, or how they are used, to get them to work.

Nurses also adapt equipment so as to be able to get the job done with as little distress for the patients as possible. For example, nurses are aware that alarms can worry patients and visitors because they do not know what the alarm means or whether or not it is for them (they may not even know that there are other patients on the ward). Therefore, they adapt alarm limits so that alarms are not going off unnecessarily. If the number of alarms going off is limited, it is also easier to detect where the alarm is coming from.

Protocols in practice

We can compare this discussion with a prominent discussion in medicine about the role of protocols. Protocols are seen as a means to enhance scientific practice, reduce variations in practice and enhance the quality of care. For example, following on from the death of a chemotherapy patient who was given an intrathecal (spinal) injection of Vincristine rather than an intravenous injection at a UK hospital, recommendations were made for an explicit procedure for the administration of chemotherapy (Toft, 2001). However, critics argue that protocols are not suitable for all situations, that unnecessary use of protocols can lead to deskilling and threaten the healthcare worker's autonomy. Berg (1997a) describes the way in which many protocols are circumvented, tinkered with and interpreted in many different ways, in the same way that the procedures of use for various devices are circumvented. Protocols reinforce a restrictive image of activities, where there is a tendency to perceive the treatment of patients as a sequence of individual, formally rational decisions (Berg, 1997b). One of the reasons protocols are so often disregarded is the clash between the formal image of health care practices embedded in protocols and the realities of ongoing, socially and materially situated work. These arguments against protocols reflect the same arguments we see in HCI against systems that do not reflect working practices and are too restrictive.

The difference between written protocols and protocols as they are implemented in equipment is that written protocols are often high level and nurses can adjust their interpretation of a protocol to fit with the work, whereas equipment forces the nurse to follow specified actions in a specified order.

In the same way that it has been argued that it is not a problem with the idea of protocols, simply that the protocols used are inadequate, one could argue that if technology were 'better' in the first place, adaptation would be unnecessary. For example, with better research, designers of the haemofiltration device would know that nurses would want to change the rate of heparin delivery; designers of the portable monitor would know that it is impractical to expect a record to be kept of how many times the monitor is used; designers of the fluid heater could specify what it can and cannot be used for. Even this would require a greater level of research on the part of manufacturers into the working practices of nurses. But, although that would solve the problems described in the examples, the possibilities for what nurses may want to adapt the device for are endless. "New patients produce new problems" and nurses are not necessarily able to specify beforehand what it is that they will require.

It is clear that adaptation happens whether it is supported by designers or not, but by providing adaptation mechanisms, we can increase the safety implications of adaptations that are made. In the same way that nurses can change alarm settings within certain parameters, we can imagine allowing variations to devices within a certain acceptable safety level. We can see that workarounds could be replaced by much easier and safer solutions, if nurses were able to change equipment. For example, rather than nurses putting heparin through the syringe driver where there is a risk of siphonage, an adaptive haemofiltration device could allow the user to change that aspect of the system so that the heparin delivery rate can be changed once treatment has begun.

The illusion of medical work as a sequence of individual, formally rational decisions affects our conception of what a safe system is. One could ask why we insist on a level of restriction in medical technologies that is not applied in the rest of medical practice. Berg (1997a) argues that 'the health care worker will often act differently in "equal" circumstances – and (s)he will not be attracted to a tool which embodies the illusion of a single answer.' The need for flexibility in medical information technologies has already been highlighted (Heath and Luff, 1996). Like protocols, rigid medical devices deny nurses the flexibility they require when problems become difficult. Certainly, protocols are an important aspect of safety for procedures where there is only one safe approach, such as with the administration of particular drugs. The bottom line is that adaptation happens whether we support it or not[7].

Limitations of adaptation

While providing more control, adaptation also implies costs to the users of the device as devices get more complex. Various modes of use increase the amount that needs to be learnt about a device in a setting where time for training is already limited and where there are already a large number of devices to understand. To adapt a device takes time, but then this has to be balanced against the time that nurses spend adapting devices that do not support such adaptation.

A much trickier problem is the question of how to certify an adaptive system in a safety-critical environment. Opportunities for adaptation increase the complexity of devices. However, again we have to balance these concerns against the fact that devices are adapted regardless.

Accountability

One of the major concerns with any adaptive system, and particularly with safety-critical adaptive systems, is to reduce the risk of antisocial behaviour by making those who carry out adaptations visible and therefore accountable. Whilst agreeing with this, our intention here is not to repeat the discussion on how to enforce accountability, but instead to demonstrate how such adaptations are already accountable[8].

The use of a device is determined not only by the technological components, which define how the system will behave, but also by the social components, which determine acceptable use and behaviour (Dourish, 1993). Technologies do not necessarily impinge on and nullify the social conventions that regulate workplace behaviours; technological spaces are just as amenable to such forces as are physical spaces. In the same way that nurses demonstrate competent behaviour through their interactions with patients, interactions with equipment are equally visible demonstrations of competence, or not. As Mackay (1990) points out in her study of user customisation of software, adaptation cannot be considered primarily an individual activity.

For example, when telling the fieldworker about changing alarm settings, a nurse said "But I must qualify that by saying that I'm experienced. How significant do I think that is? I'd say it is very significant." It would be considered inappropriate for an inexperienced nurse to set wide alarm limits, yet it is acceptable, even expected, for an experienced nurse to do this. Alarm settings are also visible to other nurses; they can come up to the monitor and see them, or see them on the main console that is placed on the nurses' desk. If the alarming settings are considered to be too narrow, i.e. the alarm keeps going off, it is acceptable for another nurse to widen the alarm limits from the main console.

Fundamental changes to the way a device is used, such as putting the heparin through the syringe driver are also unlikely to be carried out by an individual nurse without previously being discussed with other nurses. The decision to deliver the heparin in this way was a decision taken by the renal core group, where possible options were discussed. When one nurse was showing another nurse how to set up the haemofiltration device, the second nurse questioned why they were putting the heparin through the syringe driver; it was a noticeable event.

More generally, use of equipment is something that is subject to much discussion. Nurses will ask each other how to do something or why something is done a particular way. Talking about devices presents nurses with an opportunity to demonstrate their competence, as has also been observed in other professions (Orr, 1996).

---

[7] Our intention here is not to explore how such features to support adaptation could be implemented. How we can support such adaptation has already been explored in the HCI field, most notably by Dourish (1995a; 1995b; 2001) and Mackay (1990).

[8] Bellotti and Edwards (2001) provide guidelines for intelligibility and accountability in context-aware systems that are applicable to adaptive systems generally.

So, despite the very situated nature of the work, we see that adaptation, and use of equipment generally, is carried out within a specific "community of practice", where such actions are observable and reportable, and therefore accountable, making them subject to the social conventions that determine acceptable use (Wenger, 1998).

Conclusions

In this paper, we have described instances of the adaptation of medical devices by intensive care nursing staff and have described how such adaptations are made accountable. Although there are certainly situations where adaptation is not plausible or safe, we hope that through the examples we have given, and through opening up the discussion by comparing it with the arguments for and against protocols, to have encouraged designers to consider where adaptation is appropriate and how to develop medical devices that allow users to appropriate and adapt them in effective, productive and, most importantly, safe ways.

Acknowledgements

The authors gratefully acknowledge the generosity and patience of those working in the intensive care units where the observations were carried out. This work has been supported by EPSRC grant GR/M98302: Communicating Knowledge about Accidents from Synthesised Web Sites.

References

Bellotti, V. and K. Edwards (2001). Intelligibility and Accountability: Human Considerations in Context-Aware Systems. Human-Computer Interaction 16: 193-212.

Berg, M. (1997a). Problems and Promises of the Protocol. Social Science and Medicine 44(8): 1081-1088.

Berg, M. (1997b). Rationalizing Medical Work: Decision-Support Techniques and Medical Practices. MIT Press, Cambridge, Massachusetts.

Dourish, P. (1993). Culture and Control in a Media Space. ECSCW'93 (Milan), Kluwer.

Dourish, P. (1995a). Accounting for System Behaviour: Representation, Reflection and Resourceful Action. Computers in Context, Aarhus, Denmark.

Dourish, P. (1995b). Developing a Reflective Model of Collaborative Systems. ACM Transactions on Computer-Human Interaction 2(1): 40-63.

Dourish, P. (2001). Where the Action Is. The MIT Press, Cambridge, Massachusetts.

Heath, C. and P. Luff (1996). Documents and Professional Practice: 'bad' organisational reasons for 'good' clinical records. CSCW'96 (Boston), ACM Press.

Mackay, W. E. (1990). Users and Customizable Software: A Co-Adaptive Phenomenon (PhD thesis). Management of Technological Innovation. Massachusetts, MIT.

Orr, J. E. (1996). Talking about Machines: An Ethnography of a Modern Job. ILR Press, Ithaca.

Suchman, L. (1987). Plans and situated actions : the problem of human-machine communication. Cambridge University Press, Cambridge, UK.

Toft, B. (2001). External Inquiry into the adverse incident that occurred at Queen's Medical Centre, Nottingham, 4th January 2001. London, Department of Health.

Wenger, E. (1998). Communities of practice: Learning, meaning, and identity. Cambridge University Press, Cambridge, UK.

Introducing Intelligent Systems into the Intensive Care Unit: a Human-Centred Approach

M. Melles*, A. Freudenthal*, C.A.H.M. Bouwman**

* Dept. of Industrial Design, Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands.
M.Melles@io.tudelft.nl, http://studiolab.io.tudelft.nl/melles
** Dept. of Nursing Affairs, Groningen University Hospital, Groningen, The Netherlands.

Abstract: The aim of our study is to develop design knowledge about contextually based intelligent medical systems used in an intensive care unit. The basic thought is that solutions should be user-driven. This paper describes the premises and outline of our research. A conceptual framework is developed based on Vicente's and Rasmussen's ecological approach for interface design. Constraints posed by the characteristics and task goals of the intensive care nurse and posed by the context will be of main importance. Finally, an outline of the research methods is presented. For eliciting the unique and latent knowledge of the user, we propose a participative ergonomic approach. This approach is embedded in a research-through-design cycle, a method for testing theories in an iterative design cycle, with the underlying assumption that results from experiments can be generalised in the form of design guidelines for future products.

Keywords: intensive care unit, intelligent medical systems, human-product interaction, ecological approach, research through design, participative ergonomics.

Introduction

The nursing process in intensive care units (ICU) is increasingly characterised by a heavy reliance on medical equipment. The variety of equipment is large and innovations appear on the market continuously. Due to these technological developments, the profession of intensive care nursing has changed. Nurses are increasingly required to conduct complex therapeutic and diagnostic procedures using the equipment's advanced functionality. Despite this increased functionality, most (monitoring) devices still function essentially as 'single-sensor-single-indicator' devices (Effken, 1997). The task of selecting and integrating the vast amount of data into diagnostic information is still the responsibility of the nurse. Groen (1995) identified enhanced cognitive demands required by the complex equipment as one of the main stress factors in ICU nursing. On top of this, the usability of the ICU equipment is a contributing factor to human error (Bogner, 1994). Devices are not standardised, and procedures for operating and maintaining the equipment are incomplete or difficult.

But the equipment itself is not the only source of stress. Organisational and process-related factors play a role as well (e.g. Leys, 2001; Groen, 1995). The frequent occurrence of dynamic and complex situations in an ICU, combined with a high level of responsibility towards seriously ill or dying patients and their relatives places the nurse under a lot of pressure. Deciding which actions should be taken is often done under time-critical circumstances. There is a high work pace, and the cumulative work pressure combined with working in shifts results in fatigue. On top of this, there is an increasing demand on medical staff for higher efficiency. In the Netherlands this demand is especially high with a structural shortage of qualified personnel. Groen also mentions the stressing effect of working with inexperienced staff who cannot assume equal responsibility. To minimise inexperience, training is crucial. However, there is a lack of general training for nursing staff in the use of technology as well as adequate, task-specific training (Bogner, 1994). Especially older nurses will suffer from this lack of focus on device usage.

The next generation of ICU equipment should be adaptable to the needs and cognitive limits of ICU staff in relation to the constraints posed by the ICU context. Such devices, containing next generation intelligent software, should be able to adapt to the user (e.g. level of experience, preferences and physical condition) as well as to the environmental situation (e.g. level of illumination, noise and presence of colleagues). Furthermore, these products should be able to provide adaptive embedded support to users when appropriate. Several researchers claim that a successful application of modern technology depends to a large extent on its ability to function as a "team player" with human practitioners (Sarter and Woods, 2000), or, in other words, to collaborate with the user (DeKoven and Keyson, 2000). We will need to know which technological innovations can be sensibly applied and how this should be done.

Unfortunately, the development of applied interface ergonomics, which is required in the design of such devices, has not kept up with the pace of these new technological advances. There are hardly any design guidelines available about how to apply new ICT technologies in a user-centred way. Standard display ergonomics does not suffice to make the necessary changes in the selection and display of diagnostic and monitoring data. A traditional user interface, for example, is in general not dynamic through time and does not anticipate learning curves of individual users. Usually, it does not support multiple users in various times and spaces (e.g. through internet) nor does it recognise user errors or discuss the treatment with the user. It is therefore difficult to substantially increase usability in modern systems by applying traditional user interface design.

Aim of this study: The aim of our study is to determine whether, when, and how contextually based intelligent medical systems enhance the quality of interaction between intensive care nurse and the medical systems used. We aim at finding design knowledge about these systems, which should lead to an improvement of the effectiveness, efficiency, comfort, and safety of the medical care provided in an intensive care unit.

We aim on improving all four factors by increasing the ease with which the user communicates with the medical equipment. Future medical systems should be able to interact with the nurses in a more natural way, and therefore be better adapted to the comprehensive nursing process. Furthermore, these systems should be able to understand and anticipate the tasks and subsequent intentions of the intensive care nurse as well as the infuencing constraints posed by the context of use. Hence, extensive insight into the complete work process of the intensive care nurse is needed, before and during the development of these products.

The research is user-driven. By taking a participative ergonomic approach we actively involve end-users (i.e. intensive care nurses) during all stages of research and development. Collaboration with ICU staff takes place in the form of observations, focusgroup-interviews, and user tests. Besides ICU staff, management and procurements teams will also be included, as they are responsible for the short-term and long-term investments in the ICU environment.

Design knowledge should be the result of this study. This can be in the form of design guidelines, insights into user behaviour and context constraints as well as case studies in which future ICT technology is applied to show possible materialised solutions to present usage problems. These technological directions will be evaluated by the four ergonomic targets mentioned.

Position of the research: ID-StudioLab and Intelligence in Products group: This project is part of the research program of the Intelligence in Products group, one of the participating members of the ID-StudioLab in Industrial Design Engineering at Delft University of Technology. Furthermore, the project is placed in an existing framework of collaboration involving the department of Nursing Affairs at the Groningen University Hospital. The premises of this research can best be described by the three statements that ID-StudioLab is built upon (Hekkert et al, 2000): (1) Breakthroughs and innovative research require an interdisciplinairy approach (e.g. product and interface designers, psychologists, sociologists, computer scientists) ; (2) All research efforts are user-driven. Design research must pay attention to the full experience of the user. This experience not only covers the perceptual-motor and cognitive skills of the user, but draws heavily upon the social, cultural, and technological context in which the interaction with the product takes place; (3) All research efforts are designer-driven, i.e. all projects are carried out by designers or directed towards designers.

This paper describes the conceptual framework and the blueprint of our research methods for the introduction of intelligent systems into the intensive care unit.

Conceptual framework: an ecological approach

To organise our research, a conceptual framework was set up based on literature research and observations of several wards at the University Hospital of Groningen and the University Hospital of Rotterdam, Dijkzigt (i.e. thorax ICU, surgical ICU, neurosurgical ICU, and paediatric ICU). The fundamental idea behind our framework is the ecological approach to interface design as proposed by Vicente and Rasmussen (1992) and Vicente (1995). According to this approach the environment has a strong influence on the actions taken by the operator of a system, that is the characteristics and task goals of the user and his work domain interact, and are studied as a whole. This is different from the traditional organismic approach which tends to minimise the contextual influences and attributes skilled behaviour mostly to mental constructs and cognitive processes. Ecological interface design (EID) is a framework for interface design especially suited for systems involving complex human-machine interaction.

The goal of EID is to make the relevant dynamic work domain constraints visible in the interface. As a result of the influence of the dynamic environment, the actions taken by an operator to reach a specific goal will vary. Making the relevant constraining influences visible to the user should provide the user with a better insight into the state of the system and the contextual situation. For our purposes, this means: Making the relevant constraining influences visible to the intensive care nurse should provide her with a better insight into the condition of the patient as well as the situation of the entire ward (e.g. presence of colleagues, day or night, condition of other patients). As a result, the interface should provide her with the information needed to plan her actions more effectively than current equipment does.

To accomplish a certain task goal the nurse will take different actions depending on the current situation and the characteristics of the nurse (e.g. fatigue, skill level). As a consequence the information required to plan and perform these actions is dependent on both these factors. For routine tasks, like replacing a syringe, in a routine situation, lower-order variables such as the values of individual variables could suffice and could even be highly efficient. For more complex tasks, like evaluation of the prescribed treatment and planning subsequent actions, higher-order variables such as system status, situational status, and relationships between variables could be needed to support the nurse in her diagnosis (Effken, 1997), especially when the situation at the ward is chaotic as well and the nurse has to prioritise her actions. The designer has to determine which of the many constraints and subsystems are relevant and should be present in the interface. Subsequently the designer has to determine how to represent these relationships, and when which form of information is needed and should be available to the user. Hence, an elaborate task analysis is needed in which the different sources of constraints are incorporated. This task analysis results in a description of the current situation and a prescription of future situations. The descriptive task model indicates (among others) which information is used and available in present work situations. The prescriptive model describes (among others) which information should be available in which situation.

The (interface) design of ICU equipment according to the ecological approach starts by identifying the dynamic set of environmental constraints on the behaviour of the intensive care nurse. We identified five sources of work domain constraints which influence the interaction between the nurse and the medical system, namely teamwork, the situation on the ward, other medical personnel, especially clinicians, who are obviously extremely important for the nursing process, and the patient being the biological system to be controlled. Besides functioning as a system, the patient also is a seriously ill human being. The intensive care nurse is the main source of information for the patient, and for the relatives. The patient acts, like the relatives, as a passive user of the equipment as well. Besides these environmental influences, the operator characteristics of the intensive care nurse and her task goals have to be taken into account. These are the sources of constraints which have to be considered in defining the interface of future ICU equipment, and therefore define our conceptual framework as shown in figure 1. In the following, a short description is provided of these (constraining) elements.

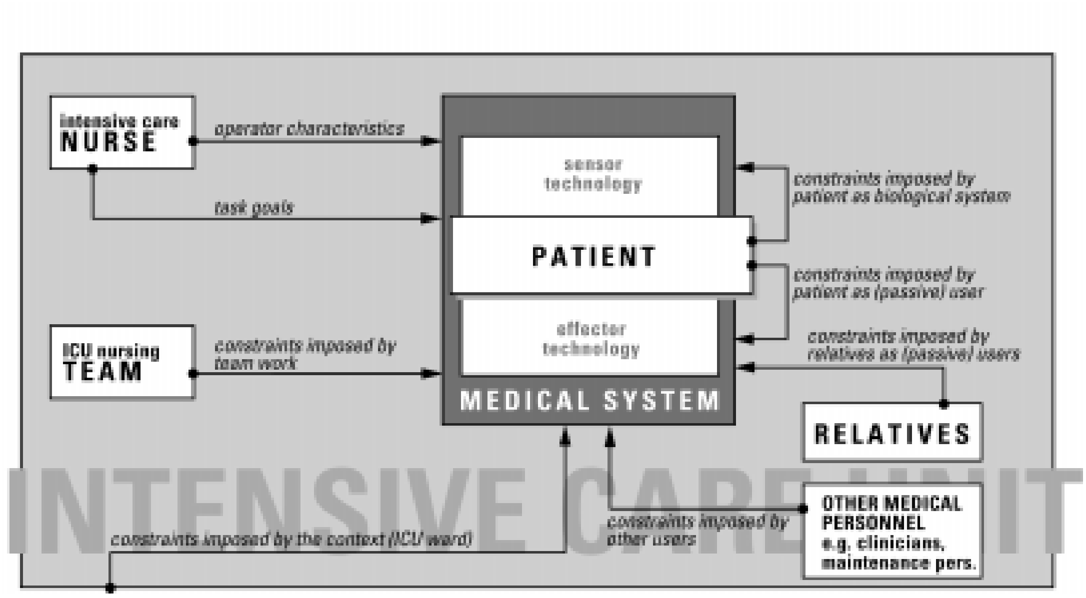


Figure 1 - Conceptual framework
Teamwork and the situation on the ward have a huge influence on the work process of the nurse. The patient is the biological system to be controlled. Besides functioning as a system, the patient also acts, like the relatives, as a passive user of the equipment. Other medical personnel, especially clinicians, are obviously extremely important for the nursing process and therefore must be included.

Medical system: Connected to the patient is the sensor and effector technology. Sensor technology (monitoring equipment) is used to measure and display a number of parameters read from the patient. Some parameters are routinely monitored for every patient (e.g. heart rate, respiration, blood pressure, body temperature, saturation (oxygenation of the blood)). All these measurements can be done either as one-off readings or as a continuous process. Effector technology (treatment or support equipment) is used for interventions to help patients recover from (sudden) changes in their conditional state. The functions that these machines perform range from completely taking over a function of the body to supporting the activity of the patient.

A lack of proper design standards for medical equipment has led to a diversity of interfaces. Bogner (1994) mentions that non-standardisation may be responsible for erroneous actions, as well as incomplete or difficult procedures for operating and maintaining the equipment. Other research confirms much of her findings (e.g. Obradovich & Woods, 1996; Bogner, 1994; Cook et al, 1992).

Groen (1995) concludes that nurses handle the information from the equipment with caution, because they do not always consider it reliable. When the image of the patient is not clear because the information from different sources (i.e. the equipment and the patient) does not converge, the role of the technology is called into question. The information that is observed directly on the patient is judged as more reliable. This set of priorities emphasises that the patient is central to the care process and that the technology is just an aid.

Patient as the biological system to be controlled: The objective of an intensive care unit in general is to control and improve the condition of the patient. In our framework, this implies that in the ICU work domain the patient is the system to be controlled. A patient could be considered as a complex biological system consisting of many highly coupled subsystems. This makes treatment of a patient (i.e. arriving at a diagnosis (problem definition), an etiology (cause or causes), and a prognosis (likely outcome of the possible treatment actions)) extremely difficult. The state of the patient is often unclear, and the precise effects of each treatment are uncertain as well: A certain treatment can solve one problem, but at the same time create a new problem or intensify a problem in another subsystem.

Miller (2000) states that ICU patients are biological systems that operate according to cybernetic rules. Critically ill ICU-patients operate in deranged states that are outside the bounds of homeostasis. This derangement affects their innate biological control system. According to her findings, it is therefore critical that models of ICU patients include and visualise these biological control systems. A user interface should make the underlying structure of the deranged patient visible.

Effkens work (1997) is based on the same premise. She evaluated the ecological approach used in interface design for a haemodynamic monitoring and control task. She also concludes that a patient should be considered as a complex system consisting of many highly coupled subsystems. Her research shows that making the relevant relationships visible in the interface result in a more accurate diagnosis and treatment of clinical problems by experts as well as novices.

Task goals of the intensive care nurse: The nurse’s tasks consist of defining and implementing care plans, execution of the treatment as prescribed by the clinician, monitoring of the patient and operating the equipment (cure). The biggest part of the task consists of evaluating medical regulation. The nurse is continuously inspecting the data provided by the equipment as well as the vital signs of the patient to see if everything is normal; whether all the data is ok, and if not, is intervention immediately required, or is there time to consult the clinician.

The main goal of an intensive care nurse is returning the patient to as healthy a state as possible, by bringing the patient in a state of homeostasis and subsequently maintaining this state. Homeostasis is defined as an internal state of dynamic balance. It is reached when all physiological variables are operating within their normal bounds. This embedded nature of control influences the relation between nurse and patient. Miller (2000) describes this very effective: because the core system of the patient follows its own logic, clinicians have to play two roles. Assuming the role of collaborator with processes tending towards homeostasis, and saboteur of processes tending away from homeostasis.

Another important factor in the use of medical equipment related to the intended tasks is that the nurses are responsible not only for the device operation, but also for the larger performance goals of the overall system. As a result, the system of people and artefacts evolve over time to produce generally successful performance, even if the usability of that system is poor. Cook and Woods (1996) call this adaptation “tailoring processes” or “user tailoring”.

Characteristics of the intensive care nurse: The characteristics and working methods of the intensive care nurse have been the object of several studies. Differences can be identified in skill level concerning information handling and decision making, and in skill level concerning the use of the equipment.

Benner (1992) has distinguished four levels of practice in critical care nursing, based on the Dreyfus Model of Skill Acquisition, shifting from advanced beginner to expert. In the process of becoming an expert, the learner moves from analytical and rule-based thinking to intuition, and from detachment of the situation to involvement. During this process knowledge becomes embedded in practice and can hardly be separated from that practice. These differences are apparent from the way information is handled and from how much information can be handled by a certain nurse. IC-nursing is characterised by unpredictability. Often there are situations in which there is no standard procedure available (unfamiliar situations as defined by Vicente, 1992). Nurses deal with this by setting priorities on the basis of images they construct, both of the individual patient and of the ward as a whole. An expert nurse is capable of dealing with more information than just that of her own patients. She knows how to prioritise and zooms in on information that needs more attention (Groen, 1995). She hypothesizes and calculates the corresponding probability of occurrence, while at the same time overseeing all the relevant evidence. The beginning nurse can only process a small part of the information provided and only in the order learned (rule-based). All information is seen as having equal importance and there is no insight in prioritising. A beginning nurse concentrates on the so-called evidence, but has difficulty with hypothesizing and calculating possibilities. She has to translate her knowledge in the form of rules to concrete action and these actions are susceptible to human error. In addition, these translations take time, while fast action is needed during emergencies.

There is not only a diversity in experience between nurses in providing care, but also in the level of experience with equipment in general and with the actual devices in use at a certain time (e.g. it might be new on the ward). Not only experience gained by practice is relevant. Age is found to be related to general knowledge on how to operate devices. It has an effect on performance which can not be fully compensated by practice. Older users have more problems when operating devices. They have more serious usability problems if devices react inconsistently or do not provide clear guidance to the user (Freudenthal, 1999). For example older users often have problems in applying the principle of 'spatial organisation' (of menus) (Docampo Rama, 2001). These age dependent performance problems are especially relevant in the intensive care unit where, in general, staff is older.

The intensive care unit: The ICU as a work domain is characterised by predictable as well as unpredictable patterns. Most of the time, the pattern of the work process is according to preset tasks and activities (Leys, 2001); dictated by shifts, schedules, visits by other personnel (e.g. fysio-therapists, clinicians), day and night rhythm and equipment procedures (e.g. change of drugs). However, these patterns can suddenly change due to the unstable and therefore often unpredictable status of most patients. Additionally, the number of patients fluctuates strongly. Sudden peaks disappear as quickly as they started and the routine patterns will be picked up again. Nurses have to deal with this unpredictability and try to solve this by prioritising the information they get (from their own patient and the ward as a whole) and hypothesising the consequences.

ICU nursing team: Team play is an important characteristic of ICU nursing. Usually, nurses are responsible for one or two patients. Nurses assist each other with physically difficult tasks, like turning the patient over. They also watch each others' patients when a nurse has to leave the ward temporally. As a consequence, devices are operated collectively. Another important aspect of team play is the trust and reliance nurses must place in each other. They consult each other in doubtful situations and give assistance in crisis situations. Preferably, medical systems should play a role as a team member as well.

The patient and the relatives (as passive users): The intensive care nurse is the main source of information for the patient and the relatives. According to Yen (2001) research has indicated that reducing stress and reducing feelings of isolation have a positive effect on the patient's sense of well-being. A familiar surrounding is an important aspect of the healing and treatment process and eases the acceptance of serious illness. Providing feedback on the treatment process can help reassure the patient. An interface design based on these patient centred care principles can stimulate these psychological effects (Yen, 2001). The design of most modern healthcare facilities ignores these principles, thereby unintentionally reinforcing the patient's idea of sickness. Naturally, the physical comfort of the patient should also be taken into account.

Research methods

It is clear from the previous discussion that our research is based on a user-driven approach as well as a designer-driven approach. These aspects translate into two empirical research methods, participative ergonomics and research-through-design respectively. Participative ergonomics is an extension to classical ergonomics whereby the end-user is actively involved in the product development process (Vink et al., 2001). This method elicits the unique and often latent knowledge of the user. Research-through-design is a method for testing theories in an iterative design cycle (Hekkert et al, 2000). An underlying assumption of this method is that results from experiments using prototypes can be generalised in the form of design guidelines for future products. To organise our research we use the method of grounded theory as an overarching structure (Strauss & Corbin, 1998). The grounded theory method (a method from the social sciences) aims at developing theories from systematically obtained data. These data are acquired in several rounds in which the resulting hypotheses are evaluated and adapted interactively. Within this grounded theory method, the two empirical methods described will be used for gathering the data. Literature is also perceived as data, which is to be analysed and evaluated. A blueprint of our research methods, is illustrated in figure 2.

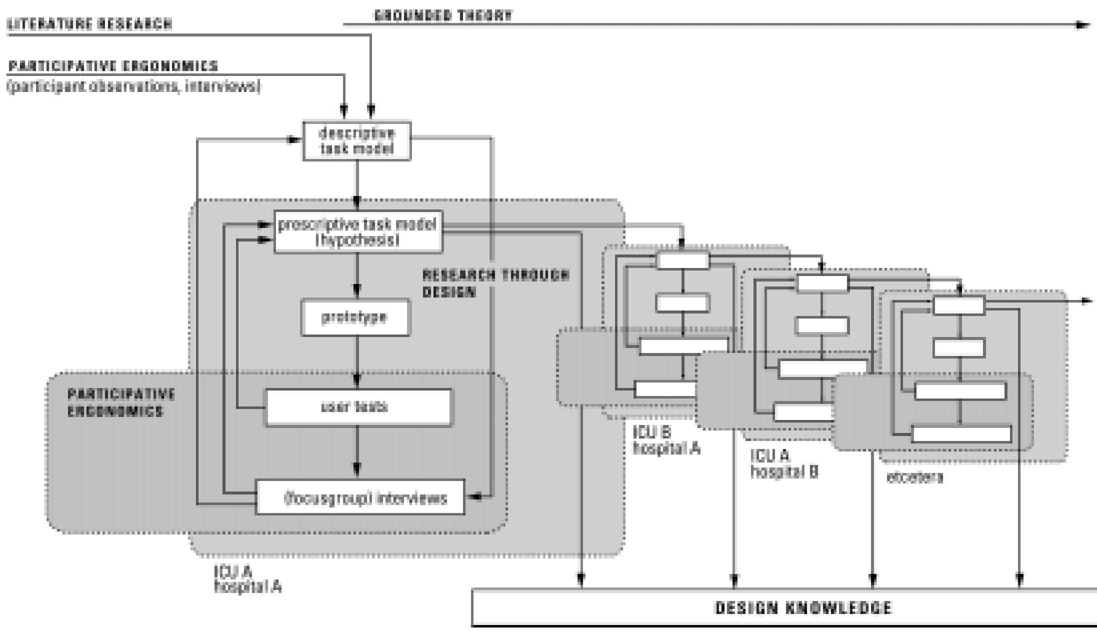


Figure 2 - Method blueprint

The method of grounded theory is the overarching method, the arrow indicating the chronological sequence of several iterations of investigations. Literature research, participative ergonomics and research-through-design are used for gathering data and testing the hypotheses. The research-cycle will be repeated across several ICU's and hospitals.

Taking our conceptual framework as basic principle, we have started with an elaborate task analysis of the nurses' characteristics and their task world. The ecological approach has important implications for the choice of task analysis method, given the emphasis on analysing the environment. The variability in the actions taken by the nurse as a result of the influences from the environment, should be taken into account by the method used. Therefore, the task analysis methodology should provide descriptions of at least the following three classes of constraints (Vicente, 1995): (a) the functional problem space in which the nurses behaviour takes place, (b) the tasks that are to be accomplished by the nurse, and (c) the set of strategies that nurses can use to carry out these tasks. We use Groupware Task Analysis (GTA) and DUTCH (Welie, 2001). The fields of application of these methods is when either the current way of performing tasks is not considered optimal, or the availability of new technology is expected to allow improvement over current methods. Moreover, GTA puts an emphasis on studying a group or organisation and their activities.

We have started with analysing the current task situation resulting in preliminary descriptive models. These descriptive task models will be presented to focus groups of intensive care nurses and according to their reactions adjusted and elaborated. Subsequently, future task situations are envisioned resulting in prescriptive task models. Again, these models will be developed in collaboration with the end-users.
These prescriptive task models (hypotheses) form the basis for initial case studies. Equipment for the ICU will be developed and tested using prototypes developed to such a level that the subjects can actually experience the interaction. The prototype(s) will be tested using participant observations, user testing in real or simulated environments and (focus group) interviews. Results of these tests lead to new design knowledge (theories) or refinement of the research issues (hypotheses). According to the grounded theory method this process will be repeated several times. Investigation will take place across several intensive care units and across multiple hospitals, thereby identifying possible local effects.

The final result of this research should be design knowledge, in the form of design guidelines, insights into user behaviour and context constraints as well as case studies in which future ICT technology is applied to show possible materialised solutions to present usage problems. Hopefully this research will make both product developers as well as hospital staff (i.e. the intensive care nurses, ICU management, and ICU procurements teams) more aware of usability problems in the ICU.

Acknowledgments

The authors thank prof.dr. C.J. Snijders, prof.dr. H. de Ridder and dr. D.V. Keyson for their helpful comments.

References

Benner, P., Tanner, C., Chesla, C. (1992). From beginner to expert: gaining a differentiated clinical world in critical care nursing. Advanced Nursing Science (14)3:13-28.

Bogner, M.S. (1994). Human Error in Medicine, Lawrence Erlbaum Associates, New Jersey.

Cook, R.I., Woods, D.D., Howie, M.B., Horrow, J.C., Gaba, D.M. (1992). Unintentional Delivery of Vasoactive Drugs With an Electromechanical Infusion Device. Journal of Cardiothoraric and Vascular Anesthesia (6)2:238-244.

Docampo Rama, M. (2001). Technology Generations handling complex User Interfaces, PhD thesis, Eindhoven University of Technology, the Netherlands.

DeKoven, E. and Keyson, D.V. (2000). Designing collaboration in consumer products. Proceedings of the Fall 2000 AAAI Symposium.

Effken, J.A., Kim, N.-G. and Shaw, R.E. (1997). Making the constraints visible: testing the ecological approach to interface design. Ergonomics (40)1:1-27.

Freudenthal, A. (1999). The design of home appliances for young and old consumers, PhD thesis, Delft University Press, Delft, The Netherlands.

Groen, M. (1995). Technology, Work and Organisation. A study of the nursing process in intensive care units, PhD thesis, Maastricht Universitaire Pers, Maastricht, the Netherlands.

Hekkert, P.P.M., Keyson, D., Overbeeke, C.J., & Stappers, P.J. (2000). The Delft ID-StudioLab: Research for and through Design. Proceedings of the Symposium on Design Research in the Netherlands:95-103.

Leys, M. (2001). Technologie, organisatie en verpleegkunde op intensive care en neonatologie afdelingen. Verpleegkunde (16)4:197-207.

Miller, A., Sanderson, P. (2000). Modeling “deranged” physiological systems for ICU information system design. Proceedings of the IEA 2000/HFES 2000 Congress:245-248.

Obradovich, J.H. and Woods, D.D. (1996). Users as designers: How People Cope with Poor HCI Design in Computer-Based Medical Devices. Human Factors (38)4:574-592.

Sarter, N.B. and Woods, D.D. (2000) Team Play with a Powerful and Independent Agent: A Full-Mission Simulation Study. Human Factors (42)3:390-402.

Strauss, A. and Corbin, J. (1998). Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory, Sage Publications Inc, Thousand Oaks, California.

Vicente, K.J. and Rasmussen, J. (1992). Ecological Interface Design: Theoretical Foundations. IEEE Transactions on Systems, Man, and Cybernetics (22)4:589-606.

Vicente, K.J. (1995). A Few Implications of an Ecological Approach to Human Factors. In J.Flach, P. Hancock, J. Caird and K. Vicente (eds). Global Perspectives on the Ecology of Human-Machine Systems:54-67 Lawrence Erlbaum, Hillsdale, NJ.

Vink, P., Pennock, H., Scheijndel, P. van and Dort, B. van (2001). Verschillende rollen bij het toepassen van participatieve ergonomie. Tijdschrift voor Ergonomie (26)3:19-23.

Welie, M. van (2001). Task-Based User Interface Design, PhD thesis, Vrije Universiteit, Amsterdam, The Netherlands.

Yen, C.C. and Wooley, M.S. (2001). Affective design solutions for Medical Equipment through PCC Principles. Proceedings of The International Conference on Affective Human Factors Design 2001:289-296.

Evaluation of the Surgical Process during Joint Replacements.

Joanne JP Minekus, Jenny Dankelman

Man-Machine Systems Group, Department of Medical Technology and Mechanics, Fac. of Design, Engineering and Production, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands. Email:j.minekus@wbmt.tudelft.nl

Abstract

A task analysis method has been developed to evaluate joint replacements. Video recordings were made of the surgical process during humeral head, total shoulder and elbow joint replacements. The actions of the surgeon, the scrub technician and the assistant were analysed off-line using a thesaurus of actions. The efficiency was defined as the relative amount of goal-oriented actions of the surgeon. The efficiency of the procedures varied between phases and was on average 61 percent in all three procedures. The main task of the scrub technician was arranging instruments and of the assistant holding clamps. The main shortcomings observed during joint replacements were repeated actions caused by the inability to align the prosthesis at once, and waiting times caused by cement hardening and caused by searching for instruments by the scrub technician.

Keywords: task analysis, surgical team, joint replacement

Introduction

In surgery, operative procedures are normally evaluated with respect to the post-operative results. Only few task-analysis studies have been achieved to evaluate the actual surgical process. Those studies were mainly performed in laparoscopic surgery and show that the operative time could be decreased and the efficiency improved (den Boer,K.T., Dankelman,J., Gouma,D.J., and Stassen,L.P., 2002; Joice,P., Hanna,G.B., and Cuschieri,A., 1998). By humeral head replacements a large variations between procedures and a large amount of shortcomings was observed (Minekus,J.P.J., Rozing,P.M., Valstar,E., and Dankelman,J., 2002). There is a demand for more efficient and shorter procedures with less people at the table because of the increased cost, the patient waiting list and the shortage of operating nurses. The goal of our study is to develop a method to evaluate joint replacements by analysing the actions of the surgeon, the assistant and the nurse. From these measurements, the shortcomings and difficulties of the procedure can be determined. We used humeral head, total shoulder and elbow joint replacement in this research.

Method

Surgical procedure: By a humeral head replacement (HH), the upper arm part of the shoulder is replaced. By a total shoulder replacement (TS), both the humeral head and the glenoid (part of the shoulder blade) are replaced. By an elbow joint replacement (TE), both the humeral and ulnar (forearm) part of the elbow joint are replaced.

The surgical procedure of a joint replacement consists of several successive phases:

– *preparation phase*: opening the skin and overlying tissues to reach the joint.
– *bone phase*: preparing the bones till the test prosthesis fits.
– *test phase*: testing the two prosthesis parts together (only TS and TE).
– *prosthesis phase*: unwrapping and inserting the real prosthesis.
– *closure phase*: closing the wound.

Surgical team: The surgical team consists of 5 team members:

– *anaesthesiologist:* checks the vital functions of the patient.
– *nurse:* unpacks new instruments and the prosthesis.
– *surgeon:* performs the actual surgical process.
– *scrub technician:* responsible for the instruments and eventually helps the surgeon.
– *assistant:* helps the surgeon by holding clamps, sucking blood and preparing tasks.

This study will focus on the team members working in the sterile area: the surgeon, the scrub-technician and the assistant.

Table 1: Thesaurus of functions for the surgeon

	Function	Definition
Goal oriented	Preparing	Dissection using e.g. a knife, a rasp, or a saw.
	Alignment and inserting of the prosthesis	Determination of the position of the prosthesis and placement of the prosthesis.
	Suturing	Placement of sutures or the drain.
Additional	Stop bleeding	Checking for bleedings and stopping them using e.g. coagulating, or swamping.
	Observing	Watching the wound, palpating or moving the arm.
	Exposing	Placement of hooks to expose the humeral head.
	Waiting	Actions that do not contribute to the procedure.
	Miscellaneous	Actions that could not be identified or classified within the other functions.

Time-action analysis: Video recordings of the procedure were made using two cameras, one giving an overview of the total operation field and one giving a detailed view of the hands of the surgeon (den Boer,K.T., Dankelman,J., Gouma,D.J., and Stassen,L.P., 2002; Minekus,J.P.J., Rozing,P.M., Valstar,E., and Dankelman,J., 2002). The images with sound were recorded simultaneously using a mixed device and were analysed off-line. The recordings did not interfere with the surgical process; the medical ethical committee of the Leiden University Medical Center approved the research.

The recordings were analysed three times. First, the actions performed by the surgeon were analysed using a thesaurus of 68 actions (Minekus,J.P.J., Rozing,P.M., Valstar,E., and Dankelman,J., 2002). The tokens in the taxonomy have been discussed with the surgeon. In a previous study, performed in laparoscopic surgery, the videos were analyzed by three different observers using a strictly defined taxonomy, showing that there was no difference in results between observers. The actions were grouped to eight different functions (Table 1). The functions preparing, aligning and inserting the prosthesis, and suturing contribute directly to the advancement of the procedure and are, therefore, classified as goal-oriented functions (Minekus,J.P.J., Rozing,P.M., Valstar,E., and Dankelman,J., 2002). The percentage of goal-oriented functions of the surgeon is used as a measure of the efficiency of the operative procedure.

Secondly, the actions of the scrub technician and the assistant were analysed using a thesaurus of functions (Table 2). The thesaurus of the surgeon was expanded with two functions and some functions got a broader meaning. The actions of all three team members can be directed towards the patient (actions like preparing and aligning), towards the instruments (actions like searching for instruments) or towards each other (actions like communication and teaching). Actions are classified as directed towards instruments if the person is solely working with instruments, e.g. cleaning them, or searching for instruments.

Thirdly, the procedures were analysed on the occurrence of shortcomings (Minekus,J.P.J., Rozing,P.M., Valstar,E., and Dankelman,J., 2002). In most procedures, repetitions and corrections are needed due to the complexity of the surgical approach, the limitations of the instruments, or the experience of the surgeon. These repetitions and corrections are called shortcomings. The shortcomings observed during the procedures were grouped into three classes: repeated actions, waiting and miscellaneous.

The functions of the surgeon and the shortcoming are evaluated in 8 humeral head replacements, 4 total shoulder and 11 elbow joint replacements. The actions of the scrub technician and the assistant are evaluated in 10 elbow joint replacements.

Table 2: Thesaurus of functions for the scrub technician and assistant.

Function	Definition
Preparing	Dissecting using e.g. scissors, saws or drills.
Prosthesis	Making cement, unpacking the prosthesis or helping to place or align the prosthesis
Suturing	Placing sutures or the drain.
Stop bleedings	Checking for bleedings and stopping them using e.g. coagulating, or swamping.
Observing	Watching the surgeon or the operative process
Exposing	Placing and holding hooks to expose the joint.
Instruments	Getting, positioning or cleaning instruments.
Helping	Helping the surgeons by actions for, which a third hand is needed.
Waiting	Actions that do not contribute to the procedure.
Miscellaneous	Actions that could not be identified or classified within the other functions.

Results

All three procedures showed a large variation in duration (Figure 1). The total shoulder replacement has the largest duration, because it is the most complex procedure. The humeral head replacement has the shortest duration, because only one joint part is replaced. The total elbow has the shortest openings phase, because the joint is more superficial and has, therefore, a shorter approach. The total shoulder replacement has the longest bone preparation phase, due to the difficult preparing of the glenoid. The total elbow replacement has the longest prosthesis insertion phase, because two prostheses are placed successively with cement, causing two waiting periods of 10 minutes for the cement to harden. The glenoid component in the total shoulder is also fixated with cement, causing only one waiting period of 10 minutes; the humeral head is placed without cement.

Figure 1: Average duration of all phases in 5 total shoulder replacements, 8 humeral head replacements and 11 total elbow replacements.

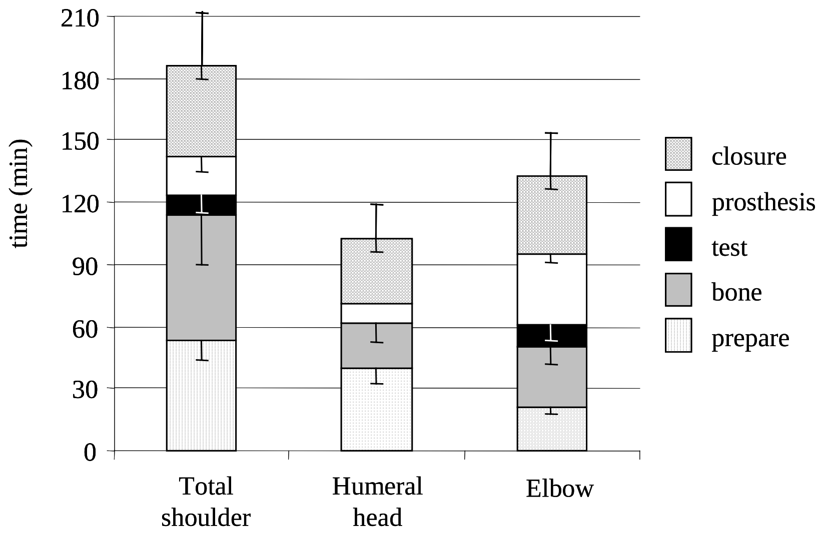
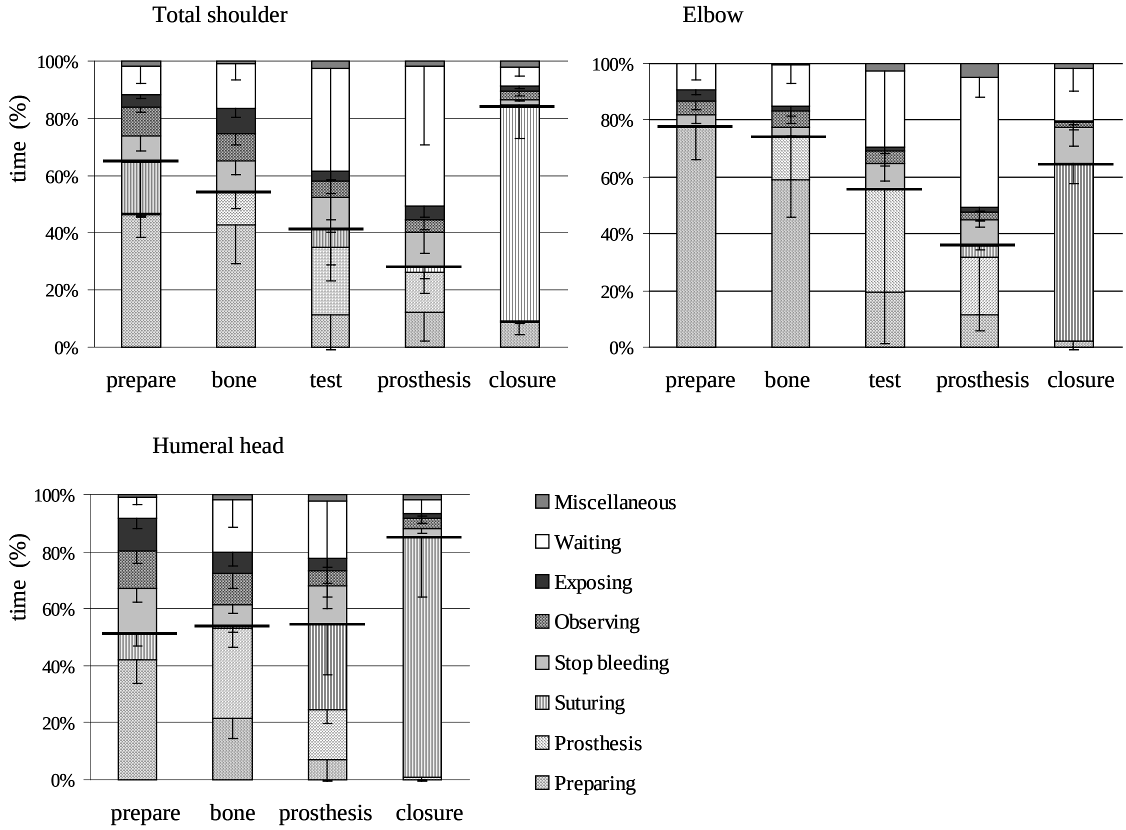


Figure 2: Relative duration of functions of the surgeon during total shoulder, humeral head and total elbow replacements. The goal-oriented functions are indicated below the dotted lines. On the X-axis the phases are indicated. (Also available online: http://www.dcs.gla.ac.uk/~johnson/eam2002)

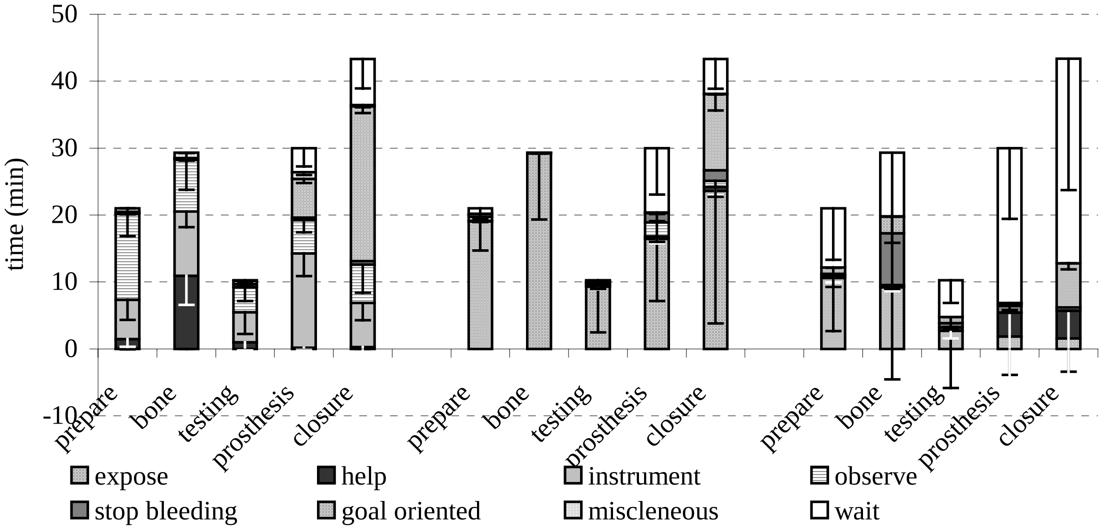


The surgeon: mainly performs goal-oriented functions. The amount of goal-oriented functions depends on the phase of the procedure (Figure 2) and is on average 61 percent in all three procedures. The main cause for non goal-oriented functions is waiting. The surgeon's actions are mainly directed (70%) towards the patient e.g. operating on the patient.

The *scrub technician:* is mainly focused on the instruments; making them ready for use, cleaning them, putting them in order and giving them to the surgeon (Figure 3). To know which instruments the surgeon will be using next, the scrub technician also spent much time observing the surgeon. During the waiting periods of the surgeon, the technician puts the instruments in order. Forty percent of the scrub technician's actions are directed towards the surgeon, 30% towards the instruments, 20% towards the patient and 10% percent is spent on waiting.

The assistant: sometimes uses his hands for two different functions, therefore, these are analysed separately. The main function of the assistant is to hold clamps for exposure (Figure 3, middle bars). Besides holding clamps, the assistant helps the surgeon with preparing, stopping bleedings and suturing. Eighty percent of the actions are directed to the patient, 10 towards the surgeon and 10 percent is spent on waiting. The assistant was a junior surgeon, who had to learn the procedure. Most teaching occurred by observing the surgeon, sometimes the surgeon explained his actions while continuing work

Figure 3: Average duration of functions of the scrub technician and the assistant during ten elbow replacements. On the X-axis the phases are indicated. (Also available online: http://www.dcs.gla.ac.uk/~johnson/eam2002)



The main shortcomings: observed during the procedure were waiting and repeated actions (Table 3). A shortcoming did not imply a complication: in none of these procedures a per-operative complication occurred. The main causes for waiting are the cementing process (10 minutes hardening); unpacking the prosthesis (because the size is determined during the procedure) and waiting for the scrub technician to find the right instruments. Both, waiting for cement and unpacking the prosthesis, happened in all procedures during the prosthesis phase. Waiting for the scrub technician occurred several times in all procedures, especially during the bone and testing phases and is the main cause for the large number of waiting times.

The main cause for repeated actions is the difficult alignment of the prosthesis. To align prosthesis, several subsequent steps depending on the prosthesis have to be performed. For most steps, special alignment instruments exist, which should help the surgeon to align the prosthesis correctly at once. But still up to 6 refinements are needed in 1 step. Some of these refinements were even made without the help of an alignment instrument. In all procedures, at least one repeated action was needed.

Table 3: The total duration of shortcomings in minutes and, between parentheses, the number of shortcomings per operative procedure.

Shortcoming Phase	**waiting**	**repeated action**	**miscellaneous**
Total shoulder	29.4 (112)	22.0 (8)	1.0 (2)
Humeral head	11.1 (53)	7.1 (6)	1.4 (2)
Total elbow	31.7 (76)	17.1 (10)	1.0 (2)

Discussion

Surgical procedures are normally evaluated mainly with respect to post-operative results and complication rate (Alund,M., Hoe-Hansen,C., Tillander,B., Heden,B., and Norlin,R., 2000). These analyses, however, provide hardly any insight into the problems of the actual complex per-operative process. Video analysing of the surgical procedure does give insight into this process. Using task analysis, we showed large variation between procedures and a huge amount of shortcomings during the placement of total elbow, total shoulder and humeral head prostheses. Besides differences in the patient condition and the surgical team, the number of needed repeated actions is the major cause for the observed variations.

Our evaluation method does give insight in the actual surgical process, but it should be used with care (Minekus,J.P.J., Rozing,P.M., Valstar,E., and Dankelman,J., 2002; Sjoerdsma,W., Meijer,D.W., Jansen,A., den Boer,K.T., and Grimbergen,C.A., 2000). The main parameter of time-action analysis, time, is not directly related to the functional outcome. A shorter and more efficient surgical procedure may have less post-operative results and is, therefore, not favourable. The evaluation may also create problems of interpretation, it might look like the nurse is searching when he/she is actually thinking or performing some other cognitive function. In our method, we cannot distinguish cognitive processes from the observations. Finally, the reason of the surgeon for a certain approach or action cannot be determined. By discussing our data with the surgeon, we found that the surgeon was aware of certain limitations and inefficiencies. Some repeated actions were even made on purpose because it improved the accuracy. The surgeon was not aware of all limitations and inefficiencies and after recognising them; the team is trying to reduce them. Therefore, for a good interpretation of the results, interaction with the surgeon is very important.

This study has only evaluated one surgical team using a specific approach and a specific prosthesis. Several approaches are possible to both the shoulder and the elbow joint and different prostheses have different alignment instruments. Different approaches and prostheses have different advantages and disadvantages during the procedure. Also, different operative teams have different cultures. Therefore, we have also analysed some operative procedures by other surgeons. In these procedures, comparable shortcomings and problems could be found.

The two main shortcomings, repeated actions and waiting, have also been found during knee replacements (Dunbar,M.J. and Gross,M., 1995). Repeated actions occur due to the inability to align the prosthesis correctly at once. The number of repeated actions is smaller in knee replacements, because this procedure is more standardised and has different alignment instruments. The waiting time in humeral head, total shoulder, total elbow replacement, and in knee replacements is mainly caused by unwrapping instruments, changing instruments and the cementing process. The comparable shortcomings in different joint replacements show the need of improving the efficiency of joint replacements by for example better alignment instruments, computer guided surgery and faster cementing techniques.

During an elbow joint replacement, most actions were directed towards the patient or the instruments and most communication occurred, e.g. teaching, while the surgeon continued working. Both the surgeon and the scrub technician were experienced and, therefore, not much communication was needed; their tasks were mainly rule based (Rasmussen,J., 1983). A less experienced team may need more communication to discuss the needed instruments or the alignment of the prosthesis. The assistant was an inexperienced junior surgeon and he used these procedures to learn. He had a mainly skill based task, holding clamps given by the surgeon. Possibly a more experienced assistant can take over some actions of the surgeon, thereby reducing the operative time, but further study is needed to confirm this.

Dutch hospitals have a shortage of operating nurses, causing an increase in waiting lists for patients. In non-academic hospitals, the assistant may also be an operating nurse. One person less in the operation theatre may be a partial solution to this shortage. The task of the assistant is mainly holding clamps, which may be done by a technical device. His tasks are then reduced to helping the surgeon and sucking blood, which will take approximately 45 minutes in a procedure. The task of the scrub technician is to cover up the instruments, which is quite complicated because of the large amount of instruments, but the timing is not critical and most work can be done in advance. Also more efficient instrument-tables can be developed, whereby the surgeon can get his own instruments and the non-sterile nurse can do the cleaning after use. This will reduce the work for the scrub technician, so the technician can take over tasks of the assistant. But if the assistant is left out, the flexibility decreases; small problems, now easily solved by the scrub technician, can become large and time consuming. Also, it will become harder to learn the procedure for inexperienced scrub technicians and assistants. Therefore, reducing the number of team members is not advisable yet for elbow and shoulder replacements. For more standard procedure, like hip and knee replacements, it may be possible to reduce the number of team members, but more research is needed to confirm this.

In summary, we developed a method to evaluate the per-operative surgical procedure. This method can be used to analyse the shortcomings of procedures. The main shortcomings of both elbow and shoulder joint replacements are repeated actions and waiting. The large variation between procedures showed that elbow and shoulder joint replacements are not standardised procedures. In the future, this method can be used to investigate whether new instruments have improved the per-operative process.

Reference List

Alund,M., Hoe-Hansen,C., Tillander,B., Heden,B., and Norlin,R. (2000). Outcome after cup hemiarthroplasty in the rheumatoid shoulder: a retrospective evaluation of 39 patients followed for 2-6 years. Acta Orthop Scand. (71) 2: 180-184.

den Boer,K.T., Dankelman,J., Gouma,D.J., and Stassen,L.P. (2002). Peroperative analysis of the surgical procedure. Surg Endosc. (13) in press.

Dunbar,M.J. and Gross,M. (1995). Critical steps in total knee arthroplasty; a method of analysing operative procedures. Int Orthop (19) 265-268.

Joice,P., Hanna,G.B., and Cuschieri,A. (1998). Errors enacted during endoscopic surgery--a human reliability analysis. Appl.Ergon. (29) 6: 409-414.

Minekus,J.P.J., Rozing,P.M., Valstar,E., and Dankelman,J. (2002). Evaluation of humeral head replacements using time-action analysis. J Shoulder Elbow Surg (accepted).

Rasmussen,J. (1983). Skills, rules, and knowledge; signals, signs and symbols, and other distinctions in human performance models. IEEE transactions on systems, man, and cybergenetics. (13) 3: 257-266.

Sjoerdsma,W., Meijer,D.W., Jansen,A., den Boer,K.T., and Grimbergen,C.A. (2000). Comparison of efficiencies of three techniques for colon surgery. J Laparoendosc.Adv.Surg.Tech.A (10) 1: 47-53.

Human Machine Issues in Automotive Safety: Preliminary Assessment of the Interface of an Anti-collision Support System

P.C. Cacciabue, E. Donato, S. Rossano

European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen
21020 Ispra (Va), Italy.
pietro.cacciabue@jrc.it, *http://humanfactors.jrc.it/*

Abstract: This paper considers the impact of Human Factors on the design of in-vehicle components. The peculiarities of the automotive environment with respect to the aviation domain are identified and the specific requisites for designing in-vehicle interfaces are discussed. Particular attention is dedicated to the User Centred Design approach. A number of special issues derived from the use of automation and Information Technology means are also reviewed.

Keywords: Design of Human Machine Interfaces, Automation, User Cantered Design.

Introduction

More than half a million people are killed world-wide each year in traffic crashes and about one person in 200 of the world's population dies from injuries received in traffic accidents (European Parliament, 1998). The European Commission considers that injuries and fatalities caused by road accidents are comparable to the effects of an annual medium-sized war and make road transport the most dangerous means of transport (AAA, 2001; European Transport Safety Council, 1995; European Parliament, 1998).

Several kinds of countermeasures have been studied and adopted in European Union in order to improve road safety. This has produced some good results (Eurostat, 2000): in the last years the number of road accidents fatalities has decreased, passing from 56,400, in 1990, to 43,500 in 1996 and to 42,600 in 1998. Nonetheless, the number of road accidents is still unacceptable. This effort continues and new and more effective road safety solutions are studied. The European Commission has set as a primary objective the reduction of the number of fatalities on road accidents to 27,000 by the year 2010 (European Parliament, 1998).

With any doubt, the main cause of accidents in automotive environment, and this holds also for most technologically advanced domains, is the so-called "Human Factor" (HF). There is, nowadays, a widespread debate about the scope of the analysis on Human Factors. Some authors consider that accident investigations should cover much more than simply the primary actors (drivers, passengers and pedestrians) involved in road accidents, looking more closely at higher socio-technical layers and organisations, such as manufacturers, designers, regulators, safety authorities and national cultures.

On the other hand, one has to be careful not to expand too much the already fuzzy and very frequently utilised connotation associated with the term "human factor", hence embracing in it the totality of causes of accidents. In this case, indeed, the exasperated use of the expression would render useless the work of many specialists of this domain.

Anyhow, independently of this debate, which defines mainly the dimension of the problem, it can be argued that the way forward to reduce road accidents is to operate on the essential contribution to hazards derived from the interaction of humans and vehicles. For these reasons, the focus of all measures and devices conceived and developed for improving road safety consist primarily on the way to support drivers in controlling vehicles and avoiding accidents, and, on a second level, to limit consequences of accidents for passengers and environment.

In this paper, we will consider the impact of HF on the design of new in-vehicle components (Cacciabue et al., 2001). We will firstly consider different kinds of components, usually subdivided in "active" and "passive" safety devices and the peculiarities of the automotive environment with respect to the aviation domain. We will then focus on the main requisites that need attention for designing interfaces of safety devices. In particular, we will focus on the combination of three existing paradigms, i.e., the Supervisory

Control model, the Systems' Usability principle, and the User Centred Design approach. We will develop an iterative procedure that enables designers to integrate them throughout the whole design process. We will conclude with a short review of some special issues that need particular attention and represent major shortcomings of the extensive use of automation and Information Technology means.

Safety Systems

Over the last years, car manufactures addressed their efforts to develop new and more effective in-vehicle safety components, with the intention of decreasing the rates of road accidents. These are usually distinguished in two different groups: "active" and "passive" safety systems.

Active safety systems help to prevent crashes by providing the driver with better means for controlling road scenario and avoiding hazards. Active safety components should help the driver in recovering control of the vehicle, by operating on the car components. Unfortunately, even though active safety systems can help in reducing the chance of a crash, not all crashes are avoidable. Once an accident become inevitable, passive safety systems aim at protecting car occupants and at minimising consequences. Therefore, passive safety systems are designed to mitigate or reduce injury and damage to vehicles, occupants and persons external to the vehicle, in a crash.

The general subdivision in active and passive safety system allows a very broad categorisation between systems that aim at preventing and recovering from collisions (active systems), and systems that aim at mitigating or minimising injury to the vehicle and its occupants (passive systems). However, this definition does not to give any visibility or consideration to the enormous variety of means and features that are currently implemented in vehicles, thanks to sophisticated electronic devices and automation. Moreover, from a Human Factors perspective, it makes difficult to clearly identify the role and tasks of humans in managing and interacting with a safety device. Therefore, a more refined subdivision needs to be considered that enables to take into consideration the progress of information technology and indicates the way in which future development may be expected, with respect to safety systems of vehicles of previous generation of technology.

A more complete and sophisticated definition of safety systems can be developed, starting from the paradigm of Automation and Supervisory Control of Sheridan (1992) that distinguishes between different sub-groups of active and passive systems (Figure 6).

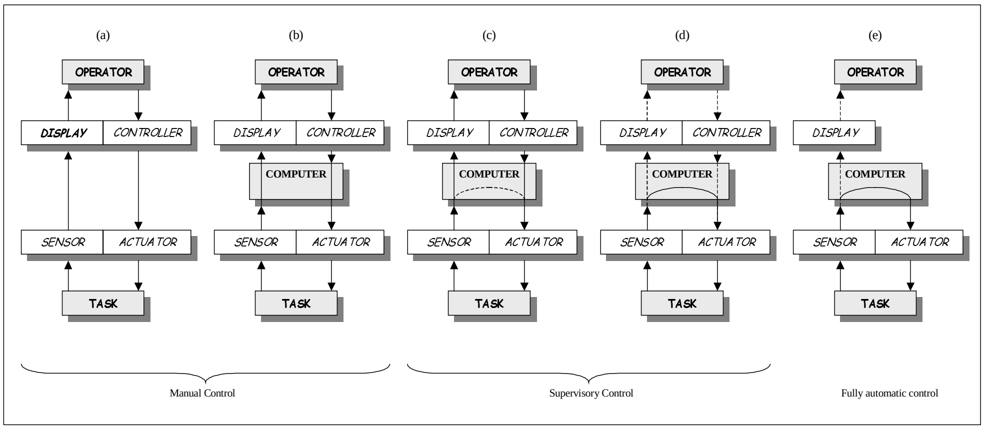


Figure 6 - Paradigms of control process in technological settings (from Sheridan, 1992).

At the first level ("manual control"), the system is fully controlled by the operator, or the operator is supported by a computerised system. At the next level ("supervisory control"), the computer becomes the major actor in controlling the system: the operator may assume direct control or may share control of some variables and processes with the computer. At the highest level, named "fully automatic control", the human operator is totally separated from the control and operates only during the initial planning of a procedure and activity to be carried out by the automated system. After this initial phase, the operator has no control over the automation, and can intervene only by turning-off the automation and restart the whole planning process ("pull the plug out").

Active Safety Systems: Focusing on *active safety systems*, it is possible to distinguish four different types of instruments, which are either indicators or controls (Table 3):

- "*Manual control systems*" that are completely controlled by the driver, such as the "brake pedal" or "sun visor".
- "*Information/alert support systems*" that support the driver with warning and information messages, such as "Anti-collision warning systems".
- "*Activity support systems*" that can be activated or stopped by driver, but, once activated or set in stand-by mode, they operate completely independent from the driver, such as the "Cruise Control" System. These systems can be made inactive by the driver or become ineffective as soon as he/she decides to return to a direct manual control of the vehicle.
- "*Fully autonomous systems*" that operate in complete autonomy and the driver can only observe the performance of the system. Examples of this sub-group are the "Electronic Suspensions". It is important to note that the systems classified in this category bear an autonomy that is even more powerful that the highest level automation envisaged by Sheridan, as they can not be made inoperative in any circumstances except by a laborious and complicated intervention on the electronic control system, carried out only in specialised workshops.

Table 3 - Classification of some Active Safety Systems and Supervisory Control

Active safety systems	**Manual Control System**	**Information/ Alert Support System**	**Activity Support System**	**Fully Autonomous System**
Active Brake System			✓	
Active Body Control (suspensions)				✓
Anti-skid			✓	
Blind Spot Monitoring		✓		
Brake pedal	✓			
Brake Assistant System			✓	
Electronic Stability Control				✓
Electronic Shock absorber				✓
Lane keeping and warning system			✓	
Lights/alarms (brake, headlight, ..)		✓		
Park Assistant		✓		
Parking Brake Driver			✓	
Warning systems (obstacles)		✓		
Systems monitoring driver's state		✓	✓	
Stabiliser			✓	✓
Sun Visor	✓			
Cruise Control			✓	

Passive Safety Systems: In the case of *passive safety systems*, it is possible to distinguish three different types of structural elements and safety devices (Table 4)

- Static systems, substantially represented by the structural elements of the vehicle (bumper, monocoque and so on).
- Dynamic systems, i.e., components activated automatically by a sensor without any driver interaction (air-bag of present generation).
- Fully autonomous systems, i.e., systems able to fit specific drivers (air bag of next generations).

Table 4 - Classification of Passive Safety Systems and Supervisory Control

Passive Safety Systems	**Static**	**Dynamic**	**Fully autonomous**
Airbag		✓	
Airbag of next generation			✓
Seat-belts		✓	
Children protection	✓		
Bumper	✓		
Headrest	✓		
Fire prevention			✓

Overview of new classification of active and passive systems: When one observes the classification of present active safety systems in Table 3, it is possible to remark that there is a tendency to go beyond systems controlled by drivers ("*Activity support systems*"), towards safety devices that are completely independent ("*Fully autonomous systems*"). The driver can only accept and adapt to their presence and performances. This trend to exasperated automation may eventually become harmful for the whole safety of the vehicle and we will observe a trend towards "de-automation" with the objective to make more human-centred the overall process of driving.

With regard to passive safety systems (Table 4), almost all systems are identified as "static", because they are structural parts of the vehicle and they can not be labelled as "intelligent systems". Therefore, it is not necessary to refer to progress in information technology. However, the tendency towards independent and autonomous systems exists also in the case of passive safety systems. As an example, in new generations of vehicles, manufactures tend to produce "intelligent" air-bags systems, able to adapt to drivers' body in order to enhance effectiveness and reduce harm caused by their blast.

The contemporary tendency towards the implementation of sophisticated systems and high automation, can be accepted and become successful only if designers take in due consideration the peculiarities of the automotive environment and apply an appropriate methodology for considering the Human-Machine Interaction in all its aspects.

Prior to discussing these issues in detail, it is important to consider the peculiarities of the automotive environment, especially with respect to aviation. This will enable to put into perspective the amount of technology and methods that may be transferable from one domain to the other, and the necessary adaptation required.

Peculiarities of the Automotive Domain

In the last years, a considerable expertise grew up in designing human-machine interfaces for sophisticated devices in the aviation, medical, nuclear and military fields. Therefore, the possibility to transfer some standards and recommendations from these domains to other areas was considered. In particular, the aviation domain, usually considered very advanced in designing and developing automated systems and Human-Machine Interfaces, is very commonly considered for borrowing technological concepts.

However, after an initial enthusiasm, nowadays there is more prudence in transferring concepts between different technological domains, especially in the case of the automotive environment (NHTSA, 1993). In particular, it is important to point out the principle differences exiting between aviation and automotive fields in terms of users, temporal demands, working and social contexts and system integration (Leibowitz, 1988; Harris and Smith, 1999):

- *Users*
  - Pilots must be approved not only within the category of aircraft for which they are licensed, but also for the type of aircraft that they fly; while drivers are free to drive any motor vehicle, which falls within the broad category that their license covers.
  - Training and retraining is compulsory during the working life of a pilot and every time he/she changes type of aircraft. Training includes also specific courses on human factors (non-technical training), in addition to airmanship skill and performance ability. These courses support pilots in managing difficult or risky situations from a Human Factor perspective (e. g., communication, leadership, team work, stress and workload, etc.). No such type of training or re-training is performed in the automotive environment, not even for professional drivers.
  - The flying licence validity depends on periodic medical checks and flying trials; while the driving license is valid almost for all the life, and only basic medical checks are required over very long periods of time, e.g., several years, depending on age of drivers.
  - Specific ratings are required to fly either at night, in poor visibility and in certain categories of airspace; while drivers are able to drive anywhere, at any time of the day and in any conditions.
- *Temporal demands*
  - Sky environment is not a time critical medium because the flying situation usually doesn't change rapidly, e.g., tens of seconds in the most critical cases of mid-air collisions; while, on the road, the driving situation may change very rapidly, e.g., few seconds.
- *Working context*
  - In the sky there are few "physical objects" to avoid; while in the roadway there are many.

- Air traffic is managed by a different group of operators, and the whole domain is strictly regulated by internationally recognised standards. In road traffic management, even though the density of vehicles on roads is much higher that in the sky, the traffic is governed by mutual interactions between drivers and very loose control by police patrols.
- Cockpits and Air Traffic Control (ATC) rooms are highly automated environments where only pilots and ATC operators have access and demand well developed skill for controlling aeroplanes activities. In automotive environments, in the context of a vehicle a variety of activity can be carried out, and the driving performance seems to assume, ironically, a minor importance with respect to comfort and pleasure.

▪ *Social Context*
- The social context of aviation environments is extremely bounded by the fact that only pilots and cabin assistants have access to cockpits and therefore organisational cultures are very important for identifying behaviour. On the other hand, in road vehicles a much wider variety of situations may be encountered.

▪ *System Integration*
- In aviation, the technology that is implemented for each type of aircraft is well known and does not vary from aeroplane to aeroplane. This is because pilots of a certain Company, certified for a specific aircraft, can fly with any aeroplane of that type without finding differences. In automotive environment, the availability of different models and options, even within the same type of car or vehicle, makes the variety of technology implanted on board, one of the most relevant diversities between cars. Therefore very little system integration exists in automotive environment, and only recently certain standards of interfaces have begun to be accepted.

For all these reasons, the transfer of technological concepts from the domain of aviation to the automotive environment can be achieved only after an accurate analysis of the differences that exist between them.

Design of Interfaces

Human Machine Interaction Integrated Approach: From the human factors perspective, the design process of a specific technology concentrates on the interfaces and controls and on the procedures for implementing their expected functions, in different working conditions, such as normal, transitory and emergency.

In order to clearly define the objectives and functionality of a control system or safety measure, the designer or safety analyst must merge the goals of prevention, recovery and containment of errors, faults and consequences, with the development of a modern technological system. In particular, three fundamental principles for designing Human Machine Systems (HMS) must be accounted for. These are the principles of Supervisory Control, Systems' Usability, and Human or User-Centred Design. They all rotate around the same concepts, which are considered from slightly different perspectives. The designer must effectively keep them into consideration and merge them in a balanced and effective manner.

The supervisory control principle, as already discussed, implies a clear understanding of functions and roles of humans and automation (Rouse, 1991; Sheridan, 1992). Systems' Usability has been defined by International Standard Organisation with the objective of enabling designers to measure the effectiveness, efficiency and satisfaction associated with the use of a product (ISO/DIS, 1993; Bevan and Mcleod, 1994). User-Centred Design (UCD) is an iterative design approach to usable systems development that requires the pre-existence of a modelling architecture of HMS and continuous user's feedback, by which it is possible to produce a sequence of constantly improved design solutions, as the design process develops and passes from initial the stages to more complete solutions and prototypes (Rouse, 1991, Billings, 1997; ISO/FDIS, 1999).

These three principles combine in such a way that the designer can include the user role from the initial stages of the design process, by considering a model of the HMS that takes into account, jointly, the type of activity of humans and the tasks and role of automation/machines, and then can verify and improve effectiveness, efficiency and usability by direct feedback with selected users (Figure 7). In other words, the designer can consider the user role already from the initial phases of the design process, by applying a model of the behaviour of operators/users in conjunction with system/supervisory control models, and then can verify and improve effectiveness, efficiency and usability by direct feedback with selected users.

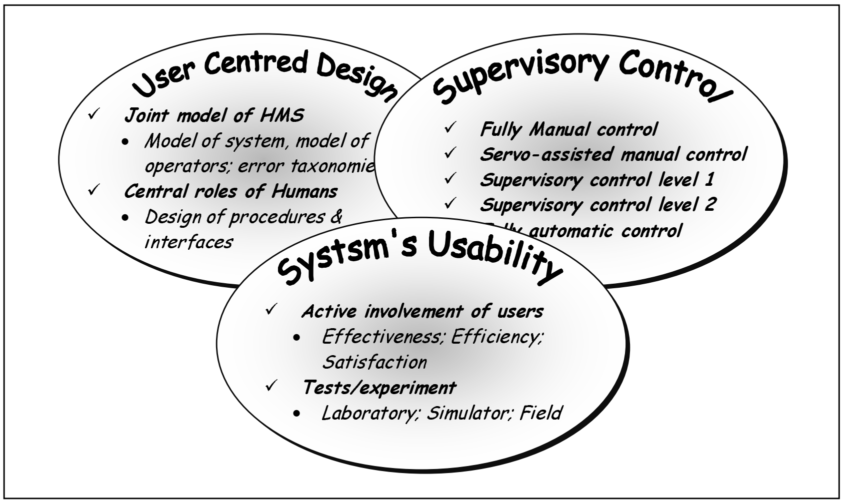


Figure 7 - Synergetic peculiarities of UCD, System's Usability and Supervisory control

The idea of Human-Centred Design (HCD) was developed in the same years of the Information Processing System (IPS) metaphor for maximising and exploiting the role of the human being in the interactions with "modern" machines (Rouse, 1991; Sheridan 1992). Given the development of automation and Information Technology, the HCD concept has been further refined and adapted to different working contexts and domains of application (Billings, 1997; ISO/FDIS, 1999). The basic idea of HCD and UCD has remained constant and, as the name says, it relays on the fact that whichever system, product, tool is designed, the role, needs, and peculiarities of the users must be accurately considered. In practice, the HCD, or "human-centred automation" (HCA), approach consists of the fact that, firstly, the role of supervisory control, assigned to the human operator, is maintained while developing the design, avoiding the trap of fully automatic control system. At the same time, the contribution of end users is considered essential, in the form of "participatory design", for giving continuous feedbacks at various stages of the design process, ensuring user friendliness and maximising exploitation of all features and potentialities of the control system.

Human Machine Interaction Design Activities: Given the above discussion, it is possible to develop a procedure that may support the designer in his/her activity. In the early phases of the design process, the designer should acquire maximum knowledge and *experience* on the environment and working context in which the system will operate (Figure 8). The collaboration with end-users is already important at this stage as it allows to familiarise with the context and tasks to be performed.

Another fundamental initial standpoint for the designer is the selection of an appropriate *model* of reference for accounting for the "joint cognitive system", i.e., the integrated system of the machine and the operator. The selection of this model of reference is of paramount importance for the whole design process, as it defines the context for simulations and experimental tests and for the taxonomy of possible inappropriate or erroneous behaviours to be evaluated during the whole design process. From this initial process, the *context of use* of the system can be developed.

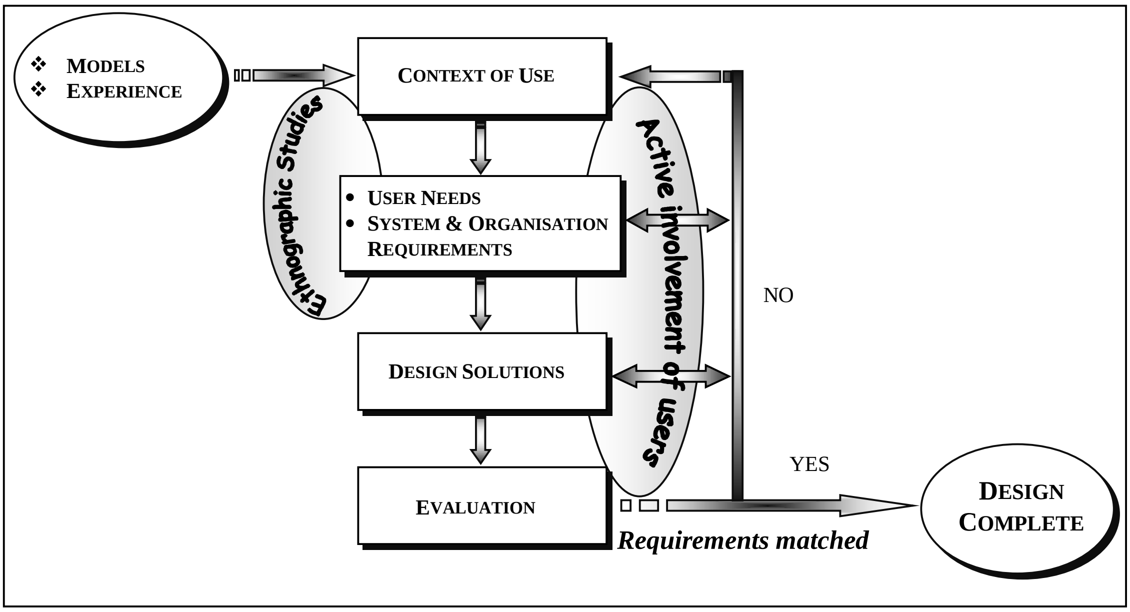


Figure 8 – Procedure for designing Human Machine Systems.

After this initial stage, observations of working context, interactions with users, and analyses of tasks and allocation of functions are necessary. End-users are the best information source for the designer, even if they often are not able to express their needs in words. The way to exploit the enormous amount of information and knowledge of end-users for improving design is through "ethnographic studies", which consist of the performance of a variety of analyses of familiarization with real working contexts that intend to be not-invasive of the normal procedures and work performance. Examples of ethnographic analysis are interviews, questionnaires, observations of normal operations, video-recording and auto-confrontation and all the techniques that intend to capture the practical performances of operators in the real working context.

Allocation of functions consists of the specification of which tasks and responsibilities are up to the user and which, instead, should be performed by the system. It is wrong to assign to the system all what is technically feasible. This allocation should be based upon several considerations, such as abilities, attitudes and limits of the user with respect to the system in terms of reliability, accuracy, speed, flexibility and economical cost. It is impossible to define user needs accurately from the beginning. They need to be taken into consideration by the joint cognitive system. Only through repeated improvements and interactions with users it is possible to know exactly which *user needs* and *system and organisational requirements* should be considered. During such an iterative process, prototypes are developed which are more and more complete and refined, and tests are performed at various levels, such as in laboratory and in simulators. Only when the ecological nature of the HMS has been fully resolved, the actual *design solutions* can be developed.

In the final phases of design, the end users involvement is also important and could result in another process of iteration. In particular, the *evaluation* of the implemented design solutions requires tests and experiments with prototypes implemented on real systems, which may well lead to further revisions of the design.

It is clear from the analysis of the activities involved in the development of a HMS Design process, that this is a lengthy and complex approach that requires the collaboration of several different types of expertise. Usually a design team is based on: end-user, purchaser, user manager, application domain specialist, business analyst, system analyst, system engineer, programmer, marketer, salesperson, user interface designer, visual designer, human factors and ergonomics expert, human-computer interaction specialist, technical author, trainer and support personnel. The success or failure of a design process depends on the profile of the design team and on the accuracy by which the above procedures is carried out. This may turn out to be a quite extensive and demanding exercise. However, any shortcuts or corner cutting, such as limited filed observations and retrospective data analyses and collection, and the extensive exploitation of technology transfer may lead to failures or poor design solutions.

Open Issues
A design development generated from an interactive and progressive process, such as the one described above, is certainly very comprehensive of most aspect concerning Human-Machine Interaction. However, some problems remain unresolved and require specific attention and analysis. In particular the issues of training, risk perception and complacency need attention and will be briefly addressed hereafter.

Training: A very important and almost obvious way to improve safety is by intensifying drivers' training. However, to date this remains an insufficiently developed or poorly exploited way for safety improvement. Nowadays, almost all cars commercialised for the general public may be driven with a basic driving licence. However, there is an enormous variety amongst cars in terms of power, number and complexity of driving devices, and differences in speed and acceleration. Diversification of training is a very reasonable way to ascertain that people in control of such vehicles are properly certified and prepared to their use.

Moreover, many support systems are imbedded in vehicles with the aim of improving vehicle control and safety. Examples of such type of modern safety devices are the traction control and the anti-collision warning systems. These types of systems have an impact on driving behaviour and, usually, demand a certain time before the driver becomes familiar with their functions and operational modes. In most cases, a specific training would be necessary for their optimal use. This training need has not yet been fully recognised, and, in addition it is not yet clear which organisations should be responsible for it.

Risk homeostasis theory: In designing warning (anti-collision) systems the human tendency to adapt to it should not be ignored. Adaptation is defined as the process of modifying own behaviour to suit new conditions. Adaptation is a manifestation of intelligent behaviour, and, normally, is not a problem. However, it has to be considered in designing warning systems, as it may lead to minimising safety improvement, which is precisely the objective of the systems under development.

The Risk Homeostasis Theory (also known as "Risk Compensation") relates exactly to this problem. It was primarily developed and validated in the area of road traffic (Wilde, 1982). However, supporting data come also from several quite different domains, like industrial settings, health protection (above all for smoking issues) and settling in flood-prone territories.

Strategies usually adopted by people do not tend to minimize risk, but aim rather to optimise it. The level of accident risk at which the net benefit is expected to maximize is called the target level of risk (Wilde, 1994). Risk homeostasis theory posits that people continuously compare the amount of risk they perceive in the present situation with their target level of risk and then adjust their behaviour to attempt to eliminate any discrepancies between the two.

In the automotive domain, the amount of risk perceived by drivers depends on the accident rate over a passed period of time. This accident rate, indeed, is the results of driving behaviours engaged by people during last months and, in turn, has an effect on drivers' subsequent behaviours in the folowing time period. Therefore, this homeostatic mechanism constitutes a case of circular causality that works like a thermostat.

Complacency: Complacency is the problem that arises when a user trusts a system to the point that ceases to sufficiently monitor it. This is not the classical "vigilance" decrement, but rather is thought to be an incorrect strategy resulting in sub optimal monitoring (Parasuraman, et al., 1993; Billings, 1997).

A complacent observer has the tendency to sample a variable less often than it is optimal, given the dynamics of the source. Unless an optimal sampling rate has been specified, no claim that sampling is too frequent or too infrequent can be sustained. Note that complacency is implied by under-sampling, not by missed signals, since the operator has no control over the detectability of signals when a source is sampled. In the automotive domain, complacency depends on two factors: the trust the driver places in the system, and the trust the driver places in himself (Cacciabue et al., 2002). It remains and open issue that requires adequate consideration, together with training and risk homeostasis, when new safety devices are designed to support drivers in their task.

Conclusions
This report has aimed at presenting the state of the art in designing Human-Machine Systems for automotive safety. The standpoint that has been assumed is that Cognitive Ergonomics is a science that can

support the designer to preserve the fundamental needs and requirements of the human in control of the vehicle, respecting the conditions for system functionality.

In order to reach this goal, we have initially examined the basic problems that result from road accidents, from a merely statistical viewpoint. Then, we have analysed the basic concepts and fundamental principles of Human Machine Interaction. Particular attention has been dedicated to the issue of modelling the integrated human-machine systems, so as to pinpoint the fundamental cognitive functions of the human in control of the system.

A procedure that aims at supporting the design process of a HMS has been developed. We have identified a spectrum of needs of drivers and requirements of systems, and we have indicated a variety of systemic and cognitive boundary conditions that should be considered by the designer when applying the HMI principles. Other critical issues, such as complacency, homeostasis and special training needs, have been discussed.

This procedure for designing HMS is being applied in practice for designing the human machine interface of a new warning anti-collision system. The efficiency and effectiveness of such interface will have to match the advanced technology applied for obstacles recognition in order to make the tool really relevant for practical application and improving safety.

References

AAA Foundation for Traffic Safety (2001). The Role of Driver Distraction in Traffic Crashes, Washington.

Bevan, N., Macleod, D. (1994), Usability measurement in context in Behaviour & Information Technology, vol. 13, NOS.1 and 2, pp. 132-145.

Billings, C. E. (1997). Aviation automation: the search for a human-centered approach. Mahwah, NJ: Lawrence Erlbaum Associates.

Cacciabue, P.C., Amditis, A., Bekiaris, E., Andreone, L., Tango (2001). The Importance of User Needs Analysis on HMI design. The EUCLIDE example. Proceedings of the Panhellenic Conference of Human Computer Interaction, December 7-9, Patras, Greece.

Cacciabue, P.C., Martinetto M., Re A., (2002). The effect of car drivers' attitudes and national cultures on design of an anti-collision system. 11th European Conference on Cognitive Ergonomics, ECCE-11. Catania, Italy, September, 9-11, 2002

European Parliament, Directorate-General for Research, (1998). The European Community and Road Safety, Working Paper, Transport Series, TRAN 103 EN. (on the web: http://www.europarl.eu.int/workingpapers/tran/pdf/103_en.pdf)

European Transport Safety Council, (1995). Reducing Traffic Injuries Resulting From Excess and Inappropriate Speed, Brussels.

Eurostat, (2000). Transport Safety in the EU, No 76/2000

Harris, D., Smith, F.J. (1999), What Can Be Done versus What Should Be Done: a Critical Evaluation of the Transfer of Human Engineering Solutions between Application Domains, in Engineering Psychology and Cognitive Ergonomics, Ashgate.

ISO/DIS 9241-11 (1993): Guidance on Usability.

ISO/FDIS 13407 (1999) (E), Human-centred design processes for interactive systems.

Leibowitz, H.W. (1988), The Human Senses in Flight. In E. L. Wiener, and D. C. Nagel (Eds.), Human Factors in Aviation, Academic Press, San Diego, CA.

NHTSA (1993). Preliminary Human Factors Guidelines for Crash Avoidance Warning Devices, NHTSA Project No. DTNH22-91-C-07004, USDOT, Draft Document, COMSIS Co.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced "complacency". International Journal of Aviation Psychology, 3, 1-23.

Rouse, W. B. (1991). Design for success. J. Wiley & Sons, New York.

Sheridan, T. B. (1992). Telerobotics, Automation and Human Supervisory Control. The MIT Press, Cambridge, MA.

Wilde, G.J.S. (1982), The Theory of Risk Homeostasis: Implications for Safety and Health. Risk Analysis, 2, pp. 209-225.

Wilde, G.J.S. (1994). Target Risk: Dealing with the Danger of Death, Disease and Damage in Everyday Decisions. Toronto: PDE Publications.

Designing Transgenerational Usability in an Intelligent Thermostat by following an Empirical Model of Domestic Appliance Usage

Adinda Freudenthal

Delft University of Technology, Faculty of Industrial Design, Intelligence in Products group,
Landbergstraat 15, 2628 CE Delft, the Netherlands, a.freudenthal@io.tudelft.nl,
http://www.io.tudelft.nl/intelligentproducts

Abstract: An intelligent thermostat was designed taking into account a model of usage strategies and how people learn. The model was based on earlier observations with young and old subjects using a brand new TV/VCR combination. Evaluation in usability trials confirmed that usage strategies were according to the model and non-compliance to the model in the interface design led to usability problems. New possibilities in ICT allow designers to meet the model based design guidelines. The result is improved usability, to a level where even senior citizens in the future should be able to program complex schedules.

Keywords: human-product interaction, learning to use, elderly users, thermostat, interface design

Introduction

Present domestic appliances, such as microwave ovens and VCR's are equipped with buttons, displays and sometimes screens with menus. Many users have problems when programming these devices. Buttons have labels, which do not indicate what to do, because in most cases several buttons are to be pressed in a certain order. Feedback is usually poor ("ERROR 4"). Feedforward about what should be done is often lacking. The correct menu needs to be opened before items can be found (or modes or screens) and therefore finding an item can be difficult. This is especially the case if the name of the required item does not match the search terms of the user. Older users have even more problems, caused by a decline of cognitive capacities (see various studies on a range of aspects needed for product usage, such as slowing of information-processing, Rabbit (1992), performance during parallel tasks (Myerson et al., 1990) and the suffering from information overload, Cann (1990), recalling the recent past and searching memory, Lovelace (1990)). Besides this elderly users of consumer products often lack relevant experience with modern interfaces. For example, they often have not learned the principle of 'spatial organization' (of menus) or have problems in applying it (Docampo Rama, 2001).

An empirical model of usage and learning to use, based on observations of usage of a combined TV/VCR set, was used in the development of a new intelligent thermostat. The thermostat was designed, using new technological possibilities in ICT and was evaluated in usability studies. In this paper we will investigate whether users behave and learn according to the model also with the new thermostat and whether designing according to the model can help increase usability.

The model was developed earlier on from observations of young and old subjects using a brand new combined TV/VCR set. In that study 5 subjects of 15-18 years of age, 5 subjects of 30-40 years of age and 10 subjects of over 59 years of age participated. The TV/VCR had just been introduced (1994) to the market and was equipped with menus on the screen and many options, such as 'VPT' and 'automatic channel search'. The subjects were allowed to approach the unknown device as they normally would. This way novice use (and learning to use) was tested. The observation and analysis method, and full results are given in Freudenthal (1999, 117-149). The summarized results and model in this paper were published earlier in Freudenthal (2000).

The general approach in novice use

Users of all ages were observed to approach the apparatus in the same way. When the subjects started using the unfamiliar TV/VCR they had an open approach and seemed to apply general knowledge about apparatus. Most started either by switching on the device or reading the manual. Their actions were 'user goal' driven. They had knowledge of possible user goals, for example 'selecting a channel' or 'changing the volume'. However, often just a few of the main functions were remembered. They seemed to strive to achieve these 'real user goals' with as limited effort as possible.

This meant that they were only inclined to make functions their goal, which represented a real user goal. For example 'programming the TV channels' would only become an intermediate goal if this was necessary to achieve another real goal. Users provided with a totally empty TV would sometimes be inclined to start programming the TV; in all other cases this would not be a spontaneous action. Elderly users would rather get help than try programming.

Users were not inclined to make 'learning' their goal. They used the product and unintentionally they learned. They did not intentionally explore. One exception we found was when occasionally an interesting or surprising feature was encountered during use. The user then sometimes briefly explored this feature. However, this was not common.

Procedures to reach user goals

Users expected that their user goals could be reached through the execution of a 'procedure', a sequence of actions in a certain order. However, they did not expect procedures to necessarily equate to those on their own equipment. In fact, they seemed to find it more logical that the new apparatus would function somewhat differently. This could be the result of previous experiences with new devices.

Users did not tend to use a 'trial and error' approach in finding out the procedure to operate functions. This was only observed with extremely simple functions, such as 'how do I switch on the TV ?'. Probably the procedures of operation were too complicated to remember the results of trial and error and be able to deduce a strategy from this. This might explain the 'step by step' approach towards their final goal: seldom more than one step at a time seemed to be planned. In deciding what to do in the next step the users expected the device to guide them by providing relevant information.

Although users expected the product and the manual to guide them in product use they were rather casual about using information supplied by the product and the manual. If they thought they knew what needed be done, they would hardly study the provided information, but would immediately act. This was even the case if they were, for instance, in the middle of carrying out a procedure in the manual. They would forget about that and carry on by themselves. This was already happening early in the learning process.

During first product use subjects used general knowledge such as 'press a button to activate a function' or 'one should first program the channels'. Later on more knowledge would be available. For example, knowledge of required sequences of main actions was used, but only if the user was aware of the fact that a correct order was needed, and this is was not always the case.

The users expected the available product functions to be organized according to how they mentally organize their user goals. For example, the subjects seemed to expect that 'programming the VCR for today' would not differ fundamentally from 'programming the VCR for tomorrow'. Problems occurred frequently when other procedures and/or other buttons were needed. One of them was that the wrong section in the manual was consulted and followed.

Mental storage of product rules

During novice use storage of information in memory was required. However, the available capacity of working memory, during the operation of the home appliance, did not seem to be sufficient to store all separate actions for the various procedures. This might be the reason why users remembered only general rules and forgot details almost immediately. They seemed to search for recurring patterns, translate them into rules and use them for predictions of future product reactions to user actions.

They seemed to deduce these 'laws', probably in the same way they have learned to explain and predict the behavior of their natural surroundings. The subjects all appeared to expect that the product would function according to constant 'laws' and would react consistently. It seemed that, consciously or unconsciously, users expected that reactions of the product to users' actions would reflect its 'laws'.

Executing a procedure that required actions, which did not conform to the general rule, would increase the number of items to be remembered in working memory. It seemed that the available capacity of working memory was usually insufficient for this. Younger users would then start to forget crucial information. Older users would easily reach a level of complete overload after which sensible actions would be rare. General product rules appeared to be rather easily stored, while exceptions were forgotten over and over again.

If a product reacts unexpectedly this can seriously disrupt the learning process. The worst cases were observed when users made mistakes while learning. Making mistakes is generally recognized by a user from the fact that feedback from the product is not according to expectations. If the user has encountered exceptions to the general product rule earlier on he may start to expect more exceptions, (which is in line with the assumption of consistency). We observed users assuming that another exception had been encountered, instead of deducing that a mistake had been made. They had to correct their developing 'mental model' again later on. For younger users this seemed to make learning difficult. For older users such problems were observed to completely frustrate further use.

A model of usage strategies and how people learn

From our observations we derived a model of novice product use of a TV/VCR combination, see figure 1 (Freudenthal, 2000). In the model we organized the way in which subjects appeared to use their internal knowledge and signs from the device.

Well-known elements, as can be found in almost every article or book with design guidelines, were recognized as being important (e.g. on feedforward and feedback, consistency, and guidelines to anticipate learned aspects in language, operation of devices and icons). Their roles in the process of novice product usage are indicated in the model. The figure indicates the relationships by arrows, which indicate the flow in time. During the usage steps the arrows are followed. Meanwhile the internal knowledge grows. After interaction has started the product rules will start to build up.

The described process of building up knowledge seems to take place only partly consciously. The user for the most part does not seem to be aware of the application of knowledge or of the laws and does not form the laws intentionally or consciously.

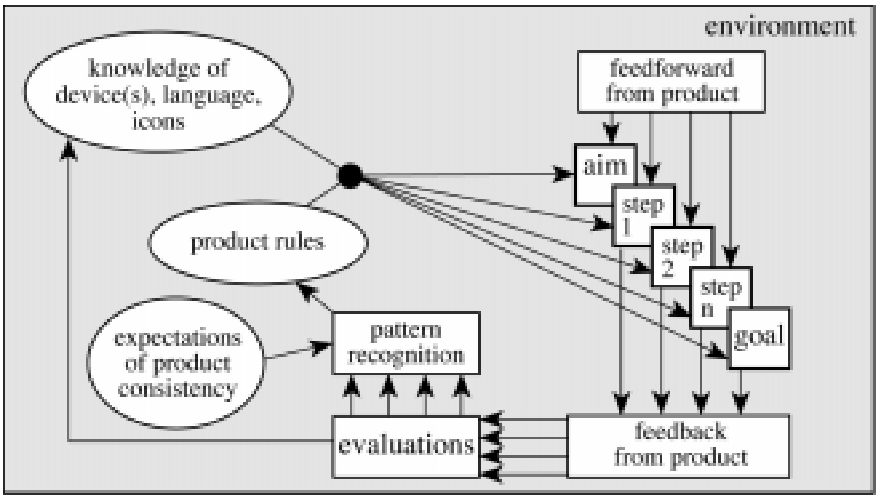


Figure 1 - Observed usage and learning to use of a brand new TV/VCR-combination (Freudenthal, 2000). The manual is seen as part of the product. Not all available feedforward is used (all the time).

A few aspects will be explained a bit more: (1) We found that subjects expect consistency always and throughout a device, even if the device has a rather inconsistent design, therefore this is indicated as a given fact. The assumption of 'consistency' is used in the process of deducing the product rules and needs not to be deduced to be a product rule. (2) We found that feedback feeds the process of developing the product rules in the mind and the growing general knowledge of devices – and therefore of the interpretation of next feedforward. (3) The environment can effect the process. Think of, for example, low lighting or other persons changing settings.

A key finding is the step-by-step manner, in which internal knowledge is used together with the feedforward from the device and the manual. Seldom more than one step at a time seemed to be planned. (The steps indicated in the model are the actual steps taken by the user, and not the steps required by the product).

The behavior of subjects observed during the use of the TV/VCR did not meet expectations based on the three levels of cognitive control according to Rasmussen (1987). If problems of use were encountered, according to these theories, an approach on the highest level – 'knowledge-based' - is to be expected. Users are supposed to plan their actions in advance and carry them out. It seemed that short-term memory was simply insufficient for the complex problems users encountered in the apparatus.

There is a mismatch between what users are willing to invest and the required effort to operate and/or operating the brand new domestic device was above available human capacities. These capacities might be relatively low in this situation, due to a low motivation to (learn to) use. The observed low motivation is probably caused by usage goals, which are rarely of major importance and a life-threatening situation does not occur if things go wrong. Therefore goals could be adjusted if other goals were satisfactory as well.

Finally we must mention that, as all models are simplifications of what really happens, so is this model. For example, sometimes a user goal is not reached or the goal is not equal to the aim and mistakes are made. We decided to not explicitly indicate intermediate goals in the model, but these can be seen as reflected in the sequence of usage steps to be made.

Research target

In Freudenthal (1999, 201-231) guidelines were presented to optimally support young and old users. Many of these guidelines were composed to meet the model in a design. The set of guidelines was tested in new appliance designs in industry and proved to be helpful to improve usability. The degree of improvement is substantial but has its limits, because several guidelines conflict or are 'impossible' to meet in currently applied technologies.

An example is the requirement to substantially enlarge typefaces on static products. This requires space. However, unknown abbreviations (technospeak) should not be used, nor parts of sentences (which often mentally are completed incorrectly). Using icons as labels is not an option, because understandability is low for older users. Nevertheless users should be completely guided in carrying out the relevant procedure, including error corrections. It should be 'shown' at all times what should be done and how. Enough guidance should be given for all users, including senior citizen.

If a designer actually aims at meeting all these requirements it does not suffice to adapt existing products. Whole new interaction principles are needed. We expected that new technologies such as speech recognition and embedded help would provide possibilities to do this.

In this investigation we would like to find out whether:

the model applies for other domestic devices (besides the TV/VCR);

to what extent the model can be anticipated in a design using new technological possibilities in ICT;

whether applying such solutions actually increases usability for young and old users.

Method

The Delft Intelligence in Products group designed an intelligent thermostat (Keyson et al., 2000, TUDelft patent) based on literature and earlier research (see Freudenthal, 1999, Keyson et al., 2000). Also practical design experience from earlier work by the team members was used. A substantial part of the interface was based on the described model of usage and learning.

Once a simulation was available user tests were carried out to find out about the quality of usability (Freudenthal et al., 2001, Freudenthal and Mook, in press). The subjects used the simulation of the thermostat, one subject at a time. (The version tested was one version before the one described below; differences will be mentioned in the next section). In the trial 7 subjects with ages between 25 and 45 and 7 subjects with ages between 60 and 73 were observed. The subject was given tasks, but they were as much as possible given as contexts for use. This was done to, as much as possible, not give away clues on how to carry out the task. No 'thinking aloud' during the tasks was asked for. There was no paper manual available. Afterwards the subjects were interviewed mainly to find out about their opinions about the thermostat. Results from the usability test were selected for this paper, which concerned design concepts according to the model and their assessed effect on usability. Also problems in interaction were analyzed to see whether it is likely that they are caused by applying the model or by not applying the model. General behavior was analyzed to detect approaches according or in contrast with the model.

For a complete record of the thermostat design, the usability test method and results we refer to Freudenthal et al. (2001) and Freudenthal and Mook (in press).

Design suppositions

The main properties of an interface, if it is to serve the user in a way, which matches the observed 'natural' way of interacting with domestic appliances, are:

- It should help the user define his (main) usage goals (it may be expected that the user can remember his goal during operation, but hardly more than that).
- It should fully guide the user in setting his first and every next step to reach his goal taking into account the level of knowledge already available. It should therefore provide sufficient guidance for users with different levels of knowledge to begin with and different learning curves.
- Information from the product during usage must be completely consistent to build up the knowledge of product rules to be used in next actions; only general rules will be remembered.

In the thermostat the richness of the various modalities was used in combination to support a range of types of information. The dialog between user and thermostat is a regular conversation in spoken language (the mother language of the user). A GUI (Graphical User Interface) on touchscreen was used to present settings and allow direct manipulations for changes with instant feedback. Sound was used to support clicking and dragging on the touch screen.

Expected foreknowledge: Crucial is the level of knowledge of users. When we started we knew that knowledge of devices, menus and icons is extremely poor with older subjects. To be able to exclude as few users as possible we assumed that users would have knowledge of their mother language (not technospeak, but everyday spoken sentences) and of known devices, which have been in use for such a long period that all users would know them. We chose an 'agenda', a 'thermometer' and a 'map of the home' (in floors) and turned these into screens for direct manipulation, feedback and information overview.

The current design can de divided into two modes. The default mode is meant to be used by anybody, also an (elderly) person without any previous knowledge. This mode can overrule all programs in the thermostats 'agenda', which allows inexperienced users to not be hindered by programs set by others. Very basic instructions will be given if a user just stands in front of the device and does nothing. The thermostat will explain that the screen should be touched or that the users should give an answer from the 'things to say' list. (See figure 2.)

The advanced mode (figure 3) can be activated by a regular user, e.g. a person who lives in the home. First time authorized users already have learned some rules about the thermostat: e.g. they know that they can speak to it and that they can set temperatures by dragging on the arrows. Later on they will learn more rules. Depending on the user's age and usage experience, assessed by the thermostat (or asked) the embedded help will adapt and be present when it might be needed. For example, a novice user might ask the thermostat 'help me switch floors' (listed under 'things to say', figure 4). The thermostat would then say: "Tap on the floor you want".

Guidance to set goals and sub goals: The possible (main) usage goals are presented to the authorized user at the beginning of (new) interaction (figure 3), so not in the middle of carrying out a procedure. They are presented as complete messages in the list of options. The user tells the thermostat his main goal and this directs the thermostat's next step or question. New sub goals will be given through a verbal dialog and direct next steps.

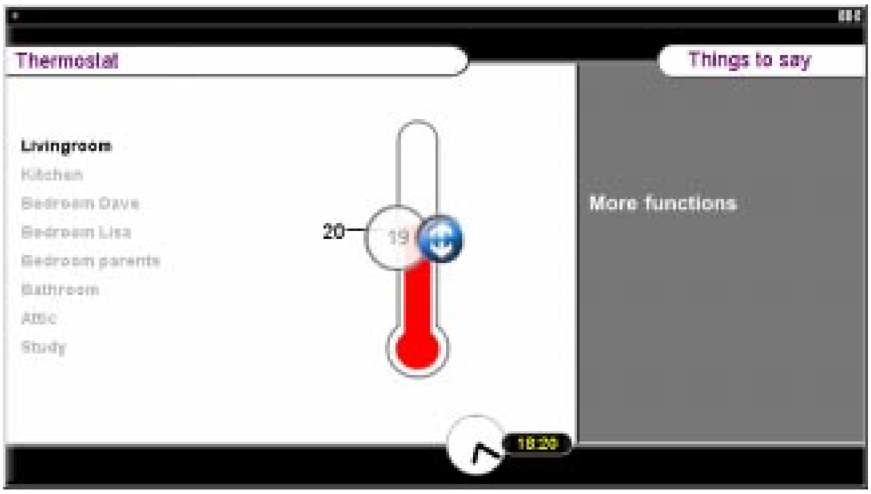


Figure 2 - Non-authorized users can use only this screen. They can program a temporary setting in separate rooms. Current (black, 20) and desired (gray, 19) temperatures in the living room are displayed. To view the figure in color, please consult http://www.dcs.gla.ac.uk/~johnson/eam2002/ or ----/iria2002/

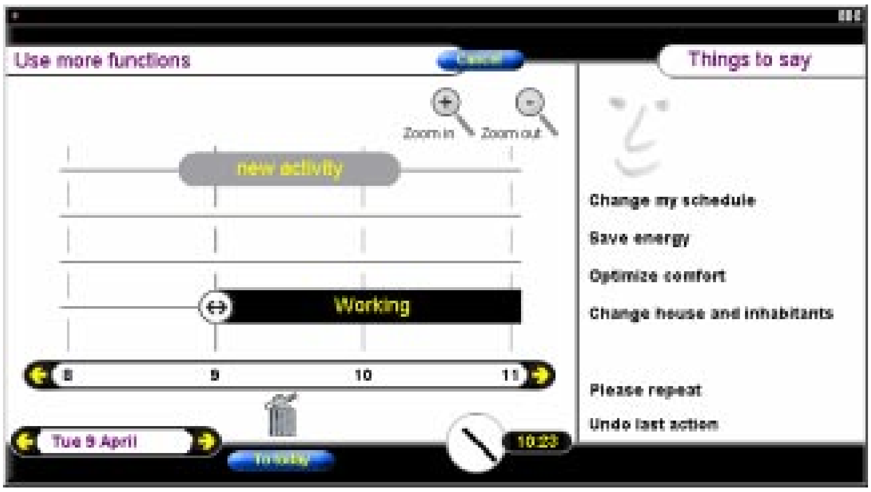


Figure 3 - The activities in the agenda can be scheduled in any desired recurrence. The user can program through direct manipulation in the agenda or main user goals can be said or touched (right part of the screen). Once a goal has been chosen next 'things to say' will be shown, depending on the situation and the next question posed by the thermostat.
To view the figure in color, please consult http://www.dcs.gla.ac.uk/~johnson/eam2002/ or ----/iria2002/

Guidance to make settings: Some users will have specific wishes to set their thermostat and others will not want to bother; they just want the thermostat to see to it that they do not waist energy. For all users, but in particular for this last group, it is important that the thermostat system will recognize recurring living patterns. It can present suggestions, based on these patterns, to the user. The user needs to discuss the suggestion through a normal verbal dialog. If the user agrees the thermostat can program the suggestion. If temperatures are needed from the user he is asked to set these in the relevant rooms in the map of the home (figure 4). Also the thermostat might need a name for the new activity in the agenda and asks the user to type that in on a keyboard presented at that time. The user will find the new activity in the agenda with a self-chosen name and can adjust it later if he wants to.
For users who have (gained) some more knowledge about the thermostat and devices in general there is the possibility to interact in a more active way. They can program intentional scheduling through a dialog. They can even manipulate in their agenda without waiting for instructions or they can overrule parts of longer procedures by acting directly, without a verbal dialog.

Consistency: Consistency has been one of the main focus points in the design. For example: the user gets an overview of what he can say, so he cannot say anything else. Speech recognition allows for more commands, but we choose to always provide explicit feedforward.

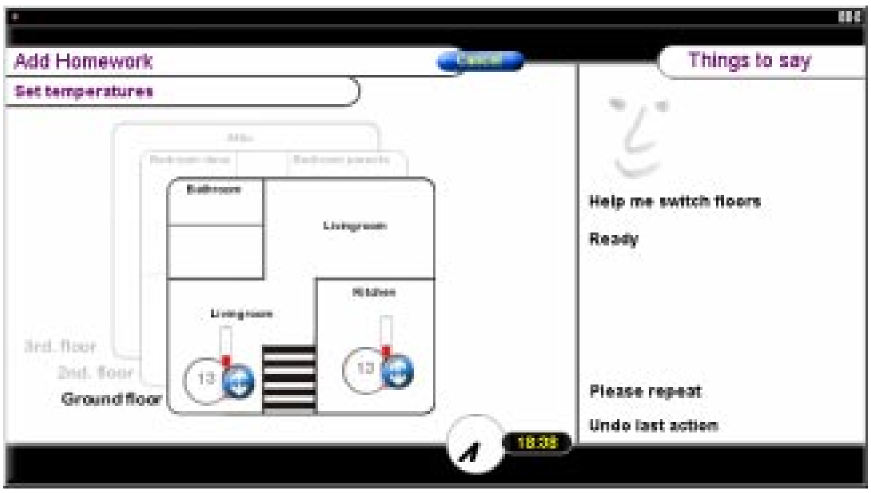


Figure 4 - The rooms of the home are presented in a map of the house. The temperatures in rooms belonging to an activity can be programmed here. The left-hand side can be operated by direct

manipulation. The options at the right can be said or touched. An example of adaptive embedded help is shown ('help me switch floors').

To view the figure in color, please consult http://www.dcs.gla.ac.uk/~johnson/eam2002/ or ----/iria2002/

Learnings from the usability trial

We found that some users indeed need instructions on how to operate through the verbal dialog, the first time they use it. Once they received instructions interaction through a spoken dialog worked very well for all subjects. Even the oldest subjects could rather effortlessly follow the suggestions. They listened to the questions and gave the appropriate answers. They appreciated the suggestions as they found the usage goals to be sensible (especially the saving of energy) and they liked the interaction style. Although the suggestions were easy to operate several younger subjects expressed their dislike of the rigidness of the wizard.

The provided guidance was sufficient, except for one family of subgoals. 'Checking the settings made' was not supported the way other subtasks were. For example, the provided final feedback was a flashing instance of the last changed setting. The user should scroll his agenda or the map of the home to check settings. This, however, was not the natural thing to do. Completely according to our own theory young and old users were not inclined to take sufficient action, even though they in many cases were clearly in need of such information. Because checking final settings is crucial to avoid errors in programming, we added spontaneous and complete final feedback in the present redesign. The thermostat's voice now tells the user what settings were made. If more freedom for the younger users will be developed the thermostat will have to know when it should give the final feedback.

All subjects understood the large metaphors to present the programming data (the thermometer, agenda and map) without any additional explaining. The feedforward to indicate how the metaphors should be operated through direct manipulation were clear, such as arrows to indicate where to drag. As a comparison we had also implemented direct manipulation without explicit feedforward (such as the stairs in the home, to be tapped to go up or down). This was unclear to all subjects.

Subjects could find an activity (birthday) in their agenda; they could all scroll their agenda and set temperatures. They were very well supported by the immediate feedback during manipulations to guide the user in adjusting the agenda or a temperature, such as a little clock, which appeared next to the pulling finger when changing the time of an activity. Activation of items was indicated by sound. Changes were immediately visible in the agenda, map, etc.

In one of the procedures – making a new activity – users were required to do more than just answer questions and carry out dictated small tasks. This type of interaction worked well for the young subjects, but it was too difficult for some elderly subjects. It was also not clear whether all elderly subjects could relate to the user goal. At least some of them did not seem to feel a need to program a schedule for temperature control.

Icon understanding proved again to be extremely low, especially with older subjects. Icons, which were supported by optional embedded instructions, were easy to use. For example 'deleting an activity' required an icon, the trashcan, see figure 3, and was successfully used.

Error recovery was not sufficiently supported yet. Elderly subjects used the item 'one step back' to navigate (several times for more steps). They did not notice that it changed their settings back. Specific support anticipating the lack of knowledge about 'spatial organization' (of menus) and error recovery principles needs to be developed for older users.

Conclusions and discussion

A model of domestic product usage was applied as design guidance for an intelligent thermostat. The model was derived from observations of users operating a brand new TV/VCR. Based on our findings with the thermostat we expect that the model also applies for the first encounter with other domestic appliances, provided that the level of motivation to achieve usage goals and perceived risks during operation is similar.

If we compare the performance observed with the thermostat with usual performance levels in usability studies with home electronic devices, the difference is remarkable. With many current devices, which provide a huge amount of possibilities, we tend to see subjects struggle (see for example Vermeeren, 1999). Elderly subjects often get completely lost when they try to program. Our thermostat provided even more functions than the most advanced thermostats on the market. Nevertheless we observed even elderly subjects actually program complex patterns into a schedule, with relative little effort.

We found a difference between young and older users. Younger users require less support to carry out tasks, because they have more knowledge available. We even expect that they might want to work in the thermostat GUI without verbal dialogue after a while, for certain subtasks. In this context maybe even 'knowledge based' behavior could occur later on (because their level of knowledge could (become) sufficient as well as their youthful working memory).

The impression is very strong that following the model in our design was for a large part responsible for the observed ease of use. It is unlikely that it is responsible for all increase in usability, because besides the aspects related to the model; many more guidelines for design were taken into account. To mention just one: we included verbal explanations about the way the smart home system (of which the thermostat is a part) works to improve user understanding and thereby acceptance of the new principles.

Nevertheless, the way users behaved was according to what we should expect, based on the model. The subjects were supported so that they should be able to operate with little effort and did so. In those situations where the design asked for some more effort, problems occurred or subtasks were not even carried out. Both by the design and by the trial set up the subjects were encouraged to learn by doing. They complied with this effortlessly. The subjects used the device from the very first encounter and did not take the time to explore or intentionally learn. Therefore they learned by using.

This way of learning by doing – mainly of how to carry out procedures to accomplish programming goals, is not common in the usage of safety critical devices. With such devices first usage tends to take place in a separate training phase, without real risks, and usually whith help from others or from simulations or training-documentation. Motivation to learn is usually higher and intentional learning is strived for. Such a learning phase will require another model of human behavior (which possibly could be an extension of this model).

Application of this other model of human behavior is not likely to mean that our model based design principles will not apply for the learning process of operation of devices for safety critical tasks. Our model based design principles comply with most well known existing design guidelines, relevant for ICT based product (interface) design. (An endless list of sources with relevant ergonomic guidelines could be mentioned here. Examples come from different disciplines, such as Pirkl and Babic (1988), for consumer electronics and appliances, http://www.useit.com, for updates on website design by Jakob Nielsen, and Schneiderman (1998) and The Mitre Corporation (1986), on desktop applications and Preece et al. (2002), who combine the various disciplines for professional and consumer products.) In many cases our guidelines are stricter or more specified for specific age groups (Freudenthal, 1999).

If young and old users can learn to program a complex thermostat by the described design aspects, it is likely that they will befit in other situations as well. Users of professional devices have many problems to operate as well, see for example Bogner (1994). Even if the model is only partly valid for safety critical devices, consequences for design requirements of such devices could be great. Present devices which are used in safety critical situations tend not to meet the requirements depicted, e.g. no complete guidance is provided, products are inconsistently designed and the burden on working memory could become high with functions not used recently, resulting in forgotten exact steps in procedures. Therefore users of safety critical devices could especially be supported by these principles with seldom-used procedures of operation.

We have demonstrated that there are ways to apply ICT in domestic appliances so that ease of use can be substantially increased for young and old users. Older residents need not be excluded from the advantages of modern devices, just because the interfaces are too complex to operate. It is even possible to give them back the control over their own devices without help from others, or have them overrule programs in devices they have never seen before - a gerontologist's dream.

Acknowledgement

The author expresses her thanks to all team members and others who have contributed to the design and research work: David Keyson, Marc de Hoogh, Elyon DeKoven, Riejanne Mook, Liek Voorbij and Henk Arisz.

References

Bogner, M.S. (1994). Human error in medicine. Lawrence Erlbaum Associates, New Yersey.

Cann, M.T. (1990). Causes and correlates of age-related cognitive slowing: effects of task loading and CNS Arousal, Proceedings of the Human Factors Society 34th Annuals Meeting 1990, 149-153, 1990.

Docampo Rama, M. (2001). Technology Generations handling complex User Interfaces, Eindhoven University of Technology, the Netherlands, PhD thesis.

Freudenthal, A. (1999). The design of home appliances for young and old consumers, Delft University Press, Delft, The Netherlands, PhD thesis.

Freudenthal, A. (2000). Improving home appliances to support the learning phase, Proceedings IEA/HFES 2000, july 29-aug. 4, 2000, CD-rom, 6.909 - 6.912.

Freudenthal, A., Keyson, D.V., DeKoven, E. and de Hoogh, M.P.A.J. (2001). Communicating extensive smart home functionality to users of all ages: The design of a mixed-initiative multimodal thermostat-interface. In: OIKOS 2001 Workshop: Methodological Issues in the Design of Household Technologies, Molslaboratoriet, Denmark, March 12-13, 2001, 34-39.

Freudenthal, A. and Mook, H.J. (in press). The evaluation of an innovative intelligent thermostat interface, Universal usability and age differences, Cognition, Technology & Work, Special Issue on Household Technologies.

Keyson, D.V., de Hoogh, M.P.A.J., Freudenthal, A. and Vermeeren, A.P.O.S. (2000). The Intelligent Thermostat: A Mixed-Initiative User Interface. In: CHI 2000 Extended Abstracts, 59-60. New York, NY: ACM Press.

Keyson, D.V., Freudenthal, A., de Hoogh, M.P.A.J. and DeKoven, E.A.M. (2000). TUDelft patent, Interface unit, March 30th, 2000, Dutch 1014792, International PCT/NL01/00257.

Lovelace, E.A. (1990). Basic concepts in cognition and aging. In: E.A. Lovelace (ed.), Aging and cognition: mental processes, self-awareness, and interventions, 1-28.

The Mitre Corporation Bedford, MA (1986). Guidelines for designing user interface software, US Department of Commerce, National Technical Information Service (NTIS), Springfield, V.A., Aug. 1986.

Myerson, J.M., Hale, S., Wagstaff, D., Poon, L.P. and Smith, G.A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing, Psychological Review.

Pirkl, J.J. and Babic, A.L. (1988). Guidelines and strategies for designing transgenerational products: An instructors manual, Center for instructional development, Syracuse University.

Preece, J., Rogers, Y. and Sharp, H. (2002). Interaction design: beyond human-computer interaction, John Wiley & Sons, Inc., New York

Rabbit, P. (1992). Memory, Chapter 8.2, In: Evans, G.J. and Williams, T.F. (eds.), Oxford textbook of geriatric medicine, Oxford University Press, Oxford, 463-479.

Rasmussen, J. (1987). Reasons, causes and human error, Chapter 26, In: Rasmussen, J., K. Duncan and J. Lepat (Eds), New technology and human error, John Wiley & Sons, Chichester, 293-301.

Vermeeren, A.P.O.S. (1999). Designing scenarios and tasks for user trials of home electronic devices, In: Green, W.S. and Jordan, P.W. (eds.), Human factors in product design, current practice and future trends, Taylor & Francis, London, 47-55.

Schneiderman, B. (1998). Designing the user interface, Strategies for effective human-computer interaction, Third edition, Addison-Wesley, Reading, Massachusetts.

http://www.useit.com, Jakob Nielsen's website.

An Introduction in the Ecology of Spatio-Temporal Affordances in Airspace

An L.M. Abeloos, Max Mulder, René (M.M.) van Paassen,

Delft University of Technology, Faculty of Aerospace Engineering,
Kluyverweg 1, 2629 HS Delft, the Netherlands.
{a.l.m.abeloos, m.mulder, m.m.vanpaassen}@lr.tudelft.nl
http://www.cs.lr.tudelft.nl

Abstract: This paper outlines a research project with a purpose to come to a prototype intelligent cockpit interface that will support building and tuning a model of the environment that is shared by the human and machine. Such improved mutual awareness of the situation would lead to less human error, thereby reducing a great cause of aircraft accidents. The innovation of the work presented here lies in the application of the ecological approach to visual perception of Gibson [1986] to come to a definition of situation awareness in terms of the spatio-temporal affordances of the environment. It is expected that an intuitive presentation of the environmental affordances will lead to a higher-level of situation awareness in the cockpit. The design of the intelligent interface is based on the principles of Ecological Interface Design (EID) as introduced by Vicente and Rasmussen [1992]. For the implementation of the intelligent interface, there is a focus on the integration of alerting systems through a multi-agent system and on the presentation of a trend in the criticality of the situation.

Keywords: Affordances, Alerting Systems, Aviation, Cockpit Display Design, Ecological Interface Design, Ecological psychology.

Introduction

Over the years, the aircraft and its environment have evolved to a complex, high technology work domain for the pilots. Some avionics systems[9] were introduced to help the flight crew deal with this complexity. These systems have shown improvements in safety and efficiency, but today's elongating time delays and the high accident rate indicate that some major challenges for improvements are still to be accepted. Moreover, with the expected traffic growth and the potential introduction of new air traffic control concepts, such as Free Flight or station keeping, the airspace environment will become even more complex. To remain a market concurrent to other types of transportation in terms of safety and efficiency, some radical changes are needed. This challenge was accepted by the Control and Simulation division of Delft Aerospace in the ongoing *2020 project*. This project investigates future developments for the three main elements in the air traffic system: Airport 2020, Airspace 2020 and Flight Deck 2020. The research presented in this paper is part of the latter, while realising that improvements on the flight deck alone will not be sufficient to achieve the safety and efficiency goals.

This paper describes the conceptual design rationale of an intelligent flight deck interface that allows an intuitive and adaptive human-machine dialogue to achieve a high level of human-machine shared situation awareness. The paper is structured as follows. First, the lessons learned so far from some Flight Deck 2020 subprojects are discussed. Next, the notion of affordances is described and how these affordances determine boundaries. This leads to a new definition of situation awareness in terms of affordances. The next section describes how the affordances are perceived in the complex and dynamic airspace environment. Then, the principles of ecological interface design are applied to the design of the intelligent flight deck. The final section describes how a realistic approach is followed to ensure the usability of the research in real life applications. The paper concludes with a brief discussion.

Flight Deck 2020 Project

Looking at the aircraft accident causes over the last decades, most accidents seem to be caused by pilot error (see Figure 1). Studies by Reason [1990] and Endsley [1995] indicate that a great deal of human error originates from a lack of situation awareness. Thus, new systems introduced in the cockpit should aim mostly at helping the pilot building situation awareness.

---

[9] Some newly introduced systems are the Flight Management System (FMS), Electronic Flight Instrument System (EFIS), and alerting systems like the Traffic alert and Collision Avoidance System (TCAS) and the Ground Proximity Warning System (GPWS).

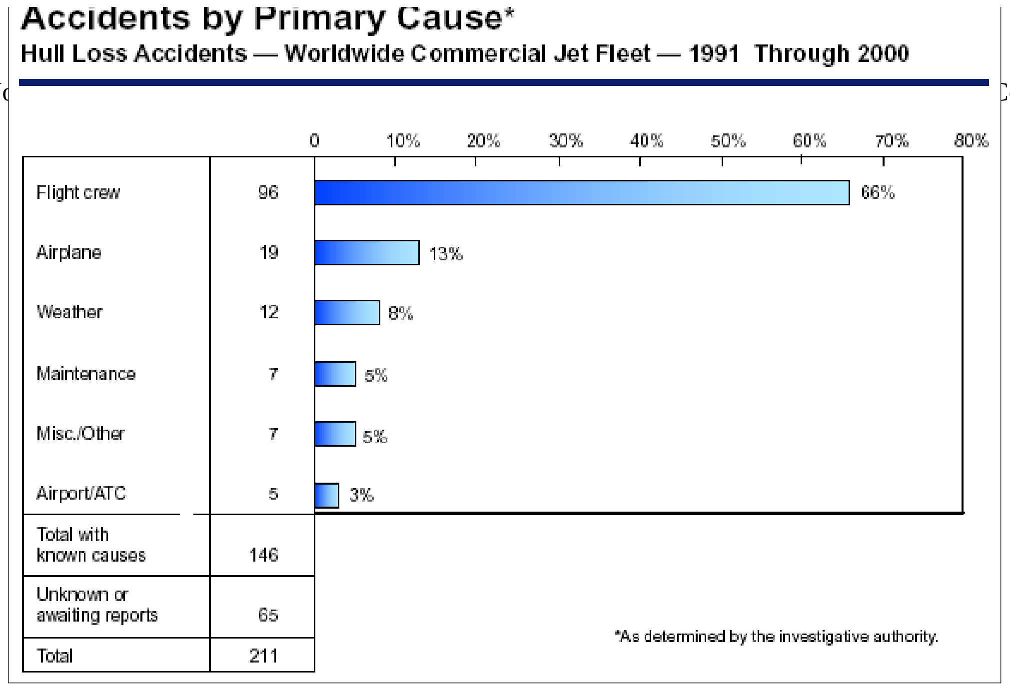


Figure 1: Aircraft accidents categorised by primary cause [Boeing 2001]

Looking at the current flight deck, we see that the *conventional* automation is dumb, inflexible and fragmented [Bainbridge 1987, Billings 1990]. The alerting systems are *independent*, bringing about some problems, such as nuisance alerts, over- and under-reliance, procedural conflicts, simultaneous and even contradicting alerts [Pritchett 2001]. Further, the current displays are *planar* representations of the aircraft's spatio-temporal situation, which is four-dimensional [Mulder 1999].

These fields for improvement provide the basis for the Flight Deck 2020 project. Some relevant subprojects are briefly described next:

– *Cockpit alerting systems:* Several studies involved the working and interaction of existing and future cockpit alerting systems. The considered warning systems all monitor some part of the aircraft's environment, e.g. traffic, terrain, weather. In [Abeloos et al. 2000a], the potential co-operations between the ASAS[10] and TCAS[11] were discussed. A study concerning the integration of aircraft warning systems is presented in [Mulder et al. 2000]. Use is made here of a multi-agent architecture.

– *The intelligent adaptive flight deck:* A literature survey was performed investigating the applicability of artificial intelligence in the cockpit in the form of an adaptive human-machine interface. Different triggers, timing, methods and levels of adaptation were considered. The interested reader is referred to [Abeloos et al. 2000b]. Currently, an experiment is set up for the testing of a prototype intelligent navigation display (see Figure 2) that adapts the human-machine dialogue to the criticality of the situation. Different levels of adaptation will be compared, ranging from fully reactive[12] to fully pro-active[13]. This study also questions the need for a trend of the criticality to be displayed (as explained in the section on EID).

[10] The Airborne Separation Assurance System (ASAS) is an aircraft warning system that would alert the flight crew when there is a potential for a loss of the required minimum aircraft separation in the near future. This system is still under development.
[11] The Traffic alert and Collision Avoidance System (TCAS) warns the crew in the event of a potential mid-air collision. It alerts on a very short notice and its resolution 'advisories' are mandatory. This system is currently installed on many different aircraft all over the world.
[12] Reactive adaptation: No automation is involved. The user performs the adaptation himself.
[13] Pro-active adaptation: The automation takes the initiative and adapts the display, without consulting the user.

Figure 2: An impression of the intelligent Navigation Display. The adaptation level shown is the interactive level that lies in between the proactive and reactive levels. It shows the operator what changes to the current display setting are required to get a good view on the conflict (in this case a terrain conflict).

– *Human-machine interface for planning four-dimensional trajectories in the cockpit:* Three interactive Navigation Displays (ND) were developed, at three levels of ease in interaction, supporting the flight crew in the task of in-flight re-planning of a four-dimensional flight plan. The high-level pilot support navigation display is illustrated in Figure 3. Through the incorporation of direct manipulation, pilots can directly perceive the consequences of their planning decisions on the 4D constraints, and act on them accordingly. The displays were tested in an experimental evaluation with professional pilots. Results of this project are yet to be published.

– *Functional modelling of airspace:* Another study considered a functional modelling technique for expressing travel possibilities in air traffic [van Paassen, 1999, de Neef et al. 2001]. A travel function for aircraft heading guidance in free flight airspace has been designed and evaluated. The function is applied to two aircraft travelling at equal velocity on collision courses with different aspect angles. It assumes a constant heading of the intruder and calculates bands of headings the ownship has to steer clear of to avoid the conflict. The off-line simulations indicated that the safety goal is achieved, i.e. sufficient aircraft separation is maintained at all times. The efficiency goal however is not achieved by simply showing heading bands, because whether the aircraft manoeuvred to the left or right (in the horizontal plane) resulted in different off-track distances. The efficiency goal demands the presentation of the preferred heading(s). Final results of this study are yet to be published. A geometrical interpretation of the heading travel function in the relative velocity field is presented in Figure 4.

Lessons learned so far within the Flight Deck 2020 project are:

– The future air traffic management system calls for the introduction of some sort of an ASAS in the cockpit. If not handled properly, the interactions and similarities of this system with the already installed TCAS can lead to a lot of confusion in the cockpit.

– It seems that an architecture based on multi-agent systems is suitable for the integration of the aircraft warning systems that monitor the aircraft's environment.

– With the increased complexity of an aircraft's environment, there is a need for an intelligent human-machine interface that adapts the human-machine dialogue to the situation at hand.

– Only a combined effort in the fields of intelligent interface design and the design and integration of alerting systems could bring about some effective results.

– It seems possible and promising to describe the airspace with a functional model. However, a new modelling technique is necessary, because the airspace functions are complex and constantly changing due to the aircrafts' locomotion.

From the above, it can be concluded that the current aircraft interfaces are not very well suited to the pilot's task, but that there are promising techniques available for improvement.

The efforts and lessons learned in the areas of alerting systems, intelligent interfaces and airspace modelling now form the departure point for the project *The Ecology of Spatio-Temporal Affordances*. The goal of this project is to build and test a prototype intelligent interface that can share with the human user a model of the direct environment, including its constraints and unexplored possibilities. The interface is intelligent by integration of information and adaptation of the presentation of this information in a manner that is most appropriate to the current situation. The ultimate goal is to come to an improved awareness of the situation by the human as well as the machine, so that their co-operation can lead to an important increase in the safety level reflected in a decreasing number of accidents. The concept presented here is based on Gibson's [1986] ecological approach to visual perception and the subsequent ecological interface design (EID) principles by Vicente and Rasmussen [1992]. Further, as mentioned before, for the implementation of the whole, use is made of multi-agent systems [Ferber 1999].

In this paper, the field of application is aviation. However, the presented theory could just as well be adopted to any vehicular locomotion such as car driving, cycling, sailing, etc.

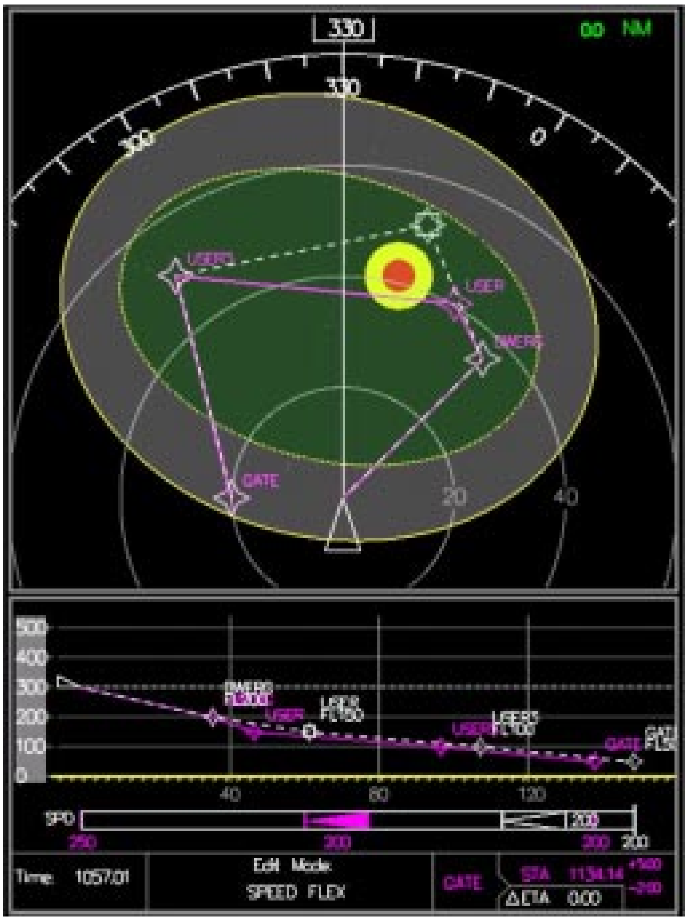
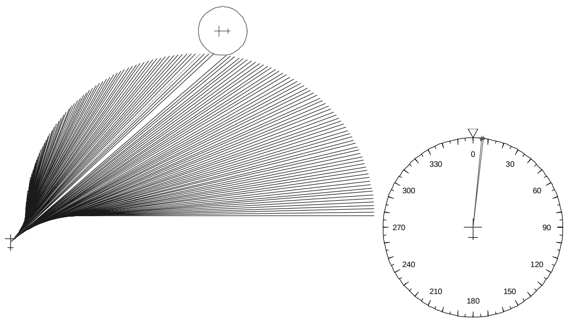
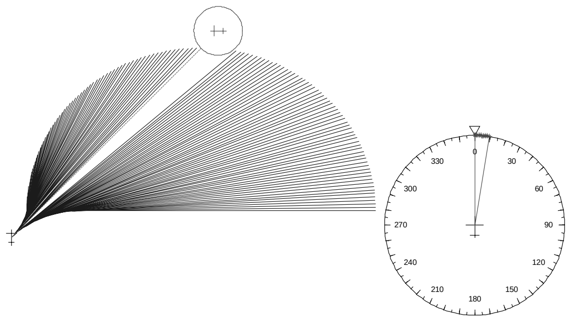


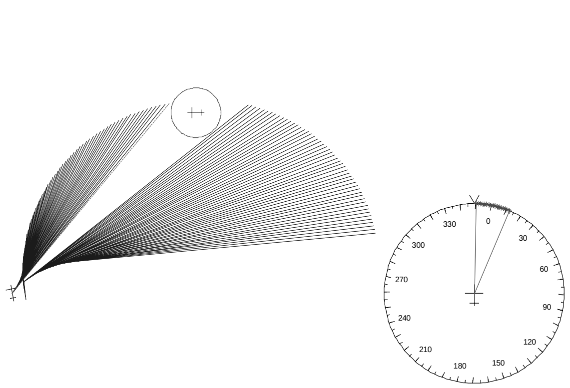
Figure 3: The 4D navigation planning display with high-level pilot support. The top square is the Horizontal Situation Display showing the conventional top-down situation of owncraft relative to waypoints. The ellipses shown are the affordance zones for the waypoint USER. They appear when a waypoint is moved with the Cursor-Control Device to avoid a severe weather area for example, as indicated by the circles. The zones indicate where a waypoint may be positioned to still arrive at GATE within a certain time delay with (light grey) or without (dark grey) changing speed. The lower square shows the Vertical Situation Display, a timeline indicating necessary speed changes and some additional information about time.



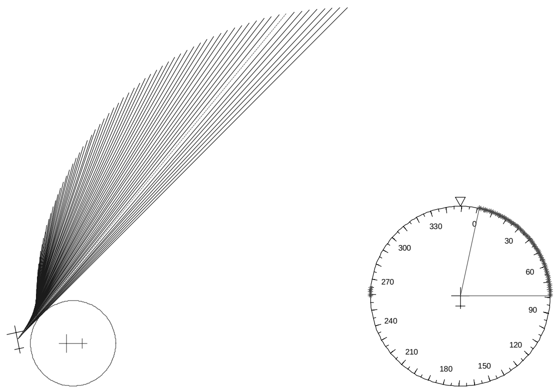
(a) A heading band emerges as the protection zone of the intruder touches the travel function of the ownship.



(b) The heading band now includes the ownship's current and desired heading. A manoeuvre is necessary.



(c) The ownship has manoeuvred to a new heading that is outside the heading band.



(d) The original desired heading is now outside the heading band. The ownship can return to the original heading.

Figure 4: Illustration of the heading travel function in the relative velocity field. The ownship is travelling north and encounters an intruder travelling west. The travel function calculates the headings that have to be avoided by the ownship. These headings are presented in a heading band.

The Notion of Affordances

The notion of affordances originates from Gibson's [1986] ecological approach to visual perception. This approach is developed within the field of *ecological psychology*, which is that part of psychology that considers the interrelationship of organisms and their environments.

The affordances of the environment are the goal-relevant properties that determine the reciprocal relationship between an animal and its environment. Gibson [1986] gives a detailed description of the elements on Earth and of what they afford to the animal or human:

The medium, substances, surfaces, objects, places and other animals have affordances for a given animal. They offer benefit or injury, life or death. This is why they need to be perceived.

[Gibson 1986, p. 143]

As an aircraft travels through airspace, it is important to consider the affordances of this airspace. A far from complete list is as follows:

Air affords unimpeded locomotion.
Oxygen affords respiration.
Other aircraft and birds afford collision.
Terrain, obstacles and a water surface afford collision.
Terrain affords support.
Water affords floating or sinking.
Clouds afford obstructing the view and icing.
Thunderstorms afford turbulence.

The above affordances are relevant in any airspace. In controlled airspace, some artificial elements were introduced by aviation organisations to regulate and control air traffic: airways, holding patterns, Standard Instrument Departures (SID), Standard Arrival Routes (STAR), etc. They provide *procedural affordances*.
Some of the airspace affordances are dependent on certain state variables of the own aircraft. For example, the amount of lift and drag generated depends on the aircraft's speed. For the apprehension of the environmental affordances, it is therefore important that the actor is aware of his own status (e.g. position, speed, configuration, engine performance, etc.).

Affordances Determine Boundaries

Even with the given examples, an affordance remains a very abstract concept. This section describes how these affordances can be imaginarily visualised through the boundaries or barriers they determine. Where the medium touches surfaces, objects, or phenomena, boundaries in the environment are defined. These boundaries indicate a transition in the airspace affordance. Most of the time, they imply a constraint for our locomotion. To provide a sufficient level of safety, usually safety margins are defined that protect us of actually coming into contact with those boundaries.

Some airspace boundaries that result from the airspace affordances are:

- *Surfaces: terrain, water*
- *Objects: aircraft, birds, obstacles*
- *Phenomena: clouds, thunderstorms*
- *Procedural boundaries*

Further, there are also some aircraft-specific boundaries, which we will call the *system boundaries*. The aircraft's *flight envelope* defines its limitations depending on engine performance, configuration, and structural load limits. The *covering range* defines the maximum look-ahead distance (and thus time) depending on the capability of the surveillance equipment on board and on the ground.
In the current airspace regulations, the boundaries are protected by specified safety margins:

- *Obstacle clearance minima*
- *Minimum aircraft separation*
- *Visual and instrument flight rules*
- *Maximum allowable deviation of procedures*

Situation Awareness in Terms of Spatio-Temporal Affordances

Locomotory motion is greatly influenced by the operator's perception of his direct environment. This perception of the world is what one calls *Situation Awareness*. Several definitions of situation awareness are circulating. The definition by Endsley seems to be the most cited and complete:

Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.

[Endsley 1995, p. 36]

Endsley distinguishes three hierarchical levels in the process of acquiring situation awareness:
1) *Perception of the elements of the environment:* First of all, situation awareness consists of knowledge of the status, attributes, and dynamics of relevant elements in the environment. For the pilot, this includes both elements from inside and outside the aircraft.
2) *Comprehension of the current situation:* The perceived elements are integrated in a holistic picture of the environment, again including the ownship. Comprehension of the situation includes an understanding of the significance of those elements in light of pertinent operator goals: *Does the environment offer me any assistance or resistance in the achievement of my goals at this moment?*
3) *Projection of future situation:* Finally, through knowledge of the status and dynamics of the elements and the comprehension of the situation, we can project the future actions of the elements in the environment.

Now that the notion of the affordances is introduced, we take a new look at the acquisition of situation awareness. A new and brief definition of situation awareness is introduced:

Situation awareness is the perception of the spatio-temporal affordances of the environment.
With this definition we claim that for the acquisition of situation awareness in a planning locomotory task, it is essential to perceive the environmental affordances. The definition implies that good awareness of the situation includes knowledge of the current affordances of the environment, the evolution of these affordances in time, and the consequences of our potential actions as reflected in a change of the relevant affordances. It is in particular the latter that makes the difference with Endsley's definition. Knowledge about the environmental affordances allows us to look *beyond* our current status. For the locomotory task, we want to move freely in the environment, so we are interested in the investigation of alternative routes for the optimal achievement of our goals. We are seeking answers to questions of the kind *What if …?* For example, what if we want to change heading? This may require a different altitude to be chosen too, because of an approaching mountain range on the new course. For efficient planning, we do not want to use trial and error. We need to know in advance whether or not the new heading will bring about consequences, without actually implementing the new heading. By directly perceiving the environmental affordances, this knowledge will be included in the situation awareness.

The Perception of Affordances
It is the perception of the affordances that controls locomotion through the environment. Gibson suggests that affordances are perceived directly, rather than after an assessment of the presented qualities:

The psychologists assume that objects are composed of their qualities. But I now suggest that what we perceive when we look at objects are their affordances, not their qualities.

[Gibson 1986, p. 134]

The perception of airspace affordances for an aircraft in flight is not as straightforward as the perception of affordances of our everyday environment due to the complex and dynamic characteristics of the airspace environment.

Complex environment: The airspace is a very complex environment. It consists of many different components: terrain, traffic, weather, airports, etc. Further, the air traffic density is continually increasing. In the next fifteen years, air traffic is expected to double! To be able to manage all this traffic, complex procedures and rules are introduced. Another point is that the direct environment that has to be taken in consideration for our locomotory task is too big for the unaided eye to perceive. Also, today practically all aircraft have the necessary equipment to allow instrument flying. This means that it is possible to fly in zero visibility conditions, thus without any visual reference from outside the aircraft, relying on cockpit instrumentation. Other components of the airspace are simply not directly visible, not even in good weather conditions. These are the airspace classifications, procedure entry or exit points, etc. Such elements are only visible on maps, charts, and perhaps also on certain cockpit displays.

Dynamic environment: The travelled airspace is a continuously changing environment. Aircraft fly here and everywhere with different routes, velocities, performances, etc. Weather is changing and every airport is different and requires different procedures. Also, since we consider travelling through this airspace, we have to take into account our own dynamics. Due to our locomotion, the relative aspects of the environment are continuously changing. Our actions in the environment have consequences for its affordances. What the airspace affords us is dependent on our location in that airspace and on time. That is why the airspace affordances are *spatio-temporal*. It has to be emphasised though that it are not the affordances themselves that change. Another aircraft always affords collision, and a runway affords a landing surface. It is rather the relative location of the boundary determined by that affordance that is changing due to our locomotion.

On the basis of the new definition of situation awareness, a new approach to the design of cockpit alerting systems and displays can be taken to make visible to the operator the affordances of its direct environment.

Ecological Interface Design (EID)
Vicente and Rasmussen [1992] argue that for complex work domain interface design, the focus should be on semantics rather than on ergonomic issues. They provide a theoretical framework that postulates a set of general, prescriptive principles for the design of a smart interface that reveals the goal-relevant higher order properties of the complex work domain and provides support for problem solving. The framework is based

on two questions. The first question requires a description of the characteristics of the *domain*, while the second one relates to characteristics of the *operator*. It is this human-environment reciprocity that is central to ecological psychology.

It is our intention to use the concept of affordances to build a four-dimensional model of the environment that can be shared by the human and machine. The interface, through which this ontology is shared, should satisfy the principles of EID. The design should *make visible the invisible*, allowing the operator to directly perceive the affordances of the work domain at any level of abstraction. Thus also the high-level goal-relevant properties of the environmental components should be reflected in the interface.

The two questions that form the basis of EID are now applied to the design of the intelligent flight deck:

1) How to describe domain complexity?
This question considers how the spatio-temporal affordances can be mapped onto the interface, taking into account the long-term goals of the system. The mapping requires the definition of a suitable *ontology*. Ontology is a term from artificial intelligence (borrowed from philosophy). It is a description of the concepts and relationships that can exist for an agent or a community of agents for the purpose of knowledge sharing.

An ontology is an explicit specification of a conceptualisation. A conceptualisation is an abstract, simplified view of the world that we wish to represent for some purpose.

[Gruber 1993]

Once the ontology is defined, it will provide the interface *content* and *structure.*

In the preliminary phase of this project, there are yet more questions than answers found. Some of the questions that are raised are discussed here:

– How to model the affordances in the shared ontology so that it is compatible with the user's mental model? To come to a model of the environment that can be shared by the human and machine for building situation awareness, it is necessary to study how the human perceives the affordances. Some of the possibilities that come to mind are time-to-contact, no-go-zones, or potential fields of travel, although a more innovative concept may prove to be more appropriate.
– Other than with the current cockpit alerting systems that work completely independently, the defined ontology must allow efficient communication between the agents monitoring some part of the environment (terrain, weather, traffic, ownship, etc.) so that a holistic model of the spatio-temporal affordances of the environment is obtained.
– What is the relationship between possibilities, constraints and affordances? While possibilities offer benefit or survival, the constraints relate to injury or death. It is this categorisation that allows us to survive in the environment. To enable a human-machine dialogue, we have to categorise the perceived airspace affordances into possibilities and constraints.
– A further analysis of the affordances requires prioritisation. It is not too difficult to determine a general priority order of threats (e.g. terrain over aircraft over weather). However, the prioritisation of the total set of affordances may depend on our short-term goal. Also, it is more difficult to prioritise the offered possibilities (e.g. alternatives to minimise fuel consumption against delays). For an efficient dialogue, the human and machine should adopt the same priority scheme.
– Another difficulty is the choice of an appropriate reference frame. It seems that the pilot uses an *Earth-referenced* system when considering departure, landing, or terrain clearance. However, when looking at an approaching aircraft, an *egocentric* reference frame is used, although the speed of another aircraft is still expressed absolutely. A pilot is not accustomed to relative speeds. This dilemma may become a difficulty when one reference frame has to be chosen for the shared environmental model.
– Is it important for the user to know what part of the environment is affording something? It seems like simply presenting an affordance of the environment without providing further information about the component that forms the basis of this affordance will not be sufficient for the required situation awareness in our locomotory task. Especially, due to the different dynamic characters of the environmental components, it is necessary to identify the concerning component to be able to predict and anticipate future behaviour of that component.
– Finally, there is an important matter of implementation: Is the model of the environment published on a high-level layer and then shared by all agents? Or, do we let the agents monitor their part of the environment and communicate the useful information to the other agents? In [Mulder et al. 2000], two different architectures for agent communication are discussed: direct agent-to-agent communication or

through a selector. Both architectures have their advantages and disadvantages. While direct communication reduces the flexibility of the system, the selector could become the bottleneck of the system.

2) How to communicate the information?
This question considers how the ontology can be mapped onto the displays, taking into account the short-term goals of the system and the criticality of the situation. It concerns the interface *form*.
Some questions are:

– How to visualise the affordances of the environment? Should the affordances be presented on a low level, thus through aircraft speeds, headings, altitudes, etc. Or, is a high level presentation more appropriate, for example a presentation of the areas in the environment that have to be avoided? If we want to provide help in the decision making process, it is necessary to present the information on a high level. However, the final execution of the flight is still performed in terms of low-level commands, such as speeds and headings. Somehow, the interface will have to deal with both levels.

– When considering an alternative conflict-free route, next to a presentation of the current airspace affordances, the display should contain information that allows us to look in the future somehow. So, how to present the consequences of our possible future actions for the affordances of the environment? For example, how to indicate that a certain heading change would require an altitude change?

– How to present the temporal characteristics of the environment? How to present the spatio-temporal affordances on just a two-dimensional display?

– How to identify and treat discrepancies between the actual affordances of the environment and what the user perceives as an affordance? Do we need intent inference? Potential discrepancies have to become visible through the human-machine interface. A two-way dialogue is necessary to come to a joint environmental model.

– How do we present affordances creating alternatives to the flight crew?

– The current alerting systems work with two or three alerting levels depending on the criticality of the situation. A level change is indicated by the use of different colours and voices. Contrary to this alerting philosophy, the intelligent system should give a continuous presentation of the evolution of a conflict. The pilot is thus informed whether his current actions are effective in solving the problem. It is expected that this continuous indication of the level of criticality will greatly enhance the situation awareness during critical events. We have however no previous experience with the presentation of a criticality trend. How to visualise this criticality and how to indicate a change in the criticality of the situation? How to warn the flight crew of an approaching boundary?

Usability
This section discusses the usability of our ecological interface design. We believe that the strength of our research will lie in the converging of two approaches: top-down and bottom-up. The *top-down* approach is high-level, fundamental research. It considers the principles of ecological psychology, the perception of information, decision-making processes, functions, goals, etc. In other words, we seek answers to the following questions: *What do we need? What do we want?* The *bottom-up* approach looks from the system's point of view: *What do we have available? What can be realised with the existing technology?* These questions focus on the existing cockpit alerting systems and on the use of artificial intelligence in the form of multi-agent systems. We believe that the artificial intelligence techniques will allow us to build a top layer over the already existing alerting systems. This would provide integration and intelligence without touching the alerting systems themselves, avoiding the burden of certification issues.
We are not considering major changes in the pilot's task. Neither will we propose revolutionary presentation techniques, but rather make better use of the sensory channels that are used today in the cockpit (vision and audition) by directly presenting the affordances of the environment.

Discussion
The Flight Deck 2020 project at the Control and Simulation group of Delft Aerospace so far looked at the following topics: integration of aircraft warning systems, design of an intelligent adaptive flight deck and functional modelling of airspace. These (and other) studies have shown that the current aircraft interfaces are not very well suited to the pilot's task (vehicular locomotion), especially when taking into account the future developments, such as increased traffic density and flexible routes. Theories and techniques from the fields of psychology as well as artificial intelligence are available for improvements.
It is argued that for complete situation awareness, it is necessary to look beyond the current situation and have knowledge about the effects the own actions may have on the situation. This information is contained in the spatio-temporal affordances of the environment. It is the perception of the boundaries determined by

the airspace affordances that complements our situation awareness. In other words, situation awareness is the perception of the spatio-temporal affordances of the environment.

It is assumed that an intelligent interface obeying the principles of ecological interface design directly presenting the environmental affordances and with a multi-agent system architecture, will lead to improved situation awareness in the human-machine system. This would ultimately lead to a decrease in the number of fatal aircraft accidents.

In the ecological interface design process, two phases are identified. First, the spatio-temporal affordances have to be mapped onto an ontology that can be shared by all agents (human and artificial) in the system. Then, this four-dimensional ontology has to be mapped onto the displays. The airspace affordances are however very complex and dynamic which brings up many difficulties.

The strength of this research project lies in the converging of two approaches. The top-down approach (as presented in this paper) is high-level research, considering the principles of ecological psychology, perception of information, functions, goals, etc. The bottom-up approach focuses on building an intelligent interface prototype based on existing cockpit alerting systems and on the use of artificial intelligence in the form of multi-agent systems.

References

Abeloos, A.L.M., Mulder, Max, van Paassen, M.M., Hoffman, E. (2000a). *Potential Co-operations between the TCAS and the ASAS.* Published in: Proceedings of HCI-Aero 2000, International Conference on Human-Computer Interaction in Aeronautics, Toulouse, France, September 27-29.

Abeloos, A.L.M., Mulder, M., and van Paassen, M.M. (2000b). *The Applicability of an Adaptive Human-Machine Interface in the Cockpit.* Published in: Proceedings of the 19th European Annual Conference on Human Decision Making and Manual Control, Ispra, Italy, June 26-28.

Bainbridge, L. (1987). *Ironies of Automation*. Published in: Rasmussen, J., Duncan, K., Leplat, J. (Eds.), New Technology and Human Error. London (UK): John Wiley & Sons Ltd.

Billings, C.E. (1997). *Aviation Automation – The Search for a Human-Centred Approach.* Mahwah (NJ): Lawrence Erlbaum Associates, Inc.

Boeing Commercial Airplanes, Airplane Safety (2001). *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Applications, 1959 – 2000*. www.boeing.com/news/techissues

de Neef, R.M., van Paassen, M.M. (2001). *Functional Modelling of Airspace.* Published in: Proceedings of the 20th European Annual Conference on Human Decision Making and Manual Control, Lyngby, Denmark, June 25-27.

Endsley, M.C. (1995). *Toward a Theory of Situation Awareness in Dynamic Systems.* Published in: Human Factors, 37 (1), pp. 32-64.

Ferber, J. (1999). *Multi-Agent Systems. An Introduction to Distributed Artificial Intelligence.* Harlow (UK): Addison Wesley Longman Ltd.

Gibson, J.J. (1986). *The Ecological Approach to Visual Perception.* Hillsdale (NJ): Lawrence Erlbaum Associates (originally published in 1979).

Gruber, T.R. (1993). *Toward Principles for the Design of Ontologies Used for Knowledge Sharing.* Published in: Formal Ontology in Conceptual Analysis and Knowledge Representation, edited by Nicola Guarino and Roberto, Poli, Kluwer Academic Publishers. Also available as technical report KSL 93-04, Knowledge Systems Laboratory, Stanford University.

Mulder, Mark, Pedroso, P., Mulder, Max, and van Paassen, M.M. (2000). *Integrating Aircraft Warning Systems.* Published in: Proceedings of the 19th European Annual Conference on Human Decision Making and Manual Control, Ispra, Italy, June 26-28.

Mulder, Max (1999). *Cybernetics of tunnel-in-the-sky displays.* Ph.D. dissertation, Delft University Press, the Netherlands.

Pritchett, A.R. (2001). *Reviewing the Role of Cockpit Alerting Systems.* Position Paper in Human Factors and Aerospace Safety 1(1), pp. 5-38.

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

van Paassen, M.M. (1999). *Functions of Space and Travellers.* Published in: Proceedings of the 18th European Annual Conference on Human Decision Making and Manual Control, Loughborough, United Kingdom, October 25-27.

Vicente, K.J., and Rasmussen, J. (1992). *Ecological Interface Design: Theoretical Foundations.* IEEE Transactions on Systems, Man, and Cybernetics, 22 (4), pp. 589-606.

Modelling Control Situations for the Design of Context Sensitive Human-Machine Systems

Johannes Petersen

Center for Human-Machine Interaction
Oersted DTU, Automation, Technical University of Denmark, DK-2800 Kongens Lyngby
jop@oersted.dtu.dk

Abstract: Safety and efficiency in process control depend on human operators being capable of identifying the state of the controlled system and its environment and assessing how this influences the *control situation*, i.e. *control action possibilities* and *control action norms.* This paper addresses important modelling problems associated with the design of human-machine systems that are sensitive to changes in control situations. Based on a discussion of control actions a generic description of control situations is proposed. Furthermore, it is shown how to specify the content of control situations and how to track changes in control action possibilities and norms based on a representation of the work domain.

Keywords: human-machine systems, context-sensitivity, situations, design, modelling.

Introduction

In supervisory control of complex dynamic systems the actions performed by human operators are shaped by the state of the system and its environment. A control action that is possible and appropriate in one situation may not be possible and/or appropriate in another situation due to a state change in the controlled system or its environment. Safety and efficiency in process control depend on human operators being capable of identifying the state of the controlled system and its environment and assessing how this influences the *control situation*, i.e. *control action possibilities* and *control action norms*. In order to support the operators' assessment of the control situation it is desirable that the human-machine system can keep track of the changes in the control situation and present this information to the operator.

This paper addresses important modelling problems associated with the design of human-machine systems that are sensitive to changes in the control situation. Based on a discussion of control actions we propose a generic description of control situations. Moreover, we consider the relation between control situations and a specific *work domain*. That is, 1) how to specify the content of control situations on the basis of a representation of the work domain, and 2) how to derive control situations from invariant structures of the work domain. The latter is a precondition for the design of human-machine systems that are sensitive to changes in control situations. Figure 9 shows a schematic illustration of the relation between control actions, control situations, and the work domain.

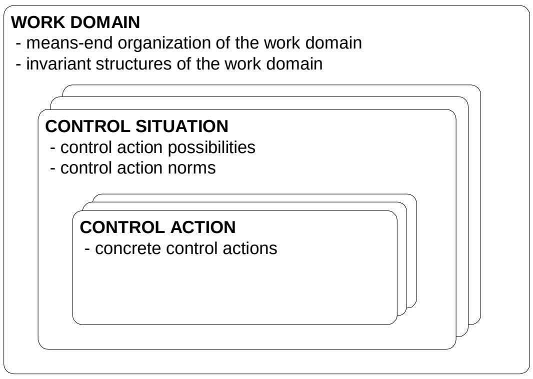


Figure 9. An illustration of the three levels of analysis discussed in this paper.

To demonstrate the importance of understanding how the dynamics of control situations shape the control actions of operators we will draw on examples from the maritime domain focusing on the control tasks performed by the navigating crew conducting a container vessel from port to port.

Control Situations
In this section we will specify what we mean by control situations in the domain of process control and hence the factors shaping the control actions performed by human operators. Our working hypothesis is that control situations comprise *actualities* (the actual state of the controlled system including disturbances), *possibilities* (the control action possibilities) and *norms* (prescriptions of appropriate state changes in the controlled system). Control situations change in response to (significant) changes in one or more of these factors. Before we can give a structured description of control situations it is necessary to discuss in more detail *control action possibilities* and *control action norms*.

Control Action Possibilities: In order to discuss the factors determining the possibilities for control we need a more detailed view on control actions. For this purpose we will take advantage of the important distinction between *doing* things and *bringing about* things proposed by Von Wright (1971) in his discussion of the conceptual relation between causality and human action.

> “It is convenient to distinguish between *doing* things and *bringing about* things, and therefore also between ability to do and ability to bring about. By doing certain things we bring about other things. For example, by opening a window we let fresh air into the room (bring about ventilation), or lower the temperature, or bring about that a person in the room feels uncomfortable, starts to sneeze, and eventually catches a cold. What we thus bring about are the effects of our action. That which we do is the cause of those effects. The cause I shall also call the *result* and the effects the *consequences* of our action” (op.cit., p. 66, emphasis in original).

Human operators perform control actions in order to *bring about* desired state changes (or non-changes) in the controlled system. Typically, the desired system state changes are not the *result* of the actions manipulating a specific part of the system but rather a *consequence* hereof. That is, operators *bring about* desired state changes (or prevent state changes) in the controlled system by manipulating system components that they believe will eventually produce desired state changes (or prevent state changes from happening). E.g. when the navigating crew on board a container vessel wants to bring about a change in the heading of the vessel they normally do this by manipulating the rudder (changing the rudder angle) or the thrusters (changing the number of revolutions of the thrusters).

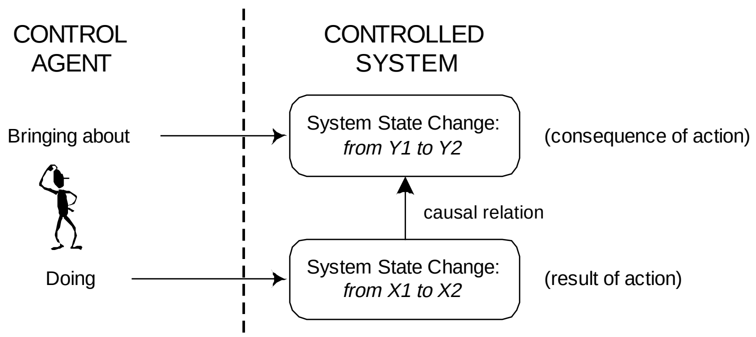


Figure 10. The doing and bringing about aspects of control actions.

In Figure 10 we have illustrated the *doing* and *bringing about* aspects of control actions performed by a control agent. Doing something results in a system state change, which in turn leads to another system state change. Actually, several system state changes may happen as a consequence of the action, corresponding to additional system state changes in the right hand side of Figure 10. Later, when discussing the relation between control situations and the work domain, we will be concerned with how to identify the actual content of the system state changes that control actions refer to.

Note that Figure 10 describes causation only in terms of a relation between two system state changes (events), not showing the underlying generative mechanism producing the second system state change. In Figure 11 we have illustrated the generative mechanism by referring to the interaction between system components (two in this case) producing the system state changes in response to appropriate circumstances – a result of the actions (doing) of the control agent.

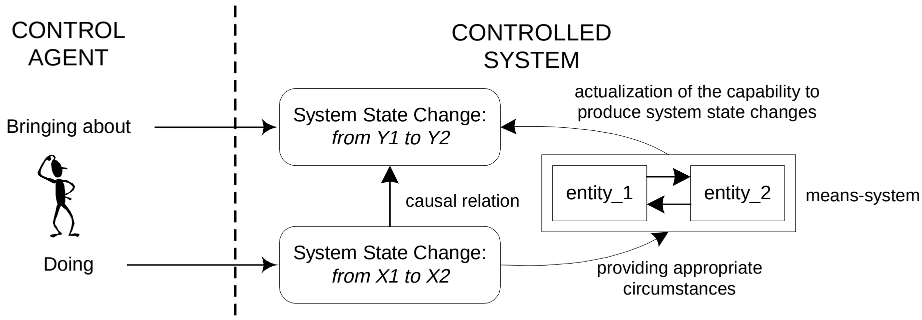


Figure 11. Illustration of the generative mechanism underlying the causal production of a system state change.

Despite the fact that the causal production of a state change is a result of several entities *interacting,* ordinary means-end thinking leads us to focus on only one of these entities – *the means* - while backgrounding the others. For instance, the rudder fitted on a vessel is typically seen as a means for transverse force production on the vessel although force production is actually a result of the interaction between the rudder *and* the water flowing past the rudder. See (Petersen and Nielsen, 2001) for a discussion of the link between means-end relations and causation.

In order to preserve the strength of means-end thinking while obtaining an adequate account of causation we propose the term *means-system* to refer to a *system* of (interacting) components or things involved in the production of a system state change (see Figure 11). Typically, the control agent manipulates only a single component of the means-system in order to bring about a system state change (this might very well be the reason why *means* are typically thought of as single entities in ordinary means-end thinking).

To specify the control action *possibilities* of the control agent it is necessary to consider both the *ability to do* and the *ability to bring about*. In a specific control situation the *ability to do* is determined by the ability to manipulate a specific system component (part of a means-system). This includes the controls provided by the interface of the human-machine system, the ability of the operator to manipulate these controls and the possibility of the control system (*control means*) to manipulate some system component. The *ability to bring about* is given by the capability of the means-system being manipulated to produce the desired consequences in the controlled system[14]. The capability of a specific means-system is determined by the properties of the components forming part of the means-system. Later we will discuss how to derive the actual capability of means-systems.

Control Action Norms: Apart from the actual control action possibilities control situations must include an account of the prevailing control action norms, i.e. prescriptions for what is an appropriate control action in the given situation (i.e. permitted, obliged or intended). Typically, such norms are expressed indirectly in terms of prescribed system states (or system state changes). That is, as control actions lead to system state changes the normative constraints on system operation turn into control action norms. The operator is supposed to see to it that the norms are not broken. That is, in a concrete situation the operator will have to take into account the actual norms related to system operation and ensure that the consequences of his or her control actions comply with these norms.

The norms related to system operation in a given control situation (as well as the upcoming norms in future control situations) determine what control actions are appropriate. Often it is a deviation (or a likely future deviation) between the actual system state and a system norm that triggers the need for control.

[14] The "interface" between *doing* and *bringing about* aspects of control actions could have been defined differently. E.g. we could have reserved the doing aspect to the manipulation of controls in the interface of the human-machine system and hence view the manipulation performed by control means as a state change in a system component (which in turn brings about another the state change in the controlled system). The main reason for upholding the proposed "interface" is that it provides a clear distinction between intervention (operator and control system action) and the propagation of state changes in the controlled system.

Norms can have different modalities. Some are *intentional* expressing a plan for what the system is intended to do (how it is supposed to function)[15]. Other norms are *legal* expressing the permitted operation of the system, or *physical* expressing how the system can behave without leading to physical damage.

The Structure of Control Situations: Based on the above discussion of control action possibilities and control action norms we can now refine our account of the structure of control situations. It is proposed that a control situation comprises the following factors shaping control actions:

- Current system state
- *Control Action Norms*
- *Ability to bring about*:
    a) The means-systems offering the capability to produce purposeful state changes in the controlled system (or preventing unwanted state changes from happing)
    b) The actual capability of the means-systems in the situation
- *Ability to do*:
    a) The control means offering the capability to manipulate system components of the controlled system
    b) The actual capability control means in the situation
- *Disturbances*

Before actually performing a control action it is important that the operator assesses the situation (identifying the actual system state, action possibilities, control action norms, etc.) and constructs a control action by establishing a match between the goal, i.e. a desired system state (which is supposed to comply with the given norms of the situation) and the means-systems offering the capability to produce the desired system state. In order to actually bring about the desired system state the operator typically has to manipulate a system component (part of a means-system) using capable *control means*.

The success of a control action can be determined only in relation to the current control situation. Below we outline a set of conditions defining the criteria for successful performance of a control action, A0. A0 is supposed to achieve the goal, G0, in the situation S0, and comprises the *bringing about* of the system state Y (produced by the means-system, MS0) and the *doing* of X (employing the control means, CM0).

1) MS0 can/will produce the system state Y in S0 when the system state X occurs.
2) Obtaining the system state Y achieves the goal G0.
3) The production of Y will not violate the norms of S0.
4) The operator is capable of performing an action that results in the system state X (by means of CM0) in S0.

In the remaining part of this paper we will focus exclusively on the "bringing about" aspect of control actions, i.e. how state changes are produced by means-systems. Consequently, we presume that the capabilities for causal production offered by means-systems can always be actualised by the operator using available control means. In other words, condition 4 is always satisfied.

A Maritime Example

Let us consider a control situation, S1 faced by the navigating crew on board a container vessel. Let there be only one norm in S1 given by the passage plan prescribing the route along which the vessel is supposed to move. In this situation it is the task of the crew to control the motion of the vessel so that it complies with the planned route. Presuming that the speed of the vessel is relatively high, the capability of the rudder-system to produce transverse force acting on the vessel is *high*, whereas the capability of the thruster-system to produce transverse forces acting on the vessel is *low*. This means, that in S1 the rudder-system is a means-system, capable of producing changes in the heading of the vessel (through changes of the rudder angle), whereas the thruster-system is incapable. A structured description of the control situation S1 is given below:

CONTROL SITUATION, S1 =

Current system state:

The speed of the vessel is high

Control Action Norms:

Follow the given passage plan (cross track error = x)

---

[15] Such plans may derive from the operators supervising the system or from another agency not interacting directly with the system.

Ability to bring about:

Means-systems for producing transverse forces on the vessel (including their actual capability):

rudder-system; the actual capability to produce transverse force on the stern of the vessel is *high*.

thruster-system; the actual capability to produce transverse force on the vessel is *low*

Ability to do:

Control means for changing the rudder angle (including their actual capability):

--

Disturbances:

No external forces from wind and current are interfering

If, however, the speed of the vessel is reduced significantly the rudder-system looses its capability to produce changes in the heading of the vessel or prevent changes from happening (at low speed the capability of the rudder-system to produce transverse force decreases, if not the inflow of water to the rudder is increased otherwise, e.g. by increasing the revolutions of the main propeller) while the capability of the thruster-system to produce changes in the heading of the vessel will increase (the capability of the thruster-system to produce transverse force increases when the speed of the vessel decreases). This marks a transition from S1 to a new control situation S2 in which only the thruster-system is capable of producing changes in the heading of the vessel.

At present, the human-machine system on the navigating bridge does not reflect the changes in the capabilities of means-systems (changes in the ability to bring about). And only in some phases of a sea passage the control action norms are made explicit, e.g. in terms of passage plans in open waters presented by a Voyage Management System. In order to increase safety and efficiency it is desirable to have supervision systems that support the crew's assessment of the prevailing control situation[16].

The Relationship between Control Situations and the Work Domain

In order to enable the design of human-machine systems that are sensitive to the dynamics of control situations we need to solve the following problems: 1) how to specify the content of system state changes referred to by control actions in relation to a specific work domain, 2) how to derive changes in the capability of means-system and the norms on system operation, from invariant structures of the work domain.

Specifying the Content of System State Changes: In order to specify the content of control situations it is necessary to come to terms with the different levels of relevant behaviour (or functions) of the system with which the operators are interacting *through* the human-machine system – the *work domain*. Only on the basis of a specification of the relevant levels of system behaviour can we specify the content of the system state changes that are supposed to be brought about through control actions on the system (see Figure 11).

Rasmussen (1986) has argued that operators supervising and controlling complex systems tend to conceive of the system at different levels of means-end abstraction. The means-end organization of a work domain is a normative functional structure ascribed to the system being controlled. This functional structure is relatively stable although the state of the work domain is changing and the task of the operator is changing. Basically, this is because the overall purpose of the system is invariant, e.g. the purpose of the power plant, to produce electrical energy, does not change.

The levels of means-end abstraction of a specific work domain can be used to determine the content of system state changes that are relevant for control actions, i.e. those system state changes that are supposed to be brought about through control actions on the work domain. Typically, the content of the system state changes that are a result of the manipulation of means-systems is determined by the lowest level of means-end abstraction, whereas the content of the system state changes that are consequences of control actions is captured by the higher levels of means-end abstraction, respectively.

[16] Andersen (1998) originally proposed the idea of having a display system on the navigating bridge that is sensitive to different standard situations. In this paper we investigate the underlying problems of how to design human-machine systems that are capable of adapting to changes in the control situation without having to define a set of standard situations in advance.

An identification of the relevant means-end abstraction levels of a specific work domain requires knowledge of the typical tasks and work activities of the operators supervising and controlling the system. This knowledge can be acquired from field studies of work practice. Figure 12 provides a representation of the maritime work domain of the navigating crew on board large container vessels[17].

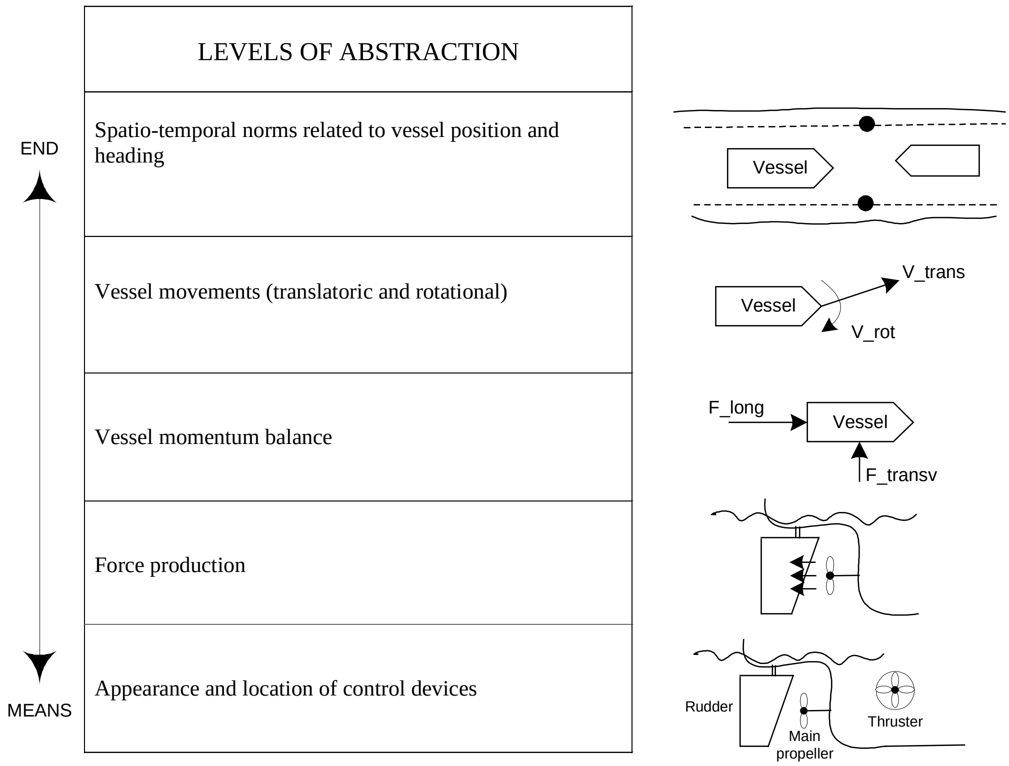


Figure 12. A means-end representation of the work domain of the navigating crew on board large container vessels.

According to Figure 12 the overall purpose of maritime operations (both long-term and short-term) is formulated in terms of spatio-temporal norms related to the vessel position and heading in the horizontal plane. These norms may derive from other types of norms, e.g. fuel minimisation and legislation. Horizontal motion of the vessel leads to changes in position and heading of the vessel and is represented at the next lower level[18]. Also for vessel motion there may be norms specifying constraints on translatoric and rotational speed of the vessel. Vessel motion is a product of physical forces acting on the vessel and at the level below we find the momentum balances of the vessel describing the relationship between controllable and uncontrollable forces and the momentum of the vessel[19]. At the next lower level there is a description of the production of forces (controllable and uncontrollable). The controllable forces acting on the vessel are produced by means-systems involving shipboard control devices (e.g. propeller-system, rudder-system and thruster-system) and devices external to the vessel (e.g. tugs), whereas the uncontrollable forces are a

[17] The given representation of the maritime work domain seems rather general and may apply to other types of maritime operations beyond those performed onboard container vessels. It is important to emphasize, however, that there is no empirical support for such generalizations. See e.g. (Chalmers, Burns and Bryant, 2001) for a work domain analysis of shipboard command and control.

[18] Disclaimer: At present only vessel motion in the horizontal plane is included. Consequently, *roll*, *pitch* and *heave* are not considered in this paper.

[19] Note that the distinction between *kinematics* and *dynamics* found in classical mechanics is preserved in the work domain representation given in Figure 12. The level of *vessel motion* is concerned exclusively with the motion of the vessel (kinematics) while the level below, *vessel momentum balance*, is concerned with the forces (mechanisms) that shape the motion of the vessel (dynamics).

result of e.g. wind and current. Finally, at the lowest level, there is a representation of the appearance and the location of physical systems and components.

When, in a given control situation, there are means-systems having the capability to produce the desired vessel motion complying with given norms then control can in principle be successfully carried out - in Simon's words the *inner* and *outer* environment are mutually appropriate to each other (Simon, 1996).

According to Vicente and Rasmussen (1992) the abstraction hierarchy is supposed to capture the so-called goal relevant constraints at different levels of means-end abstraction, ranging from the basic physical components of a system to its overall operating requirements ascribed to the system. In this paper we make a distinction between different kinds of constraints. At the higher levels (vessel position and heading and vessel movement) the constraints are actually *norms*, i.e. something prescribing purposeful, lawful or safe operation of the vessel, whereas at the lower levels (vessel momentum balance and force production) the constraints are interdependencies among properties of system components based on physical laws.

Deriving the Capability of Means-Systems: In order to derive the actual ability to bring about a system state in a specific control situation, and predict how these are changing we need to track changes in the capability of means-systems to produce such system states (defined relative to the means-end abstraction levels of the controlled system).

The capability of means-systems is determined by the properties of its interacting components. In the rudder-system the water acting on the rudder produces the transverse force acting on the stern of the vessel. The magnitude of the transverse force being produced is determined by the properties of the rudder (area, rudder angle and lift-coefficient) and the water flow past the rudder surface.

The following expression provides an approximation of the transverse rudder force Y_R (Simonsen, 2000):

$$Y_R \cong \tfrac{1}{2}\,\rho\, C_L\, (A_S V_S^2 + A_A V_A^2) \quad (1)$$

ρ is the density of the water, C_L is the lift coefficient of the rudder (proportional to the rudder angle, α), A_S is the lateral area inside the propeller slipstream, V_S is the velocity of water inflow inside the propeller slipstream, V_A is the velocity of water inflow outside the propeller slipstream and A_A is the lateral area outside the propeller slipstream.

The expression (1) is a typical example of a function expressing interdependencies among the properties of interacting entities. Functions express constant relations among the numerical values of (metrical) properties and may be used to state that something is invariably associated with something else. Although mathematical functions make it possible to symbolize and to give precise quantitative descriptions and predictions of connections they fail to state the one-sided genetic connection that characterizes causation (Bunge, 1959). These have to be stated in an extra set of (semantic) propositions.

Functions, together with semantic rules stating the meaning of the variables tied by them, are often useful to tell what happens and why it happens; if a causal meaning can be attached to some the symbols intervening in a function, such an interpreted function will reflect a causal connection. "...functions, which are syntactic forms, cannot *replace* causal propositions; at most, they may take part in the description of causal connections." (Bunge, 1959, p. 95, emphasis in original).

We may view (1) as a description of the causal connection between the flow of water past the rudder (the independent variables V_S and V_A) and the transverse force being produced (the dependent variable Y_R). According to this causal interpretation of (1) the force production capability of the rudder-system is determined by the value of the area of the rudder (A_S and A_A) and its lift coefficient (C_L). Consequently, the connection between the rudder angle α and the lift coefficient C_L describes changes in the force production capability of the rudder-system. Below the causal connections of this interpretation of (1) are described using the notation from Forbus' Qualitative Process Theory (Forbus, 1984) [20]:

[20] (Q1 qprop+ Q2) expresses that Q1 is *qualitative proportional* to Q2, and (Q3 qprop- Q4) expresses the Q3 is *inversely qualitative proportional* to Q4. If a quantity Q1 is qualitative proportional to another quantity Q2, it means that there is a functional relationship between Q1 and Q2, and that Q1 is increasing monotonic in its dependence on Q2 (inversely qualitative proportionalities are defined similarly, with the function being decreasing monotonic) (Forbus, 1984).

Primary connection: Y_R Qprop+ V_S
Y_R Qprop+ V_A

Secondary connection: Y_R Qprop+ α $(0° < \alpha < 35°)$

This interpretation of (1), however, is not compatible with the normal use of rudders, where the rudder angle is the independent control variable (input). When the rudder angle is the input it is preferable to focus instead on the causal connection between the rudder angle and the force being produced (the *primary* causal connection), and hence view the connection between water inflows and the force produced as a second order causal connection describing changes in the force production capability of the rudder-system. It is the changes in the force production capability of the rudder described by this second order connection that we want to track across control situations. Below the causal connections of this interpretation of (1) are shown:

Primary connection: Y_R Qprop+ α $(0° < \alpha < 35°)$

Secondary connection: Y_R Qprop+ V_S
Y_R Qprop+ V_A

Deriving Changes in Control Action Norms: Also the norms imposed on the operation of the controlled system might change across different control situations. In relation to the maritime work domain, the spatio-temporal norms tend to change during the passage of the vessel, e.g. when leaving the harbour and entering the lane leading the vessel towards the open sea. In the harbour there are physical norms given by the harbour area, legal norms on speed of the vessel etc., whereas in a lane there are other types of norms given by traffic separations (legal) and water depth contours (physical).

It is clear that such norms cannot be derived from interdependencies among physical properties. Instead these have to be derived using other types of invariant structures of the work domain, such as maps. To some extent information indicating these norms are already present in the man-machine systems found on modern ship bridges (e.g. via Voyage Management System). This information could, however, be improved by making explicit the order of these norms (physical, intentional or legal).

Conclusions

This paper has addressed important modelling problems associated with the design of human-machine systems that are sensitive to changes in control situations. It was proposed that control situations include the following contextual factors shaping the control actions of a control agent: 1) the current state of the controlled system, 2) control action norms, 3) the ability to bring about (determined by the capability of means-systems of the controlled system), 4) the ability to do (determined by the capability of control means) and 5) disturbances. Especially, *the ability to bring about* was discussed at length in this paper. Based on an account combining a causal and a means-end view we were able to show that the ability to bring about is given by the capability of the means-systems to produce state changes in the controlled system.

Furthermore, the relation between control situations and the work domain was discussed. That is, how to specify control situations in relation to a representation of a specific work domain (specifying the content of the state changes referred to by control actions) and how to derive changes in control action possibilities and the norms imposed on control actions from different kinds of invariants of the work domain.

Acknowledgements

This work is funded by the Danish National Research Foundation, Center for Human-Machine Interaction.

References

Andersen, P.B. (1998). Analysis of Maritime Operations 4. Center for Human-Machine Interaction. CHMI-2-98

Bunge, M. (1959). Causality. The place of the causal principle in modern science. Harvard University Press, Massachusetts.
Bunge, M. (1977). Treatise on basic philosophy (Vol. 3). Ontology I: The furniture of the world. D. Reidel Publishing Company. Dordrecht-Holland/Boston-USA.
Chalmers, B.A., Burns, C.M. and Bryant, D.J. (2001). Work Domain Modeling to Support Shipboard Command and Control. 6th International Command and Control Research and Technology Symposium, June 19-21, 2001, U.S. Naval Academy, Annapolis, MD.
Forbus, K.D. (1984). Qualitative Process Theory. Artificial Intelligence, Vol. 24, pp. 85-168.
Harré R. and Madden, E.H. (1975). Causal Powers. A theory of natural necessity. Basil Blackwell. Oxford.
Petersen, J. and Nielsen, M. (2001). Analyzing Maritime Work Domains. Proc. of 8th Conference on Cognitive Science Approaches to Process Control. 24-26 September 2001.
Rasmussen, J. (1986). Information Processing and Human-Machine Interaction. An Approach to Cognitive Engineering, Amsterdam: North-Holland.
Simon, H.A. (1996). The sciences of the artificial – third edition. The MIT press. Cambridge, Massachusetts.
Simonsen, C.D. (2000). Rudder, Propeller and Hull Interaction by RANS. Ph.D. thesis. Department of Naval Architecture and Offshore Engineering, Technical University of Denmark. ISBN 87-89502-33-7.
Vicente, K.J. and Rasmussen, J. (1992). Ecological Interface Design. IEEE Transactions on systems, man, and cybernetics. Vol. 22, No. 4, July/August 1992.
Von Wright, G.H. (1971). Explanation and Understanding. Cornell University Press. Ithaca, New York.
Woods, D.D. (1991). The cognitive engineering of problem representations. In: Weir, G.R.S. and Alty, J.L. (Eds.), Human-Computer Interaction and Complex Systems.

A Formative Approach to Designing Teams for First-of-a-Kind, Complex Systems

Neelam Naikar[1], Brett Pearce, Dominic Drumm, and Penelope M. Sanderson[2]

Defence Science and Technology Organisation and [2]Swinburne University of Technology
[1]PO Box 4331, Melbourne, VIC 3001, Australia; neelam.naikar@dsto.defence.gov.au.

Abstract: Standard techniques for team design, which are based on normative and descriptive approaches to work analysis, cannot readily be applied to first-of-a-kind, complex systems during the early stages of system development. In this paper we present a formative approach to team design based on Cognitive Work Analysis (CWA). We also discuss how we have applied this approach to design a team for a new military system called Airborne Early Warning and Control. This case study shows that the CWA-based approach to team design is both feasible and useful. By designing a team during the early stages of system development, the CWA-based approach helps to guard against the possibility that the team design for a new system is simply a default of the technical-subsystem solution. Instead, the CWA-based technique offers a means for ensuring that the technical-subsystem solution supports the proposed team design.

Keywords: team design, cognitive work analysis.

Introduction

When we invest in the development of new systems, our expectation is that the new systems will offer more effective and efficient ways of working than older systems. Our experience with military acquisitions, however, is that the design effort during system development is heavily focussed on the technical subsystem. Little attention is given to the design of the work structure and work processes of the people that will be manning the system. Yet, these elements of system design are just as critical as the technical solution for fulfilling system goals. As a result, although a system design may look attractive on paper, the potential offered by a new technology may not be fully realised in practice. To avoid this situation, a systematic and pre-emptive analysis of the new ways of working with a new technology is essential (Rasmussen, 1991).

Our focus in this paper is the design of teamwork for new, complex systems. In particular, we focus on first-of-a-kind systems (Roth & Mumaw, 1995). These are systems that have no close existing analogues because, for example, technological advances have led to vastly improved functionality compared to older systems. Thus, the behaviour of workers in first-of-a-kind systems cannot be inferred from workers in older-generation systems. Second, we focus on systems when they are at the early stages of development. During these stages, detailed information about the behaviour of workers in the new system is unavailable. In the following sections we demonstrate that systems with these characteristics require novel approaches to team design.

Standard Techniques for Team Design

When a system will be populated with several workers a number of issues must be addressed about the team design that is best for fulfilling system goals. Some of these issues are: (1) team size; (2) number of levels of hierarchy; (3) number of subteams; and (4) whether workers should have dedicated roles and responsibilities or whether they should be multi-skilled. Typically, these decisions are based on an analysis of the work requirements of the new system. The aim is to design a team that will best fulfil the work requirements of the system. Thus, the work analysis itself is critical for designing effective teams.

Standard techniques for team design (e.g., Davis & Wacker, 1982; 1987; Hackman & Oldham, 1980; Lehner, 1991; Medsker & Campion, 1997) are based on descriptive or normative forms of work analysis. Descriptive techniques identify the work requirements of a system by observing or measuring workers in their physical work context. However, during the early stages of development, the new system and its workers don't exist in a physical form. Descriptive techniques could be used to design teams for new systems by studying older-generations of the proposed system. However, basing the design of new teams on older systems is potentially dangerous because previously unproductive ways of working may be inadvertently incorporated into the new design. Moreover, potentially effective ways of working that are offered by the new technology may be left uncovered (Vicente, 1999; Woods, 1998).

Normative techniques describe the work requirements of a system in terms of a stable set of tasks or procedures. However, workers will usually develop new ways of using a system as they gain experience with it, and they will also invent new ways of working to deal with unexpected contingencies. It is therefore difficult to specify a complete set of tasks or procedures ahead of a system being put into operation. Normative approaches to work analysis may therefore lead to team designs that are not well suited to work requirements that could not be specified up front. Normative approaches may also lead to team designs that do not provide workers with the necessary flexibility for dealing with unanticipated situations (see Vicente, 1999 for a full description of normative and descriptive forms of work analysis).

Cognitive Work Analysis – A Formative Approach

Cognitive Work Analysis (CWA) offers a formative approach to work analysis because it focuses on the fundamental boundary conditions on system safety and performance (Rasmussen, Pejtersen & Goodstein, 1994; Vicente, 1999). These boundary conditions or constraints shape workers' behaviour by imposing limits as well as offering possibilities for safe and effective action. In most work systems there is a very large number of possible sequences of behaviour that do not violate the boundaries or constraints of the workspace. Thus, an analysis of the sequences of tasks or behaviour that typically happen or that should happen is bound to be incomplete. Rather, workers' behaviour is more robustly described by the constraints that shape or will shape the sequences of behaviour in the first place. In addition, an analysis of constraints does not rely exclusively on details about physical-system implementation and workers' behaviour. Hence, CWA can be conducted prior to a system being developed and prior to populating it with workers.

CWA has five analytic techniques that focus on different types of boundary conditions or constraints: (1) Work Domain Analysis identifies the high-level purposes, priorities and values, functions, and physical resources of a work domain, (2) Activity Analysis or Control Task Analysis focuses on the activity that is carried out in the work domain, (3) Strategies Analysis identifies different strategies for carrying out the activity, (4) Socio-organisational Analysis focuses on who carries out the work and how it is shared, and (5) Worker Competencies Analysis identifies the competencies required by workers to carry out the work of the system.

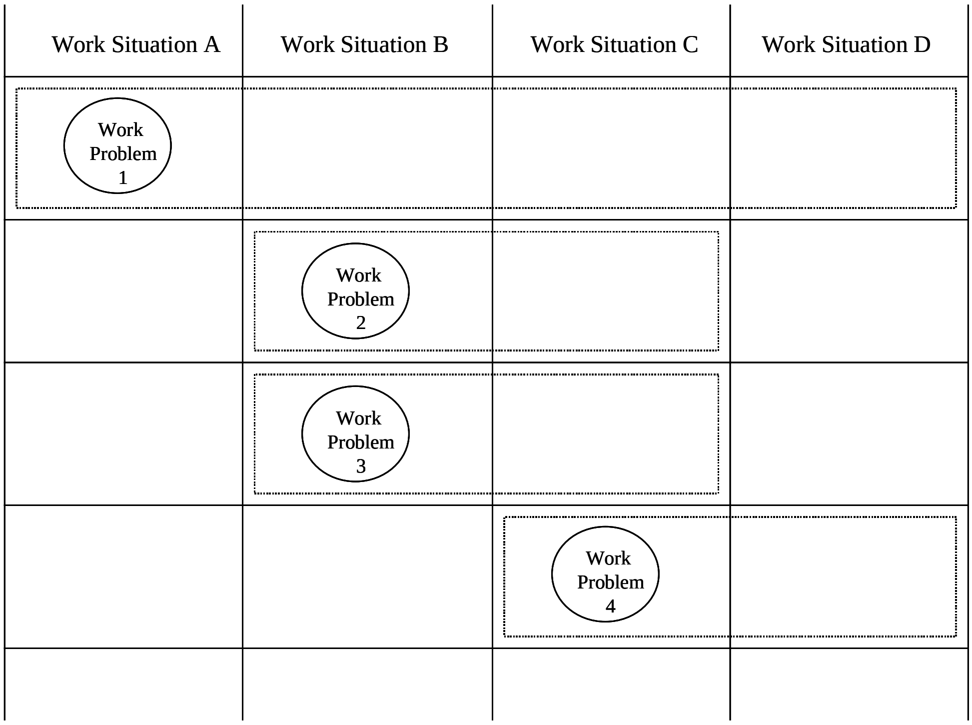
A CWA-Based Approach to Team Design

The CWA-based approach we have developed for designing teams is based on Work Domain Analysis and Activity Analysis and the use of a walkthrough technique to explore the feasibility of alternative team designs for a proposed work system. The first step is to conduct a Work Domain Analysis in order to identify: (1) the functional purposes or high-level objectives of the proposed work system, (2) the priorities and values that must be preserved during system operation, (3) the general functions or everyday functions that the system must coordinate and/or execute to fulfil the functional purposes, (4) the physical functionality afforded by the physical devices of the system, and (5) the physical devices themselves.

The second step is to conduct an Activity Analysis in work domain terms (Rasmussen et al., 1994). Here, the aim is to identify the activity that is required in a work domain for a system to fulfil its functions, priorities and values, and purposes, given a set of physical resources. Rasmussen et al. (1994) suggest identifying activity in terms of a set of work situations that workers must participate in or the set of work problems that workers must solve in order to fulfil work-domain constraints. In our experience with military systems we have found it useful to combine the two (Figure 1).

The third step involves using a walkthrough technique and the products of the Work Domain Analysis and the Activity Analysis to explore the feasibility of alternative team designs for a proposed work system. More specifically, the work problems from the Activity Analysis are used to model workers' activity as a function of alternative team designs in different scenarios. The Work Domain Analysis is used to evaluate the alternative team designs in terms of how well the different patterns of activity support the functions, priorities and values, and purposes of the system, as well as how effectively the physical resources of the system are employed.

Figure 1: Generic illustration of work problems against a backdrop of work situations; each work problem can occur over several work situations.



Application of CWA-Based Approach to Team Design

In this section, we discuss how we used the CWA-based approach to design a team for a new military system called Airborne Early Warning and Control (AEW&C). AEW&C is a complex airborne system that is currently being manufactured by Boeing for the Australian Defence Force; this contract was signed in December 2000 and the first aircraft is scheduled for delivery in 2006. When it is operational, AEW&C will be manned by a team of people in the cabin of the aircraft, who will be responsible for developing a situation picture of an allocated area of operations, and for coordinating the activities of defence assets in the area. Thus, this role is similar to the roles of the Airborne Warning and Control System (AWACS) of the United States Air Force and the E2C system of the United States Navy. A key concern of the AEW&C System Program Office (the organisation within the Australian Department of Defence that is responsible for the AEW&C system acquisition) is the design of the AEW&C team that will best facilitate system performance and safety.

AEW&C Work Domain Analysis: To conduct the AEW&C Work Domain Analysis we relied on various defence documents and input from subject matter experts, including military experts, scientists, operations analysts, and engineers (see Naikar & Sanderson, in press). The AEW&C Work Domain Analysis describes: (1) the Functional Purposes or high-level objectives of the AEW&C system (e.g., early warning and control of events in an allocated area of operations), (2) the Priorities and Values that will be preserved during AEW&C operation (e.g. knowledge edge), (3) the Purpose-related Functions that will be executed and coordinated on AEW&C missions (e.g., representation of the tactical situation), (4) the Physical Functions afforded by the physical devices of the AEW&C platform (e.g., information sensing, information exchange), and (5) the Physical Devices themselves (e.g., radar, radio voice links).

AEW&C Activity Analysis in Work Domain Terms: For AEW&C, we identified the set of work situations that workers must participate in as a set of mission phases, for example, enroute to station, on station, return to base. In addition, we identified 10 work problems including manage crew, develop the Recognised Air Surface Picture (RASP), and manage asset disposition. As in Figure 1, the AEW&C work problems were represented against a backdrop of mission phases (Naikar & Pearce, 2001). We also developed definitions for each work problem. For example, for 'manage crew' the problem is to distribute tasks, assets, and other resources among crew in order to support the aims of the mission under changing tactical and environmental conditions.

AEW&C Walkthrough: Once we had completed the AEW&C Work Domain Analysis and Activity Analysis we were able to use the products of these analyses together with a walkthrough technique to explore the feasibility of alternative team designs for AEW&C. The walkthrough itself involved a further five steps.

The first step of the walkthrough involved identifying the team design variables to examine. Some of the issues that the AEW&C System Program Office were particularly concerned about included the size of the team, the number of levels of hierarchy in the team, whether the team should be decomposed into subteams, and the skill sets of workers (i.e. whether workers should have dedicated roles and responsibilities or whether they should be multi-skilled). We then identified the set of values for each team-design variable that were plausible for AEW&C. So, for example, for the variable of team size, values between six and ten were considered plausible whereas for the variable of number of levels of hierarchy, values of two and three were considered plausible.

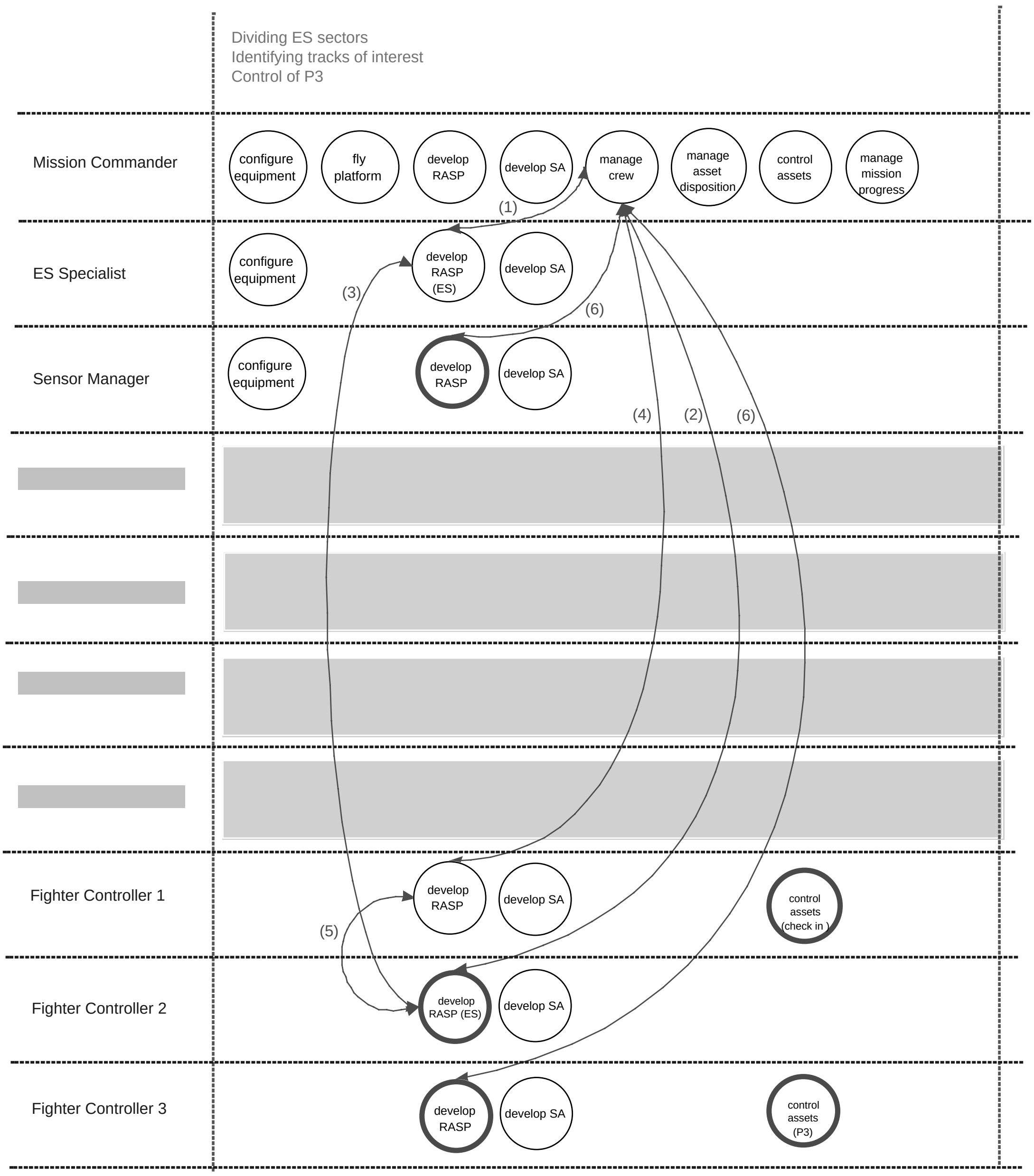
The different combinations of values for the team-design variables then specified the alternative team concepts to examine. For example, four of the team concepts that we examined were: (1) a ten-person crew with two levels of hierarchy, no subteams, and workers with dedicated roles; (2) a ten-person crew with two levels of hierarchy, no subteams, and workers with multiskilled roles; (3) a six-person crew with two levels of hierarchy, no subteams, and workers with dedicated roles; and (4) a six-person crew with two levels of hierarchy, no subteams, and workers with multiskilled roles.

The second step of the walkthrough involved working with Australian Defence Force personnel to develop air defence scenarios that were representative of routine missions for AEW&C (e.g., conducting general surveillance in the northern regions of Australia) as well as missions representing more exceptional circumstances (e.g., supporting allied forces during battle). Each of the mission scenarios were divided into segments or epochs during which a coherent set of mission events occurred (e.g., strike package enters no fly zone, hostile firing of friendly ships). One of the scenarios that we used to examine the four team concepts specified above was an 8 hour mission involving focal area surveillance and broad area surveillance. There were 48 major assets in the blue force and significant hostile activity.

The third step of the walkthrough involved working through each scenario with subject matter experts and asking them to outline the activity of each crew member, as a function of every epoch in a scenario, assuming a particular team concept. We kept a record of crew activity on a whiteboard, and we also took notes to supplement the record on the whiteboard. In particular, we noted how work allocations would be negotiated and implemented and the criteria used for allocating work to crew members (e.g., likely workload of crew members; requirements for information sharing; likely future demands of the mission). Where there was a reallocation of responsibilities among crew members, the communication and coordination requirements were also noted.

The fourth step of the walkthrough involved translating the data from the whiteboard and the analysts' notes into a representation of crew activity in terms of work problems (from the AEW&C Activity Analysis in Work Domain Terms) for every epoch in a scenario. Our representations of entire scenarios are too large to be reproduced legibly here. So, in Figure 2, we focus on just one epoch of a scenario to illustrate the kinds of information that were captured. In essence, the representations describe: the critical events at each epoch of the scenario (top row); the roles of each of the crew members (first column); the work problems that crew members are preoccupied with (circles); the reallocation of work problems and associated coordination requirements (arrows); and the added responsibilities of crew members, given a reduction in team size from 10 to 6 (bold circles). The rows that are shaded represent those individuals in the 10-person team who were not part of the 6-person team that is illustrated in the figure.

Figure 2: Representation of crew activity in terms of work problems for one epoch in a scenario.



(1) ES Specialist discusses the need for additional ES support with the Mission commander
(2) The Mission Commander directs Fighter Controller 2 to assist with the ES analysis
(3) The ES Specialist negotiates the distribution of ES analysis with Fighter Controller 2
(4) The Mission Commander asks Fighter Controller 1 to assist with develop RASP
(5) Fighter Controller 2 hands over the work problem of develop RASP to Fighter Controller 1 (briefing on outstanding tasks, identifying significant details etc)
(6) The Mission Commander informs the Sensor Manager and Fighter Controller 3, who are also performing develop RASP, of this change to tasking.

In the fifth step of the walkthrough, we used the profiles of crew activity to examine how the work demands of a mission would be distributed and managed by crews with different team designs. In particular, we were able to compare: (1) the number of crew members allocated to performing each work problem, (2) the number and types of work problems allocated to individual team members, (3) the number and types of assets (e.g., fighters, strike aircraft) controlled by each crew member, (4) the number and types of instances in which work was reallocated across crew, and (5) the coordination requirements associated with each reallocation. For example, in comparing a six-person team with multiskilled crew and a ten-person team with multiskilled crew, we found that there were fewer instances of work reallocation in the ten-person team than in the six-person team.

We then examined the impact of the different patterns of activity on the functions, priorities and values, and purposes of the AEW&C work domain. For example, in comparing a six-person team with crew that had dedicated roles and responsibilities and a six-person team with crew that were multi-skilled, we observed that at least one crew member in the dedicated team was solely concerned with developing the Recognised Air Surface Picture (RASP) throughout the scenario. In contrast, individuals in the multi-skilled team who had this responsibility were also controlling and managing assets at critical points in the scenario. Given the high workload associated with coordinating and protecting a large number of assets in a hostile airspace, these results suggest that the multi-skilled team may find it more difficult than the dedicated team to fulfil the AEW&C Purpose-related Function of representation of the tactical situation and the AEW&C Priority and Value of maintaining a knowledge edge.

On the basis of these types of results we were able to generate requirements for a new team design for AEW&C and, subsequently, to specify a team design that fulfilled these requirements. When we presented this team design to the AEW&C System Program Office, military experts (including those with backgrounds in AWACS and E2C operations) judged that this team design was better than the designs they had independently generated for AEW&C in the past. As a result, the AEW&C System Program Office adopted the team design that we developed for AEW&C operations. In addition, our analyses led to modifications of the AEW&C technical-system specification so that it better supports AEW&C team work. These alterations were made prior to the contract being signed by Boeing and therefore at no cost to the Commonwealth of Australia.

Conclusion

In this paper we have presented a new approach to team design based on CWA, and we have illustrated that this technique is useful for designing teams for new systems when detailed information about the technical-subsystem solution and workers' behaviour is still unavailable. By defining a team concept during the early stages of system development, the CWA-based approach helps to guard against the possibility that the team design for a new system is simply determined by default, after the technical subsystem is put into place. Rather, this technique can be used to guide the development of the technical-subsystem solution so that it supports the proposed team design.

We acknowledge, however, that there is currently no empirical support for our CWA-based approach to team design. While this too is part of our future research program, the reality is that, like AEW&C, the development of many new systems will proceed in the absence of this data. Moreover, some have argued that within disciplines such as human factors engineering, many of the important results are qualitative rather than quantitative and that within these areas the truly significant research results are often concepts rather than data (Rouse, 1985). On the basis of the arguments presented in this paper, we believe that it is reasonable to assume that a CWA-based approach to team design will offer better results for first-of-a-kind, complex systems than team designs based on intuition, informal analyses, or conventional approaches to team design.

Acknowledgements

We thank: the AEW&C System Program Office, in particular Squadron Leaders Antony Martin and Carl Zell for their support; Tracey Bryan, the principal subject matter expert for the Cognitive Work Analysis; and Anna Moylan of the Defence Science and Technology Organisation for her assistance.

References

Davis, L. E., & Wacker, G. L. (1982). Job design. In G. Salvendy (Ed.), *Handbook of Industrial Engineering* (pp. 2.5.1-2.5.31). New York: Wiley.

Davis, L. E., & Wacker, G. L. (1987). Job design. In G. Salvendy (Ed.), *Handbook of Industrial Engineering* (pp. 431-452). New York: Wiley.

Hackman, J. R., & Oldham, G. R. (1980). *Work Redesign.* Reading, MA: Addison-Wesley.

Lehner, P. E. (1991). Towards a prescriptive theory of team design. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics: Decision aiding for complex systems* (pp. 2029-2034). New York, NY: Institute of Electrical and Electronics Engineers.

Medsker, G. J., & Campion, M. A. (1997). Job and team design. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 450-489). New York: John Wiley and Sons Inc.

Naikar, N. & Pearce, B. (2001). Analysing activity for new, complex systems with cognitive work analysis. *Proceedings of the 37th Annual Conference of the Ergonomics Society of Australia*, pp217-222. Sydney, Australia. Nov 27-30.

Naikar, N., & Sanderson, P. M. (in press). Evaluating design proposals for complex systems with work domain analysis. *Human Factors.*

Rasmussen, J. (1991). Modelling distributed decision making. In J. Rasmussen, B. Brehmer, & J. Leplat (Eds.). *Distributed decision making: cognitive models for cooperative work* (pp. 111-142). Chichester: John Wiley & Sons Ltd.

Rasmussen, J., Pejtersen, A., & Goodstein, L. P. (1994). *Cognitive Systems Engineering*. New York: Wiley.

Roth, E. M., & Mumaw, R. J. (1995). Using cognitive task analysis to define human interface requirements for first-of-a-kind systems. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 520-524). Santa Monica, CA: Human Factors and Ergonomics Society.

Rouse, W. B. (1985). On better mousetraps and basic research: Getting the applied world to the laboratory door. *IEEE Transactions on systems, man, and cybernetics: 15(1): 2-8.*

Vicente, K. J. (1999). *Cognitive work analysis: Towards safe, productive, and healthy computer-based work*. Mahweh, NJ: Lawrence Erlbaum & Associates.

Woods, D. D. (1998). Designs are hypotheses about how artifacts shape cognition. *Ergonomics, 41*(2), 168-173.

Qualitative Analysis of Visualisation Requirements for Improved Campaign Assessment and Decision Making in Command and Control

Claire Macklin, Malcolm J. Cook, Carol S. Angus, Corrine S.G. Adams, Shan Cook and Robbie Cooper

QinetiQ, Rm 1024, A50 Building, Cody Technology Park, Farnborough, Hants., GU14 0LX, England.
E-mail: cmmacklin@qinetiq.com
University of Abertay Dundee, Dundee, DD1 1HG, Tayside, Scotland.
E-mail: m.cook@abertay.ac.uk

Abstract: This research aims to develop visualisations to support military command and control teams in situation assessment, facilitate their consequent decision making and support the selection of appropriate courses of action. By understanding the cognitive and social processes underlying campaign planning and situation assessment, visualisation techniques can be designed that support users' mental models. These visualisation techniques will result in decision dominance via improved campaign assessment (Macklin and Dudfield, 2001). Design processes should be supplemented by psychological information in order to ensure that technology supports people in the way they *actually make decisions* and provides them with the information and cues needed in a complex, dynamic environment to support naturalistic decision making (Klein, 1993). Critical Decision Method (CDM) interviews were carried out with a number of senior military decision-makers. The interview transcripts were coded using a framework adapted from previous research into socio-cognitive processes in command teams (Cook and Cooper 2001a, 2001b). The issues identified in the transcripts provide a basis for design of campaign visualisation tools as they identify the important processes, structures and activities that ensure effective functioning of command team members.

Keywords: visualisation, social, cognitive, command and control, decision-making

Introduction

Visualisation is about the representation of information to facilitate the formation of an internal representation. More formally information visualisation has been defined as "the use of computer-supported, interactive, visual representations of abstract data to amplify cognition" (Card, Mackinlay and Schneiderman, 1999, pp7). Representations and visualisations (using technologies such as information visualisation, graphics databases, multimedia, animation) can aid human cognition because they provide a visual framework that allows information and knowledge to be organised. They are of particular use in complex and dynamic domains (such as the military) where understanding relationships among sets of variables is crucial in order to comprehend important domain-relevant concepts (Goldman and Petrosino, 1999). Knowledge is retained only when it is embedded in some organising structure, such as a mental model (Resnick, 1989). Presenting information in multiple ways facilitates the process of linking it to existing knowledge, applying it to other domains and retaining it, particularly in complex and ill-defined environments (Spiro, Vispoel, Shmitz, Samarapungavan and Boerger, 1987). The benefits of alternative representations are that they demonstrate different aspects of the same situation, highlight critical functional relationships and create multiple encodings and deeper processing (Goldman and Petrosino, 1999).

Humans are very effective at representing large amounts of highly complex and multivariate information visually. Mental representations are accurate and durable (Anderson, 1995; Spiro et al, 1987). Typical experiments show results where subjects' memory for previously shown visual material is very accurate and much better than that for textual material, but it is the meaning of a picture that is remembered rather than its actual, specific detail (Anderson, 1995). Incoming information embedded in a coherent framework and linked to existing knowledge enables people to assign meaning to it (and thus construct mental models) (Resnick, 1989). Therefore providing people with a visual framework (a visualisation) around which to organise incoming knowledge, will aid them in processing and therefore truly comprehending that information. Mental models give a form to knowledge and a framework on which to build – an organising structure through which further information can be interpreted (Resnick, 1989). Mental models are abstractions of knowledge (imperfect knowledge) that act as an "interpretative framework" to give people a simpler way of predicting, explaining and interacting with complex systems (Moray 1996b). According to mental model theory (Moray, 1996b), the mapping of knowledge forms mental models in long-term memory (also called schemata). These long-term models are activated by a particular task or environment, and loaded into working memory where they become dynamic, running models. Mental models allow people to decrease the amount of information they have to deal with, lightening the cognitive load to get information into a manageable form (a model).

One device used to transfer a reader's concrete knowledge about a familiar object to an unfamiliar object being described is the metaphor. This process of transference can be used in interface design to allow existing knowledge to be applied to a novel situation to aid users' understanding and support construction of a mental model. Metaphors and visual images are used to support mental models in displays due to the strong visual component of mental models. It is for this reason that visual metaphors (if they are congruent with users underlying mental models) can be helpful in supporting interactions with a complex system (MacMillan, Getty, Tatum and Ropp, 1997). Therefore it is important that artefacts used in system design are presented in a form which has meaning to the user, allowing them to interpret data or system output in a way that is meaningful in the domain context (see Figure 1).

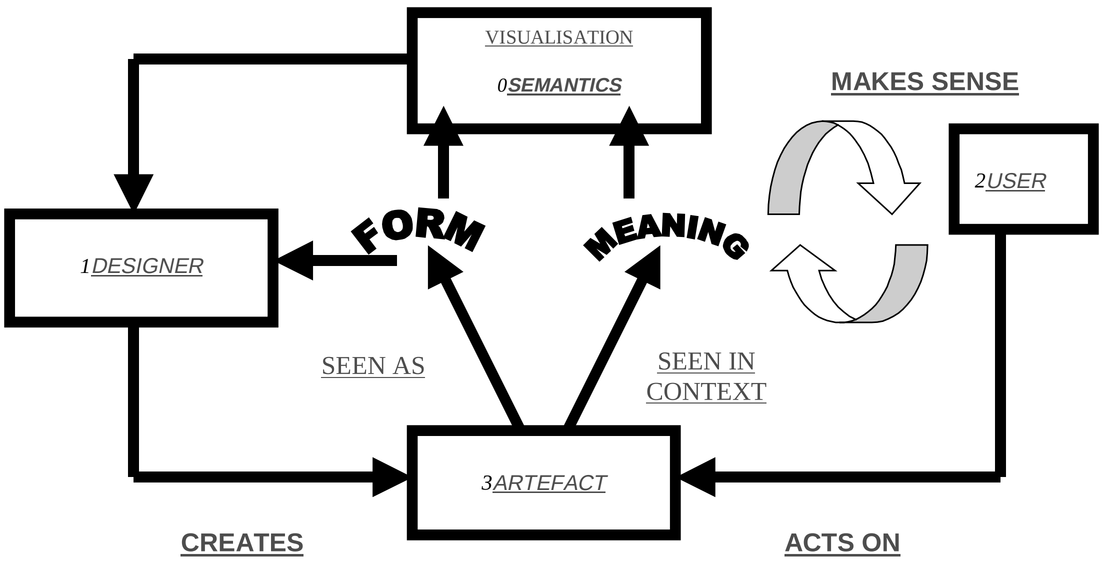


Figure 1 – Use in context as a stimulus to meaning (amended from Krippendorff, 2000)

The key to developing techniques that successfully support users in task performance is ensuring that the technology supports them in the task they are trying to achieve. Not only should visualisations be supportive of users' mental models, but it is important that they support the task in the way it is really done, providing cues needed by the user in that domain to facilitate accurate situation assessment and lead to better decision making. *Appropriately designed* visualisations will support users engaged in situation assessment and command and control but only if designed with an understanding of the cognitive and social processes underlying the actual tasks performed (Macklin and Dudfield, 2001). Poorly designed visualisations could lead to an inadequate or inaccurate understanding of the data leading to sub-optimal situation assessment. Decisions based on this assessment could in turn be negatively affected. Visualisations that do not meet users' information needs, or which interfere with operators' decision and inference strategies can lead to operational failures, incidents and accidents (Klein, 1993; Macklin and Dudfield, 2001). A key aspect of the current difficulty in processing and using information efficiently is the lack of capability to articulate visualisation requirements (Alexander, 2000). This research attempted to address these two crucial issues by gaining an in-depth understanding of the decision requirements of military commanders as the basis of the visualisation design process.

Method

To design visualisation tools to support campaign assessment in command teams an understanding of the decision requirements of military staff in a command and control setting is necessary. It is important to know what cues command teams use in understanding a situation that will contribute to later decision making. Therefore a Naturalistic Decision Making (NDM) framework has been applied (Klein, 1993; Klein 1997; Diggins, 2000). Naturalistic models of decision-making capture the adaptive characteristics of real world behaviour in a dynamic and ill-structured environment. They focus on the way people actually make decisions and how knowledge structures are created and adjusted, which is a key part of the campaign assessment process (Cohen, 1995). Research into command and control decision making suggests that naturalistic theories account for over 80% of all decisions studied (Henderson, 1999). Rational models of decision making have been used as a basis for tools to support command and control functions such as decision making and selection between alternative courses of action. However, while these compensate for the biases in decision making that humans exhibit when making judgements under uncertainty (Fischhoff, 1988), they fail to support military inventiveness. Command and control systems designed as “decision aids” often do not allow the exploitation of commanders’ experience in inferring the enemy’s intent. Nor do they support experienced military commanders in retaining agility and surprise by making bold decisions which involve elements of calculated risk (Lanir, Fischhoff and Johnson, 1988; Klein, 1993). Tools need to be designed which support command teams in making decisions in the way military decisions are actually made, rather than dictating how decisions should be made. According to NDM theory, situation assessment is the most important aspect of decision making, so when designing interfaces for systems it is crucial that the operator should be assisted in gaining a good situational understanding. In order to understand *how* the command team make their situation assessments, their decision-making strategies and inferences need to be thoroughly understood. However, this is often difficult especially where expert knowledge has become ‘tacit’ due to proceduralisation of skills. It has also been argued that people are inaccurate when asked to speculate about their own cognitive processes or are unable to do so (Fischhoff, 1988). Cognitive task analysis methods therefore use critical incidents as a basis. These methods are concerned with critical incidents, which are more memorable than routine tasks (that may have been performed in a largely automatic manner) and lead to fairly accurate recall (Klein, 1993). Cognitive task analysis methods can be used to understand and elicit how a person actually experiences and performs a task, allowing the generation of decision requirements which can inform system/interface design (Klein, 1993). CTA is particularly appropriate to the complex, dynamic and uncertain command and control environment with its requirement to perform multiple (often ill-structured) tasks requiring a large conceptual knowledge base (Gordon and Gill, 1997). Klein (1993) recommends that identifying decision requirements and using these to guide system design should result in systems and interfaces that work better in supporting users in their tasks.

Various methods of conducting cognitive task analysis exist (critical incident, interview, observational and analytical methods). Critical incident techniques such as the critical decision method (CDM) are useful for capturing details in context and providing a rich source of data (Klein, 1993). CDM is an interview conducted in four cycles, which works best when the interviewee has a certain amount of expertise and personal experience in the domain. It uses a number of cognitive probes, which can be modified as appropriate, to understand the processes underlying decisions made. The cues allow the capture of unconscious or implicit knowledge because of the manner in which the interviewee is required to recall the incident to answer these questions. They capture information that it may not be possible to elicit just by asking for a recollection of the incident such as: key decisions and cues entering into the decision; types of inferences involved and strategies used for making them; sources of confusions; and types of knowledge gained through experience (Klein, 1993). CDM has been used successfully in a number of applied settings (Klein, 1993; Chandler, Emery and Routledge, 2000). Henderson and Pechy (1999) give examples of its successful implementation in the design process and suggest that using CDM as a basis for design may significantly improve the effectiveness of new technologies to support command teams. Therefore a CDM interview appropriate to command and control was developed and refined based on the processes and cognitive probes used by Klein (1993); Hoffman, Crandall and Shadbolt (1998); Chandler, Emery and Routledge (2000). To ensure that the methodology considered all the pertinent psychological factors influencing campaign assessment, it was informed by the framework developed by Cook and Cooper (2001a). This framework identifies 11 first-order categories (and associated sub-categories) of social and cognitive characteristics involved in campaign combat information management for command and control teams.

Analysis

The summary of the analysis provided here is explained in more detail in Angus, Adams, and Cook (2002a; 2002b).

Formatting the Transcripts: To enable the transcript analysis, the transcripts were presented anonymously and in the same format, which was well defined to aid analysis using a qualitative analysis framework developed in an earlier project (Cook and Cooper, 2001a, 2001b).

- For easy reference, the interviewees' conversation was presented in italic in comparison to normal type used for the interviewer.
- The interviewees' conversational paragraphs were coded using an ordinary numbering sequence. This was methodically completed and included any utterances of sounds or simple one word answers such as 'yes'.
- Each interviewee was allocated a participant code for reference purposes.

Once the transcripts were formatted to experimental standard they were presented initially to two independent raters, one of whom had more experience in military command issues. This preliminary analysis was cross-checked by a third independent rater who had initially helped to construct the proposed analytical framework for qualitative analysis (Cook and Cooper, 2001a, 2001b). The higher order factors were derived from clustering sub-categorical information (in the primary categories shown in Figure 2) along key dimensions related to psychological and operational constructs (shown in Figure 3).

Higher Order Category	**Primary Categories Within**
Cognitive Factors	Decision Making Information Analysis **(Information and Knowledge)** Situation Awareness Implicit Cognition **(Gut Instinct)** Perceptual **(Display Technology)**
Social, Organisational & Role Related Factors	Hierarchical Position **(Hierarchy)** Organisational Structure **(Organisational)** Team and Group Factors (includes **Behavioural factors)** Communication
Contextual Factors	Knowledge Domain (Intelligence) **(Enemy Characteristics)** Operational Domain (Environment) **(Environmental)** Factors in Warfare
Time and Resource Management Factors	Time Frame Co-ordination
Campaign Management	Any remaining utterances directly related to military protocols that could be initially coded under other sub-categories but which did not fall directly into the four higher categories of cognitive, social, contextual or time related factors above.

Table 1 - The basic and higher order categorisation used to classify comments from transcripts. The labels in bold indicate the alternative label used in Figure 2.

The same, more experienced, third rater developed a final qualitative analysis which sorted material with reference to four higher order categories, which subsumed the original categories developed for the qualitative analysis. In addition, any comments specifically addressing communication and display technology issues were grouped according to two sub-categorical features of phase of campaign management, being concerned with planning or execution.

Coding the Conversational Paragraphs: As noted above, each of the interviewees' conversational paragraphs was systematically numbered, no matter the response given, whether that be a simple sound response or a one word answer. The only time the conversation may not have a paragraph number within the transcripts is where an interruption may have occurred and the conversation was of no relevance. The raters were requested when categorising the transcripts to present the selected using the allocated letter code, followed by the actual text and then the paragraph number at the end of the text in brackets.

Interviewee Coding: The coding occurred in two parts, that of category and individual. The category was the interviewee's position held at the time of the interview and the letter simply was a method of identifying between the six transcripts. In this case, the initial of the last name was used.

Scripts: The six interview transcripts completed for this study included interviewees taken from all three services and from the following ranks; Wing Commander, Squadron Leader, Lieutenant Colonel, Commander and Major.

Results

The qualitative analysis produced a large number of entries that filled most of the original categories developed in the qualitative framework from previous research (Cook and Cooper, 2001a, 2001b). This supports the view that the visualisation process in command teams is likely to be a socio-cognitive process along the lines identified by Vertzberger's (1998) model of military decision making. Cook and Cooper (2001a) identify the various social processes that have an influence on team cognition such as groupthink and diffusion of responsibility. They identify military incidents that demonstrate the effects of these processes on team cognition. Other authors (e.g. Hosking and Moreley, 1991) have also argued that social and cognitive processes cannot be separated, particularly in a team environment. The socio-cognitive nature of the process is indicated by the frequency of utterances related to social and cognitive processes contained within transcripts, that may not be immediately obvious in the primary encoding shown in Figure 2. However, the encoding of higher order categories in Figure 3 indicates the predominance of cognitive and social issues in the use of the campaign visualisation processes. The figure for social factors would be further increased by a further 5%, to a figure of 20.6 %, if the communication issues contained in utterances codified as campaign management were included. Thus, nearly 70% of the utterances are related to the combined categories of social and cognitive processes.

The second most important category was labelled contextual issues and the lesser category related to time and resources was fourth largest. It is interesting that the operational requirements of the visualisation process are categorised as lesser issues than social and cognitive issues. This may reflect the strong domain knowledge of the participants in the study and their confidence in their expertise. Contextual, time and resource factors were identified as highly significant predictors of success and failure in historical analyses used to generate the original framework developed by Cook and Cooper (2001a). It is re-assuring to find that they are still strong factors in the transcript analysis, but the ability to use knowledge is considered secondary. It is the issues of sharing knowledge across the command team in the form of a mental model and the processing of the information by the individuals that dominates the thinking of the experienced command team members.

However, a number of deficiencies in the qualitative framework were noted because additional levels of detail were needed to discriminate adequately information presented in a number of cases. There was also a need to cluster or group the detailed categories into higher order categories, that might be used to aid and inform the design process for which the qualitative data was captured (Angus, Adams and Cook, 2002b). The categorisation used is noted above in Table 1. The codified utterances in Figure 2 were an attempt to extract detailed information based upon the dominant themes within the short passages of text identified as a coherent passage of related text. While the categories appear close three judges made clear distinctions regarding the theme of the utterance. The transition to higher order categories in Figure 3 used an assignment based on the original categorisation and where ambiguity existed an analysis of the overall context of the speech utterance.

Discussion

There is no doubt that the visualisation process used to support campaign command teams is both a social and cognitive process. The importance of this is borne out by the breakdown of the profile of utterances in the low-level categorisation (shown in Figure 2) and in the clustering of material in high level categories (shown in Figure 3). To appreciate fully the way in which the task structures and the communication protocols surrounding the use of the campaign visualisation are used, empirical evidence is needed but simple frequency counts cannot represent the richness of the data in the detailed qualitative analysis. The design requirements for visualisation need to address the way that the proposed design will influence the social interaction across the command team and the availability of free cognitive resources to process the available information. Poor or clumsy visualisation could undermine the process it aims to support, by demanding additional cognitive resources and interrupting the free exchange of information across the command team, the decision-making performance may be reduced in quality.

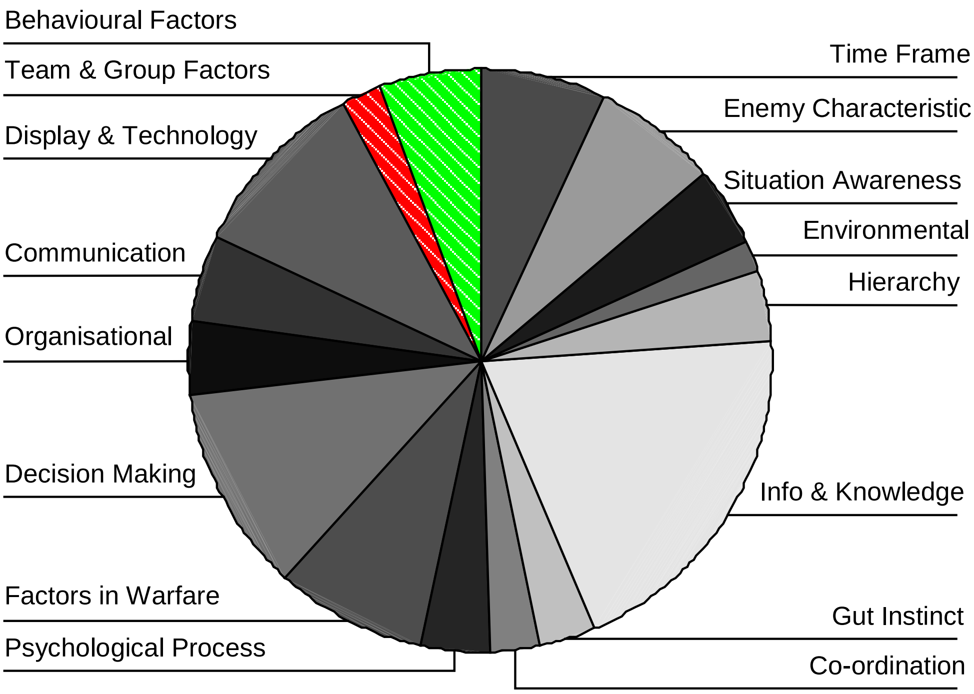


Figure 2 - Pie chart of codified utterances according to detailed categorisation.

The design of any artefact used to extend cognitive capabilities in a team or group environment is difficult in itself because the role definitions for the team members may not be uniform. In addition, the privileges accorded the team members may be strongly defined by hierarchical positions within the team, and the combination of individual information requirements and authority gradients can create a difficult environment in which to enable decision making. Added to these very obvious problems the military environment varies from periods of planning to frenetic periods of activity during execution. Thus, any design requirements model will meet a number of antagonistic and competing demands that could move the design towards any number of antithetical solutions. This general view was substantiated by comments from the participants in this study with many comments referring to factors associated with organisational structure, team behaviour and hierarchical position.

It would be perfectly possible to develop a mechanistic approach to the analysis of the design problem that simply counted the frequency and type of comments made to focus design effort onto those areas which attract more comments. However, many skilled operators use both explicit and implicit cognitive and social mechanisms that may be more or less effectively represented in their comments. Thus, the frequent events are recalled with ease and these may paradoxically be the easiest to accomplish. The less frequent events are only identified with some difficulty and often represent extreme deviations from average expectations, illustrating the tendency to recall unique and novel experiences in memory. The analysis of the transcripts supported this view, and often topics with limited direct references represented problematic issues in design visualisation.

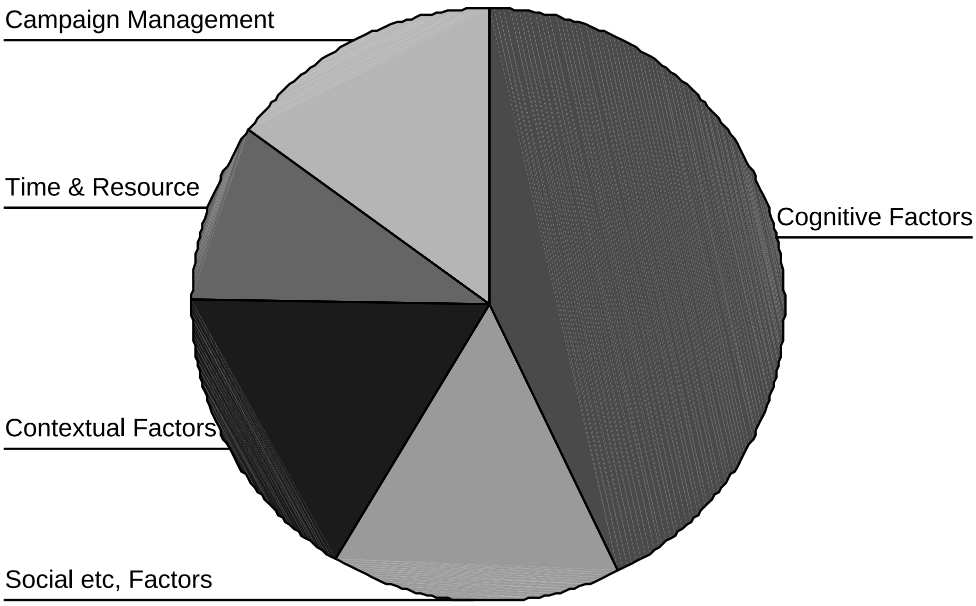


Figure 3 - Pie chart of utterances clustered into higher order categories used in Table 1. Remaining utterances associated with display technology and communication coded in campaign management factors.

The analysis of the transcripts supported the view that cognitive factors were an important determinant in the use of a system that was intended to mediate sense-making activities for an unseen operational environment. One of the most significant additions to the categorisation in this higher order category of cognitive factors was what would colloquially be described as ‘gut feeling’, and in modern parlance would be described as implicit cognition. In such a structured and methodical environment as the military, it seems strange to find references to implicit cognition but the sheer weight of information available to the command team undermines normal processes of information assimilation. In addition, the higher order patterns which could drive decision making in a routinised procedural manner are lost in noise and attempted deception by enemy forces, the so called ‘fog of war’.

The direct reference to time and contextual factors by command team members is less than surprising because of the high degree of impact those factors have on operational success. More surprising is the manner in which these are currently linked into group activities and processes that support an effective shared mental model through a series of structured activities. This has been identified as a significant weakness in the analysis of systems engineering approaches to command and control, and design (Cook, 2000). Currently it is impossible to embed the richness of the mental model inside the visualisation, and the fragile nature of the social construction of the world is subject to the vagaries of forgetting, fatigue and stress that are endemic to the battlefield management role. Future visualisation will enable richer databases of information to be visualised in synoptic form to solve problems with implementation of plans more easily.

Conclusion

In the final analysis, display and visualisation technologies must create new opportunities for sharing information across the team to enable the appreciation of command intent and to critique the current formulation of events, with the greatest body of knowledge. The large number of references to communication, directly or indirectly, in the transcripts suggests that this is the socio-cognitive medium by which the command team achieves superlative performance. Any poorly designed visualisation tool that robs the decision-making process of cognitive resources, and robs the social process of time to communicate, will potentiate the conditions for failure. The overall impression is of a socio-technical system requirement for supporting shared cognitive activities to plan and implement scripts of actions that will achieve goals in a highly dynamic high-risk environment.

This brief summary of issues identified in the transcript analysis provides the basis for designing a campaign visualisation tool by identifying important processes, structures and activities that ensure effective command team function in the perception of active command team members. This analysis has led to the ability to derive the important dimensions for visualisation design that impact on the campaign assessment and decision-making performance of command teams. From these dimensions, it has been possible to derive recommendations for visualisation design solutions, both in terms of presentation of the visualisations and use of them by the team. Full details of the visualisation recommendations, along with likely process changes to be gained from application of the appropriate technological solutions, are given in Cook, Angus and Adams (2002). These results are being used to inform the next stage in this process. The recommendations are being implemented in the design of prototype visualisations for evaluation in cognitive walkthroughs involving representative end users interacting with the prototypes to perform simulated command and control tasks. Feedback from these evaluations will be incorporated into the iterative visualisation design process before the prototypes are evaluated and compared to other representational formats in experimental trials. This ensures a fully user-centred design process where the campaign assessment needs of military command teams are considered *throughout* the design process of the visualisations in order to obtain the maximum possible performance benefit from the use of visualisations in campaign assessment.

This work was funded by the Chemical and Biological Defence and Human Sciences Domain of the MOD's Corporate Research Programme.

References

Alexander, K. B. (Brigadier General) (2000) Keynote Speech. In *RTO meeting proceedings 57 – Usability of information in battle management operations*. (pp K-1 – K-3) NATO.

Anderson, J. R. (1995) *Cognitive Psychology and its Implications* (4th Ed). New York: W. H. Freeman and Company

Angus, C.S., Adams, C.S.G., and Cook, M.J. (2002a) *Experimental presentation of military transcripts*. Unpublished report produced for QinetiQ

Angus, C.S., Adams, C.S.G., and Cook, M.J. (2002b) *Categorisation of military transcripts based on a thematic framework*. Unpublished report produced for QinetiQ

Card, S., Mackinlay, J., and Schneiderman, B. (1999*) Readings in Information Visualisation: Using Vision to Think*. San Francisco, California: Morgan Kaufmann

Chandler, M., Emery, L., and Routledge, E. (2000) *FSTA: Flight Deck Task Analysis for Air-to-air Refuelling*. Unpublished DERA report

Cohen, M. S. (1995) Three paradigms for viewing decision biases. In G. A. Klein, J. Orasanu, R. Calderwood and C. E. Zsambok (Eds) *Decision Making in Action* (2nd Edition). Norwood, New Jersey: Ablex

Cook, M.J. (2000) Neanderthal approaches to command and control : A response to cavemen in command. In C. Sundin and H. Friman (Eds.) *ROLF 2010 : The way ahead*. Stockholm: Swedish National Defence College.

Cook, M.J., Angus, C.S., and Adams, C.S.G. (2002) *Campaign Combat Information Management for Future Command Teams : Functional Requirements for Command Team Visualisation*. Unpublished report produced for QinetiQ

Cook, M.J. and Cooper, R. (2001a) *Campaign Combat Information Management for Future Command Teams: After Project Report*. Unpublished report produced for QinetiQ

Cook, M.J. and Cooper, R. (2001b) *Campaign Combat Information Management for Future Command Teams: Amended After Project Report*. Unpublished report produced for QinetiQ

Cook, M.J., Stapleton, G., and Artman, H. (2000) Battlefield Information Systems in Air Warfare. In *Usability of Information in Battle Management Operations* RTO-MP-57, AC/323 (HFM) TP/29. Nieully-sur-Seine, France : NATO-RTO (pp. 2-1, 2-12).

Diggins, S. L. C. (2000) The estimate and the emperor's new clothes *The British Army Review* (124). Spring 2000. (pp 4- 13)

Fischhoff, B, (1988) Jusdgement and decision making. In R. J. Sternberg and E. E. Smith (Eds) *The Psychology of Human Thought* (pp 155 – 178). Cambridge: Cambridge University Press

Goldman, S. R. and Petrosino, A. J. (1999) Design principles for instruction in content domains: Lessons from research on expertise and learning. In F. T. Durso (Ed), *Handbook of Applied Cognition*. (pp. 595-628). New York: John Wiley and Sons

Gordon, S. E. and Gill, R. T. (1997) Cognitive task analysis. . In C. E. Zsambok and D. E. Klein (Eds) *Naturalistic Decision Making*. (pp 131–140). Mahway, New Jersey: Lawrence Erlbaum Associates.

Henderson, S. and Pechy, A. (1999) *Shared situational awareness in PJHQ teams* Unpublished DERA Report

Hoffman R.R., Crandall B., and Shadbolt, N. (1998) Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors*, 40 (2) (pp 254-276)

Hosking, D. M. and Moreley, I. (1991) A Social Psychology of Organising: Persons, Processes and Contexts, London: Harvester Wheatsheaf

IST/TG-002 (1999) *Visualisation of massive military datasets: Human factors applications and technologies*. (Interim Report). NATO, RTO.

Klein, G. (1993) *Naturalistic decision making: Implications for design*. Crew System Ergonomics Information Analysis Centre (CSERIAC Report Number 93-01). Wright Patterson Air Force Base, Ohio

Klein, G. (1997) An overview of naturalistic decision making applications. In C. E. Zsambok and D. E. Klein (Eds) *Naturalistic Decision Making.* (pp 49-60). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Krippendorff, K. (2000) On the essential contexts of artifacts or on the
proposition that "Design is making sense (of Things)". In V. Margolin and R. Buchanan (Eds) *The idea of design*. Massachussetts, U.S.A.: MIT Press.

Lanir, Z., Fischhoff, B. and Johnson, S. (1988) Military risk taking: C3I and the cognitive functions of boldness in war. *Journal of Strategic Studies* 11 (pp 96 – 114)

Macklin, C. and Dudfield, H. (2001) Campaign assessment visualisation techniques for command teams. *Proceedings of People in control: Second international conference on human interfaces in control rooms cockpits and command centres* (481) (pp 73 – 78). IEE

Macmillan J., Getty D. J., Tatum, B. C. and Ropp, G. A. (1997) Visual Metaphors and Mental Models in Display Design: A Method for Comparing Intangibles. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting.* (pp 284-288).

Moray, N. (1996b) Model of models of….mental models. In T. B. Sheridan (Ed) *Liber Armicorum in Honour of Henk Stassen*. Cambridge Mass: MIT press

Resnick, L. B. (1989) *Knowing, Learning and Instruction: Essays in Honor of Robert Glaser*. Hillsdale, New Jersey: Lawrence Erlbaum Associates

Spiro, R.J., Vispoel, W. L., Schmitz, J. G., Samarapungavan, A., and Boerger, A. E. (1987). Knowledge acquisition for application: cognitive flexibility and transfer in complex content domains. In B. K. Britton and S. M. Glynn (Eds), *Executive Control Processes in Reading*. (pp 177-200). Hillsdale, New Jersey: Lawrence Erlbaum.

Vertzberger, Y.Y.I. (1998) *Risk taking and decisionmaking: Foreign military intervention*. Cambridge: Stanford University Press.

Model-based Principles for Human-Centred Alarm Systems from Theory and Practice

Steven T. Shorrock[1], Richard Scaife[2] and Alan Cousins[3]

[1]Det Norske Veritas (DNV) Ltd., Highbank House, Exchange Street, Stockport, Cheshire, SK3 0TE, UK. steven.shorrock@dnv.com
[2]National Air Traffic Services Ltd., Air Traffic Management Development Centre, Bournemouth Airport, Christchurch, Dorset, BH23 6DF, UK. richard.scaife@nats.co.uk
[3]National Air Traffic Services Ltd., London Area Control Centre, Swanwick, Southampton, Hampshire, SO31 7AY. alan.cousins@nats.co.uk

Abstract: The burgeoning use of 'soft-desk' alarm systems employing visual display unit (VDU) technology has resulted in various problems in alarm handling. Many control rooms are transferring from a 'hard-desk' system, with the migration of many alarms to a limited display space. One of the problems for alarm systems appears to be the lack of an 'alarm philosophy'. This paper describes efforts to develop high-level principles for the design of soft-desk alarm systems that could contribute to such a philosophy. The principles were derived via the distillation of bottom-up and top-down approaches. The bottom-up approach involved two studies regarding the design and evaluation of one bespoke system and one adapted commercial-off-the-shelf (COTS) system designed for the control and monitoring of air traffic management (ATM) software and hardware. These evaluations utilised a comprehensive database of human-machine interface (HMI) development guidelines (MacKendrick, 1998; Shorrock, et al. 2001). The guidelines that were relevant to alarm handling, and put into context by the studies, were extracted and grouped into higher-level sets to help form preliminary principles. The top-down approach involved reviewing the implications arising from a model of alarm-initiated activities (Stanton, 1994). The resultant set of human-centred principles were structured around the model, and illustrated with examples from one of the studies.

Keywords: alarm systems; principles; guidelines; air traffic management; system control.

Alarm Systems

Alarm systems represent one of the most essential and important interfaces between human operators and safety-critical processes, yet often one of the most problematic. Engineering psychology has paid considerable attention to the design of alarm systems, particularly in the process industries. This has been spurred by several major accident investigations, including Three Mile Island (1979), Milford Haven refinery (1994) and Channel Tunnel (1996). In the UK, investigations by the Department of Trade and Industry and Health and Safety Executive have found significant human factors (HF) deficiencies in alarm handling (Health and Safety Executive, 2000). Alarm flooding, poorly prioritised alarms and 'clumsy automation' have prevented operators from detecting important alarms, understanding the system state, and reacting in a directed and timely manner. Indeed, poorly designed alarm systems can hinder rather than help the operator (Swann, 1999).

Bransby and Jenkinson (1998) define an alarm system simply as "a system designed to direct the operator's attention towards significant aspects of the current plant status" (p. 7). More specifically, the Engineering Equipment and Materials Users Association (EEMUA) (1999) defines alarms as "signals which are annunciated to the operator typically by an audible sound, some form of visual indication, usually flashing, and by the presentation of a message or some other identifier. An alarm will indicate a problem requiring operator attention, and is generally initiated by a process measurement passing a defined alarm setting as it approaches an undesirable or potentially unsafe value." (p. 1). EEMUA (1999) summarise the characteristics of a good alarm as follows:

- Relevant - not spurious or of low operational value.
- Unique - not duplicating another alarm.
- Timely - not long before any response is required or too late to do anything.
- Prioritised - indicating the importance that the operator deals with the problem.
- Understandable - having a message which is clear and easy to understand.
- Diagnostic - identifying the problem that has occurred.
- Advisory - indicative of the action to be taken.

- Focusing - drawing attention to the most important issues.

The tasks of the ATM system controller have little overlap with the tasks of the air traffic controller. In fact, ATM system control (SC) has more in common with a modern UK nuclear power station (see Marshall and Baker, 1994). An integrated centralised computer system (control and monitoring system (CMS)) is used to monitor and control engineering systems within an air traffic control (ATC) centre to help engineers to maintain the ATC service. Engineers monitor alarms from dedicated workstations, and remedy faults either remotely (via software) or locally. Coupled with the increasing computerisation of ATM generally, ATM SC has adopted a 'soft-desk' environment, with VDU-based presentation of alarms.

HF Design Guidelines

HF guidelines have proven to be a popular and important method to support HF integration in alarm system development. Guidelines are an informal tool used to support designers or to aid usability evaluation of a system. Chapanis and Budurka (1990) state that guidelines can help to put HF directly in the main stream of development and make HF more directly responsible and accountable for the usability of systems.

Shorrock et al. (2001) describe the development of a database of human-machine interface (HMI) guidelines for ATM. The computerised database contains around 1,600 guidelines, integrated from over 30 established sources from a variety of industrial areas, including ATM, aviation, military, nuclear, and petro-chemical. The guidelines database is structured around seven areas: Visual Displays; Controls and Input Devices; Alarms; Interpersonal Communication; Workspace Configuration; Workplace Layout; and The Environment. An accompanying audit tool allows the user to select and store guidelines that are applicable to a particular purpose, and rate HMI components in terms of compliance and priority. Additional functionality allows the user to structure a report, by allocating guidelines, rating components, and adding comments or recommendations.

The guidelines have been applied to a variety of prototype and operational systems, including:

- Service Management Centre (SMC) - the SMC monitors and controls NATS infrastructure systems (i.e., surveillance and communications systems) across all NATS centres.
- Future Area Control Toolset (FACTS) - provides advanced conflict detection and resolution information (Evans, et al., 1999).
- Control and Monitoring Systems (CMS) - two CMS HMIs for two major air traffic control centres, focusing on alarm handling (Shorrock and Scaife, 2001, one further described below).

It is recognised that there are problems with HF guidelines. Campbell (1996) asserts that, despite increased interest in the development of design guidelines, there remains considerable uncertainty and concern regarding the utility of such information. It is paradoxical that few guidelines have been evaluated in terms of their validity, comprehensiveness, reliability, utility and usability. Shorrock et al. (2001) provide one attempt to overcome these criticisms, including an evaluation of the aforementioned guidelines.

Carter (1999) agrees that guidelines must be usable for developers before usability for end users can be improved. To get over these problems, developers need help in integrating the guidelines within their development process in a usable manner. What is needed is a set of guiding principles, preferably grounded in a human-centred model of alarm handling.

This paper describes efforts to develop high-level principles for the design of soft-desk alarm systems that could contribute to a philosophy of alarm handling. The principles were derived via the distillation of information from two studies regarding the design and evaluation of one bespoke and one adapted commercial-off-the-shelf (COTS) system designed for the control and monitoring of Air Traffic Management (ATM) software and hardware. These evaluations utilised the aforementioned HMI guidelines database (MacKendrick, 1998; Shorrock, et al. 2001). The guidelines that were relevant to alarm handling, and put into context by the studies, were extracted and grouped into higher-level sets to help form preliminary principles. The top-down approach involved reviewing the implications arising from a model of alarm-initiated activities (Stanton, 1994). The resultant set of human-centred principles was structured around the model, and illustrated with examples from one of the two studies.

Case Study - Evaluation of a Bespoke Control and Monitoring System

Introduction: One of the systems which helped in the formation of alarm handling principles is a bespoke CMS designed specifically for an en-route ATC centre in the UK (study further described in Shorrock and Scaife, 2001). The HMI comprises a number of tools and components. The main components include:

- Alarm List - displays alarms and cleared alarms for monitored systems with colour coding to denote priority, for response or re-allocation. The alarm list is the primary alerting mechanism.
- System Alarms - displays alarms for the system associated with the selected alarm (or mimic entity) with colour coding to denote priority.
- Mimics - block diagrams of monitored systems, colour coded by system state.
- Commands - allows the user to send commands to monitored workstations.
- Message and Event Logs - displays messages sent and received, and events (analysed messages) detected by the system.
- Performance Data - displays information pertaining to the performance of SC components.
- Macro - allows the user to create, edit and execute macros.
- Aide Memoire - allows the user to make and read free-text electronic notes, presented in a chronological list.
- Allocation - allows systems to be allocated to specific users.
- Responsibilities - displays systems currently allocated to the SC workstation.
- Support Information System (SIS) - displays supplementary information (e.g., telephone numbers).

The alarm list displays time-based, event-driven entries. The alarm system uses a 'ring-back' philosophy; both alarms and alarm clears must be acknowledged using buttons on the alarm list. The alarm list also provides access to system alarms (alarms for a specific system) and to mimic windows (block diagrams of monitored systems). These components are shared between two 19-inch VDUs. The alarm list is positioned on the right hand screen, and is unmoveable and unobscurable. Other components may be moved between the two screens using a mouse. A keyboard is available for entering commands and searches.

The mimic diagrams form a set of grouped and hierarchical block diagrams showing the status of monitored system components. The 'top level' diagram shows an overview of all systems. The CMS mimic displays use colours to represent system state, in contrast to other schematic displays, which use colours on the mimics to represent priority directly (e.g., Dicken, 1999). Nine colours are used to indicate generic states: OK, failed, redundancy loss, disabled, out of scan, maintenance, standby, not installed, and unknown.

'Derived entities' show summary states of lower level 'reported entities', with the 'worst' state tending to propagate upwards. A problem was identified with the way that CMS presented faults. The user's mental picture could become overloaded and degraded due to the amount of information being presented. Whilst CMS alarms were prioritised on display, further mental prioritisation was required of the user, prior to remedy. Under certain conditions, such as cascading alarms, the alarm list could become flooded. There was a danger that due to the number of alarms received, priorities could be mis-apportioned, and responses to alarms from critical systems could be delayed.

Method: The work was approached in two stages. First, CMS was audited using the ATM HMI guidelines database. This stage produced a number of observations and recommendations. Second, a multidisciplinary team reviewed the recommendations to select those that were of high-priority. These stages are described below.

Data for the guidelines audit were derived from user training materials and observations made during a series of 'walk-throughs' and over 20 hours of simulation and operation. These activities helped to generate a log of issues of concern. These issues were checked against the database of HMI guidelines. Observations were coded according to how the system complied with the guidelines (non-compliance; partial compliance; full compliance).

During this audit, recommendations were generated for the possible improvement of the system. Since it was not possible to implement every recommendation due to financial and temporal constraints on the CMS project, it was necessary to prioritise the recommendations for further consideration. This was achieved via a group review of the HF recommendations - a form of 'cost-benefit' analysis. The team represented the following: operational engineering (5 participants); engineering day-team (non-operational) (4); management (2); engineering safety (1); and human factors (1). Two sessions were held, each lasting approximately three hours. Seven to eight participants attended each session. The groups rated each recommendation on a number of criteria, with each rating receiving an agreed score, as follows:

Safety significance	Yes	No		
Potential benefit	High = 3	Med. = 2	Low = 1	None = 0
Feasibility	High = 3	Med. = 2	Low = 1	None = 0

The scores for each recommendation were multiplied to give a priority score, with a minimum of zero and a maximum of nine, thus:

Priority score	Potential benefit	$\times$	Feasibility

The 'potential benefit' ratings took account of any trade-offs and knock-on impacts. The 'feasibility of implementation' ratings were made based on ease of implementation *within* the following categories: HMI change, procedures, training or further investigation. This ensured that the scale remained sensitive, and that those recommendations that were relatively difficult to implement (mostly HMI changes) did not automatically receive a low score. 'Safety-related' and 'non-safety-related' recommendations were dealt with separately. The following sections detail the HF review of the CMS HMI, followed by the group recommendations review.

Results: The results of the HMI audit identified 65 cases in which the design was not compliant with the guidelines and 18 cases of partial compliance. The partial- and non-compliances provided indications of a number of issues that would need to be addressed in some way in order to improve the usability of the system. The main issues fell into the categories shown in Table 1.

Table 1 - Distribution of recommendations over human factors areas and modes of implementation

HF area	*Example Issues*	*# Rec's*
Text format & display	Truncation of text, Number of alarms presented, Meaningfulness of abbreviations, Alarm detection	19
Text input & search	Lack of 'cut' and 'paste' functions, Keyboard shortcuts and alternatives	15
Button selection & feedback	Lack of selection feedback	14
General usage	HMI layout	10
Colour & prioritisation	Prioritisation strategy, Consistency of colour coding, Use of saturated colours, Number of colours	4
Text highlighting	Contrast between text and background	4
Symbology & icons	Symbol size	3
Window handling & title bars	Window overlap	3
Masking & suppression	Suspension of alarms when systems out of scan Clearing alarms when fault still exists	2
Blink coding	Overuse of blink coding	1
Response times	Feedback of system status (e.g., 'busy', 'searching')	1

In total, 76 recommendations were made; 65 concerned HMI changes, six concerned procedures and training, and five concerned further investigation. Table 1 also shows the distribution of recommendations over HF areas. The majority of recommendations pertained to alarm lists, mimic diagrams, and message and event logs. Fourteen recommendations (18% of the total number) were classed as 'safety-related'. Thirteen recommendations had a priority score of nine, and 10 recommendations had a priority score of six. The remaining recommendations were either low priority, or a priority could not be calculated at the time due to unknown scores.

The review selected 13 high priority recommendations (eight safety-related, five non-safety-related) for action. Two further recommendations had unknown potential benefit, and five had unknown feasibility, and were put forward for further investigation. The review led to a series of change requests being raised by SC staff, to address the selected recommendations, and further work to address recommendations concerning procedures, training, and further investigation.

The Distillation of Principles from Guidelines

The study described in this paper, and another concerning a COTS system for another ATC centre, made it apparent that a set of alarm handling principles would greatly assist designers of 'soft-desk' alarm systems. At present, relatively few companies adhere to a 'design philosophy' for alarm systems, and hence the design process can become protracted, error-prone, and costly.

This stage of work used the studies to put the guidelines into context and draw upon insights from the real-world application of such guidance. Collectively, the two studies resulted in approximately 100 recommendations on the design of alarm handling systems. This meant that there was a great deal of material from which real examples of the application of guidelines could be drawn. The fact that HF specialists operated as part of interdisciplinary teams in both cases meant that they were also a rich source of information on the pragmatic application of HF guidance in major projects.

This final list, embellished with the outputs from the two studies, was then analysed independently by two of the authors to generalise the list of guidelines and recommendations into a smaller set of design principles. The results of both researchers' deliberations were compared and any disagreements were resolved through discussion.

The final list of principles contained 43 items covering HMI design, including issues such as:

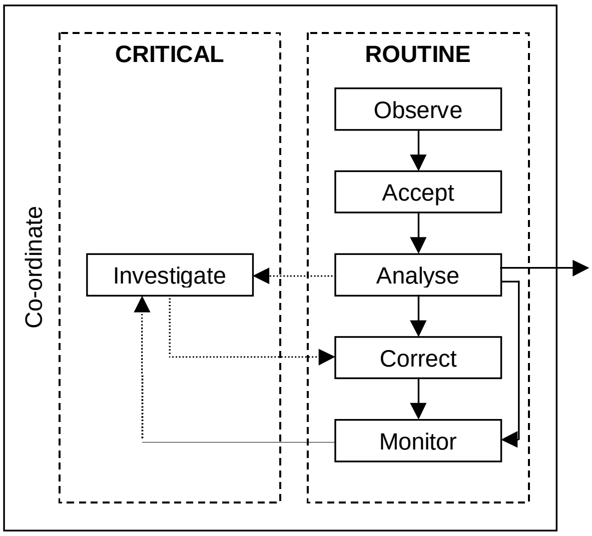
- visibility, appearance and distinctiveness
- meaningfulness and contextual relevance
- priority and status
- acknowledgement and acceptance
- multiple operator protocols.
- order of presentation and timeliness of detection
- information quantity and filtering
- grouping
- interaction and navigation, and

The last stage in this process was to use a model of alarm management to structure the principles in a meaningful way. Stanton's (1994) model of alarm-initiated activities was used for this purpose. Initially this was used to brainstorm high-level principles from a top down perspective. This was complemented by the bottom-up generation of principles from guidelines and case studies. The following section describes Stanton's model, along with the human-centred alarm handling principles elicited.

Model-based Alarm Handling Principles

Stanton's (1994) 'model of alarm initiated activities' was used to help formulate and organise principles to support the alarm handling process for future projects. Stanton's model comprises six activities (Observe, Accept, Analyse, Investigate, Correct, and Monitor), and is illustrated in Figure 1.

Figure 1- Model of alarm initiated activities (adapted from Stanton, 1994).



Observation is the detection of an abnormal condition within the system (i.e., a raised alarm). At this stage, care must be taken to ensure that colour and flash/blink coding, in particular, supports alarm monitoring and searching. Excessive use of highly saturated colours and blinking can de-sensitise the operator and reduce the attention-getting value of alarms. Auditory alarms should further support observation without causing frustration due to the need to accept alarms in order to silence the auditory alert, which can change the 'alarm handling' task to an 'alarm silencing' task.

In the case study described, urgent alarms were presented in red, but non-urgent alarms in brown rendering the urgent (red) alarms less distinctive. It was recommended that an alternative colour for non-urgent alarms should be used to increase discriminability of urgent alarms.

Acceptance is the act of acknowledging the receipt and awareness of an alarm. At this stage, operator acceptance should be reflected in other elements of the system providing alarm information. Alarm systems should aim to reduce operator workload to manageable levels - excessive demands for acknowledgement increase workload and operator error. 'Ring-back' design philosophies double the acceptance workload; careful consideration is required to determine whether cleared alarms really need to be acknowledged. The issue of single versus multiple alarm acknowledgement does not yet seem to have been resolved (Stanton, 1994). Group acknowledgement may lead to unrelated alarms being masked in a block of related alarms. Single acknowledgement will increase workload and frustration, and may begin to resemble group acknowledgement as task demands increase. Two alternatives may be suggested. Operators may select individual alarms with the mouse with a 'Control' key depressed. This would form a half-way house between individual and block acknowledgement, countering the effects of a block selection using 'click-and-drag' or a 'Shift' key, whilst reducing workload. A still more preferable solution may be to allow acknowledgement for alarms for the same system.

With the CMS system described, it was possible for operators to clear an alarm message manually when the fault still existed. This was possible for individual alarms or for a whole system, in which case the mimic remained the same status but the alarm entries were deleted. This was not recommended as it increased the risk of error (i.e., the operator could forget that the fault still existed). It was recommended that the user should not be able to clear an alarm if the fault still existed.

Analysis is the assessment of the alarm within the task and system contexts, which leads to the prioritisation of that alarm. Whilst many (often justified) criticisms have been levelled against alarm lists, if properly designed, alarm lists can support the operator's preference for serial fault management (Moray and Rotenburg, 1989). Effective prioritisation of alarm list entries can help engineers at this stage. Single 'all alarm' lists can make it difficult to handle alarms by shifting the processing debt to the engineer. However, a limited number of separate alarm lists (e.g., by system, priority, acknowledgement, etc.) could help operators to decide whether to ignore, monitor, correct or investigate the alarm.

In the CMS system, insufficient space was given to the fault description, while too much space was given to the system name. This resulted in a truncated fault description, making it difficult to understand the nature of the fault. It was recommended that more space for the fault description be provided by increasing the width of the alarm list and/or abbreviating system names (e.g., Workstation = WKS).

Investigation is any activity that aims to discover the underlying causes of alarms in order to deal with the fault. At this stage, system schematics can be of considerable use in determining the underlying cause. Coding techniques (e.g., group, colour, shape) need to be considered fully to ensure that they support this stage without detracting from their usefulness elsewhere. Furthermore, displays of system performance need to be designed carefully in terms of information presentation, ease of update, etc.

Using the CMS system, users were able to suppress or block updates from a reported system state from propagating up through higher order mimics. Whilst this had a purpose for the user, it meant that the user could forget that the updates from an entity were blocked because there would be no further visual indication that the entity was not reporting. This could have a serious detrimental effect on the investigation of underlying causes of a fault. It was recommended that a unique, black symbol be designed and propagated through higher-level entities to remind the user of the block (e.g., '✖').

Correction is the application of the results of the previous stages to address the problem(s) identified by the alarm(s). At this stage, the HMI must afford timely and error-tolerant command entry, if the fault can be fixed remotely. For instance, command windows should be easily called-up, operator memory demands for commands should be minimised, help or instructions should be clear, upper and lower case characters should be treated equivalently, and positive feedback should be presented to show command acceptance.

Users were required to input text into the CMS HMI after taking control of a system to complete certain tasks. One such task included entering processor information (e.g., serial number), which required the user to access a separate 'performance data' screen. This could be a lengthy process as the user needed to open a new window indirectly and copy the text manually, which could be more than 9 characters. Ideally, such information should be available directly from the entity block (e.g., through an associated menu), with no typing required. In the absence of this feature, it was recommended that a cut and paste function be added to reduce user workload and input errors during fault correction.

Monitoring is the assessment of the outcome of the Correction stage. At this stage, the HMI (including schematics, alarm clears, performance data and message/event logs) needs to be designed to reduce memory demand and the possibility for errors of interpretation (e.g., the 'confirmation bias').

On the CMS HMI, all alarms were presented in the same list. During periods of alarm flooding (high numbers of alarms, e.g., cascading alarms due to a common mode failure), this made it difficult for users to identify urgent alarms and confirm when faults had been cleared. It was recommended that cleared alarms be presented in a separate bay with the ability to filter alarms based on priority (i.e., temporarily to show only high priority alarms).

The case study reported in this paper represents a multiple user system in which a team of operators could work collaboratively to attend to system faults. Such collaboration could take the form of delegating authority for specific parts of the system to work colleagues, or the co-ordination of effort for faults that permeate several different parts of the overall system. In order to account for the increasing prevalence of multiple user systems in control room operations within safety related industries, it was felt that an additional module needed to be added to Stanton's model of alarm handling. Hence, *'co-ordination'* is the transfer of information between operators and the application of collaborative efforts to observe, accept, analyse, investigate or correct faults. This is not a final stage of the model, but rather an additional activity that permeates all stages.

In the study described, team members within the control room used a large overview display at the front of the control room, to maintain a group awareness of the status of the overall system, and the progress of other team members towards the correction of any faults. Recommendations were put forward regarding the content of this display, to ensure that 'group situation awareness' was maintained at the optimum level.

The principles of alarm handling derived from the analysis of the model, synthesis of guidelines, and consideration of the studies were examined in relation to Stanton's model to identify the relevant stages of the alarm handling process. It was found that in most cases, the principles were applicable primarily in one stage of the process, but would also have a bearing on other stages, depending on the system in question.

The principles were therefore classified in terms of their primary relevance within the model. Table 2 shows the resulting principles sub-divided by model area.

Conclusion

This paper has argued that alarm system design activities need to be guided by user-centred principles of alarm handling, preferably structured around a valid model of alarm handling. Such a set of principles has been proposed based on several strands of work, and the distillation of bottom-up and top-down approaches.

The paper has identified a number of considerations that should be borne in mind, and has demonstrated the usefulness of a model of ‘alarm initiated activities’ (Stanton, 1994) in the design and evaluation of alarm systems. Whilst HMI guidelines have helped considerably to define solutions for the study described in this paper, it is believed that consideration of the proposed principles and formulation of alarm philosophies will assist further in the design and development of usable and safe alarm systems.

Acknowledgements

The authors wish to thank the NATS System Control engineers and others involved in the study for their enthusiastic involvement and support. In particular the NATS System Control Design Group Design Authority; the Operational and Engineering Authority who conducted the trials and provided the data. The views expressed in this paper are those of the authors and do not necessarily represent those of the affiliated organisations.

Table 2 - Model-based principles for human-centred alarm systems.

Observe	
1	Alarms should be presented (time stamped) in chronological order, and recorded in a log in the same order.
2	Alarms should signal the need for action.
3	Alarms should be relevant and worthy of attention in all the plant states and operating conditions.
4	Alarms should be detected rapidly in all operating conditions.
5	It should be possible to distinguish alarms immediately, i.e., different alarms, different operators, alarm priority.
6	The rate at which alarm lists are populated must not exceed the users' information processing capabilities.
7	Auditory alarms should contain enough information for observation and initial analysis and no more.
8	Alarms should not annoy, startle or distract unnecessarily.
9	An indication of the alarm should remain until the operator is aware of the condition.
10	The user should have control over automatically updated information so that information important to them at any specific time does not disappear from view.
11	It should be possible to switch off an auditory alarm independent of acceptance, but it should repeat after a reasonable period if the fault is not fixed.
12	Failure of an element of the alarm system should be made obvious to the operator.
Accept	
13	Reduce the number of alarms that require acceptance as far as is practicable.
14	Allow multiple selection of alarm entries in alarm lists.
15	It should be possible to view the first unaccepted alarm with a minimum of action.
16	In multi-user systems, only one user should be able to accept and/or clear alarms displayed at multiple workstations.
17	It should only be possible to accept the alarm from where the sufficient alarm information is available.
18	It should be possible to accept alarms with a minimum of action (e.g., double click), from the alarm list or mimic.
19	Alarm acceptance should be reflected by a change on the visual display, such as a visual marker and the cancellation of attention-getting mechanisms, which prevails until the system state changes.
Analyse	
20	Alarm presentation, including conspicuity, should reflect alarm priority, with respect to the severity of consequences associated with delay in recognising the deviant condition.
21	When the number of alarms is large, provide a means to filter the alarm list display by sub-system or by priority.
22	Operators should be able to suppress or shelve certain alarms according to system mode and state, and see which alarms have been suppressed or shelved, with facilities to document the reason for suppression or shelving.
23	It should not be possible for operators to change priorities of any alarms.
24	Automatic signal over-riding should always ensure that the highest priority signal over-rides.
25	The coding strategy should be the same for all display elements.
26	Facilities should be provided to allow operators to recall the position of a particular alarm (e.g., periodic divider lines).
27	Alarm information such as terms, abbreviations and message structure should be familiar to operators and consistent when applied to alarm lists, mimics and message/event logs.
28	The number of coding techniques should be at the required minimum, but dual (redundant) coding may be necessary to indicate alarm status and improve accurate analysis (e.g., symbols and colours).

29	Alarm information should be positioned so as to be easily read from the normal operating position.
Investigate	
30	Alarm point information (e.g., settings, equipment reference) should be available with a minimum of action.
31	Information on the likely cause of an alarm should be available.
32	A detailed graphical display pertaining to a displayed alarm should be available with a single action.
33	When multiple display elements are used, no individual element should be completely obscured by another.
34	Visual mimics should be spatially and logically arranged to reflect functional or naturally occurring relationships.
35	Navigation between screens should be quick and easy, requiring a minimum of user action.
Correct	
36	Every alarm should have a defined response and provide guidance or indication of what response is required.
37	If two alarms for the same system have the same response, then consideration should be given to grouping them.
38	It should be possible to view status information during fault correction.
39	Use cautions for operations that might have detrimental effects.
40	Alarm clearance should be indicated on the visual display, both for accepted and unaccepted alarms.
41	Local controls should be positioned within reach of the normal operating position.
Monitor	
42	No primary principles. However, a number of principles primarily associated with observation become relevant to monitoring.
Co-ordinate	
43	Provide high-level overview displays to show location of operators in system, areas of responsibility, etc.

References

Bransby, M.L. and Jenkinson, J. (1998). The Management of Alarm Systems. HSE Contract Research Report CRR 166/1998. HSE Books: Norwich.

Campbell, J.L. (1996). The development of human factors design guidelines. International Journal of Industrial Ergonomics, 18 (5-6), 363-371.

Carter, J. (1999). Incorporating standards and guidelines in an approach that balances usability concerns for developers and end users, Interacting with Computers, 12 (2), 179-206.

Chapanis, A. and Budurka, W.J. (1990). Specifying human-computer interface requirements. Behaviour & Information Technology, 9 (6), 479-492.

Dicken, C.R. (1999). Soft control desks - producing an effective alarm system. In IEE, People in Control: An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres. 21-23 June 1999, pp. 212-216.

EEMUA (1999). Alarm Systems: A Guide to Design, Management and Procurement. EEMUA Publication No. 191. The Engineering Equipment and Materials Users Association: London.

Evans. A, Slamen A.M. and Shorrock S.T. (1999). Use of Human Factors Guidelines and Human Error Identification in the Design Lifecycle of NATS Future Systems. Paper presented to the Eurocontrol/Federal Aviation Administration Interchange Meeting, France, 27-29 April 1999.

Health and Safety Executive (2000). Better Alarm Handling. HSE Information Sheet - Chemicals Sheet No. 6. March, 2000.

MacKendrick, H. (1998). Development of a Human Machine Interface (HMI) Guidelines Database for Air Traffic Control Centres. R & D Report 9822. National Air Traffic Services Ltd.: London.

Marshall, E. and Baker, S. (1994). Alarms in nuclear power plant control rooms: current approaches and future design. In N. Stanton (Ed.), Human Factors in Alarm Design. Taylor and Francis: London, pp. 183-192.

Moray, N. and Rotenburg, I. (1989). Fault management in process control: eye movements and action. Ergonomics, 32 (11), 1319-1342.

Shorrock, S.T., MacKendrick, H. and Hook, M., Cummings, C. and Lamoureux, T. (2001). The development of human factors guidelines with automation support. Proceedings of People in Control: An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres, UMIST, Manchester, UK: 18 - 21 June 2001.

Shorrock, S.T. and Scaife, R. (2001). Evaluation of an alarm management system for an ATC Centre. In D. Harris (Ed.) Engineering Psychology and Cognitive Ergonomics: Volume Five - Aerospace and Transportation Systems. Aldershot: Ashgate, UK.

Stanton, N. (1994). Alarm initiated activities. In N. Stanton (Ed.), Human Factors in Alarm Design. Taylor and Francis: London, pp. 93-118.

Swann, C.D. (1999). Ergonomics of the design of process plant control rooms. In IEE, People in Control: An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres, 21-23 June 1999.

.

Toward a Decision Making Support of Barrier Removal

Zhicheng Zhang, Philippe Polet, Frédéric Vanderhaegen,

Laboratoire d'Automatique, de Mécanique et d'Informatique industrielles et Humaines (LAMIH)
University of Valenciennes - UMR 8530 CNRS
Le Mont Houy - 59313 Valenciennes Cedex 9 - France.
{zzhang, ppolet, vanderhaegen}@univ-valenciennes.fr.

Abstract: This paper presents a model to study human reliability in decision situations related to operational Barrier Removal (BR). BR is a safety related violation, which usually results from a decision making. A comparative modelling approach for decision reliability studies on BR is firstly proposed in this paper. The comparative analysis between the data prior to the barrier removal and the ones after their removal allows us to understand the mechanism better when human operator removes a barrier, and to predict finally the consequences better on human actions for a new barrier. The further development of this model can provide designers as tools to support the anticipation of removal for a barrier. This tool may be used to analyse both the qualitative (subjective) and the quantitative (objective) data. Finally, a case study with the data in terms of both BR indicators and indicators on performance criteria on the railway experimental simulator is presented.

Keywords: HRA, Barrier Removal, Violation, Simulator studies, Decision making support.

Introduction

Decision making has been widely studied by many scientific disciplines, e.g. Svenson (1998) and Dougherty (1998) discussed the role of decision making in PSA related cases, especially for the Error Of Commission (EOC). However, human operators' decision making is not usually explicitly included in human reliability analysis (HRA) (Holmberg J, Hukki J, Norros L, et al, 1999; Pyy, P., 2000; Basra, G., Dang V., Fabjan L. et al, 1998; De Gelder, P., Zhang, Z., Reiman, L., 1998).

Decision making on a kind of particular violation, Barrier Removal (BR), has been aware recently in HRA field. Usually, designers of a complex human-machine system (HMS) specify barriers to protect the system from the negative consequences of errors or failures (Kecklund L. J., Edland A., Wedin P. et al, 1996). Nevertheless, Hollnagel (1999) also allocated a prevention function to barriers and defines them as obstacles, obstructions, or hindrances in order to decrease risk, in term of prevention (occurrence of hazardous event) and protection (severity of hazardous event).

Sometimes, users of such specified HMS voluntarily do not respect the prescriptions of designers. This phenomenon can be observed in various process industry fields (Vanderhaegen, F., Polet, P., 2000; Polet, P., Vanderhaegen, F., Amalberti, R., 2001; Zhang Z., Polet P., Vanderhaegen V. et al, 2002). The BR discussed in this paper concerns the particular violations which are made without any intention to damage subjectively the HMS. When a barrier is removed, the users are facing a risk from which they were protected, but obtain an immediate benefit by compromising between performance criteria. As a matter of fact, these operational situations and real conditions of use differ from the prescribed ones (Vanderhaegen, F., Polet, P., Zhang, Z. et al, 2002). A development of a risk analysis method to analyse such migrations is then important to introduce the notion of violations and their associated risks at the design phase as early as possible.

The aim of this paper is to present a comparative decision making model and apply it to decision situation of BR. Representing barriers as a constraint network and simulating relaxation of constraints related to barrier removal can provide designers as tools to support the anticipation of barrier removal. It may be used to analyse both the qualitative (subjective) and the quantitative (objective) data. Data from a series of railway simulator experiments are used in the study. The paper ends up with conclusions and some perspectives about the study.

Comparative model to analyse human reliability in BR

Based on the identification of two barrier groups[21] according to the physical presence or absence of four classes of barriers (Polet P., Vanderhaegen F., Wieringa P.A., 2002), a model of the removal of a given barrier integrated with three distinct attributes has been developed by considering both the positive and the negative consequences of such a human behaviour (Polet P., Vanderhaegen F., Millot P. et al, 2001).

The consequences in the model of BR consist of the **direct** consequences and the **potential** ones. Firstly, the **direct** consequences include:

The immediate cost of removal: in order to cross a barrier the human operator sometimes has to modify the material structure, and/or the operational mode of use. That usually leads to an increase in workload and may also have negative consequences on productivity or quality. These negative consequences are immediate and acceptable by the users.

The expected benefit: a barrier removal is goal driven. Removing a barrier is immediately beneficial and the benefits outweigh the costs, i.e. they are high enough to accept to support the cost.

Secondly, a barrier that is removed introduces a **potentially** dangerous situation, i.e. there may be some potential consequences. So, the removal of a barrier has also a possible deficit considered by users as a latent and unacceptable negative consequence.

Usually, the decision on the removal of a barrier is made among three indicators (benefit, cost and possible deficit), a compromise or a combination of these three attributes determines the decision of either removing or respecting the barrier.

In order to analyse the technical or operational reasons for removing barriers, and propose technical solutions to reduce the risks associated to barrier removal, an original model of barrier removal has been developed (Vanderhaegen, F., Polet, P., Zhang, Z. et al, 2002). It is both a descriptive and an explanatory model that presents a barrier removal as a combination of direct and potential consequences: the immediate benefit obtained after the removal, the immediate cost when the barrier is removed and the possible deficit due to the absence of the barrier. Both a prospective model (compared with designer's viewpoint) and a retrospective one (compared with the user's viewpoint) of the removal of prescribed barriers have been discussed. Preliminary results on comparison between the point of view of one of the platform's designers and the average point of view of 20 human operators have been provided.

During the BR analysis, for each barrier class, the levels of all three indicators (benefit, cost and possible deficit) may be provided in terms of, e.g. four performance criteria: productivity, quality, safety and workload (there are also another criteria, e.g. individual advantage criteria such as motivation, free time, etc.). The identification of the function giving the BR probability regarding benefit, cost and potential deficit is not evident. Normally, for the given barriers (e.g. during the simulator experiment), removal of a barrier can be observed, which means that all these barriers can be divided into two groups — the removed barriers and the respected/non removed barriers. However, it isn't very easy to know which barriers are similar ones among all the barriers; Finally, when the (re)design of a new barrier needs to be implemented, it's better to predict, first of all, its final removal probability, then retrospectively, to integrate with the user's viewpoint during the early phase of the (re)design.

To solve this kind of problem, a method integrating the Self-Organizing Maps (SOM) **(**Kohonen T., 1998; Kohonen T., 2001; Rousset P., 1999) algorithm has been proposed (Zhang Z., Polet P., Vanderhaegen V. et al, 2002). As an artificial neural network model, the SOM is designed for multidimensional data reduction with topology-preserving properties. It has been applied extensively within fields ranging from engineering sciences to medicine, biology and economics (Kaski, S., Kangas, J., and Kohonen, T., 1998). In various engineering applications, entire fields of industry can be investigated using SOM based methods (Simula O., Vasara P., Vesanto J. et al, 1999). The SOM algorithm is based on the unsupervised learning principle (Kohonen T., 2001; Simula O., Vasara P., Vesanto J. et al, 1999). The SOM can be used for, e.g. clustering of data without knowing the class memberships of the input data, which is a clear advantage when compared with the artificial neural network methods based on supervised learning which require that the target values corresponding to the data vectors are known (Simula O., Vasara P., Vesanto J. et al, 1999).

The method of integrating the SOM algorithm has been used to analyse the BR data in terms of benefit, cost and possible deficit. The experimental data have been analysed in term of mono-performance mode and multi-performance mode, predictions (removal result for each time is a removal status, i.e. either removed or non removed) for additional human operators have been implemented (Zhang Z., Polet P., Vanderhaegen V. et al, 2002).

[21] Four barrier classes: material, functional, symbolic and immaterial barriers (Hollnagel E.,1999).

Moreover, it was found that the direct application of SOM algorithm to the BR study didn't give the satisfied prediction results. In order to improve the prediction accuracy of the removal of a barrier, an improved Hierarchical SOM (HSOM) method has therefore been proposed (Zhang Z., Vanderhaegen F., 2002). By comparing between the unsupervised SOM, the supervised SOM and the HSOM algorithms for same BR data set, the proposed HSOM provides highest prediction rate. Based on the SOM map obtained from the training, predictions can be made prospectively for a changed barrier, even if some barrier indicator data for the barrier are incomplete or missing.

Anyway, the analysis of barrier removal has been implemented only for the human behaviour data after the barriers were removed so far. In order to understand the mechanism better when human operator removes a barrier, and finally predict the consequences better on human actions for a new barrier, it is necessary to study human reliability in decision making situation by comparing between the operator behaviour data before the removal of barrier and the ones after the removal.

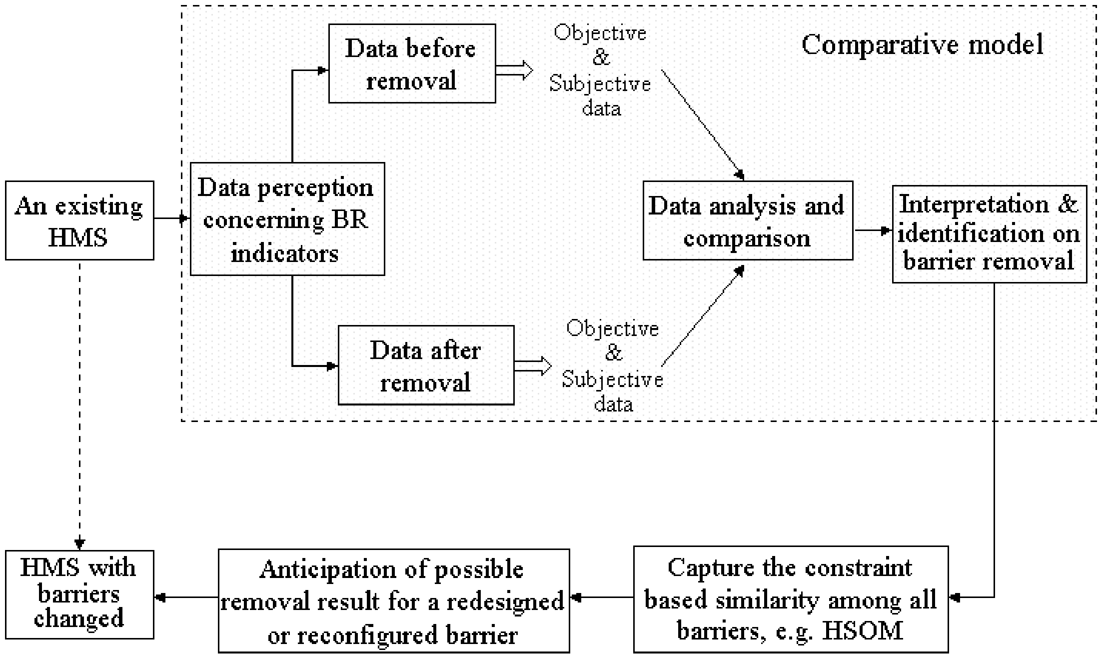


Figure 1 - Comparative model and its application to the anticipation of the possible removal consequence of a new barrier

A comparative model can be used to study the human operator behaviour data before and after the removal of a barrier for a given system, Figure 1:

- First, the data set without the removal of barrier for a given HMS include subjective data (qualitative judgment by the human operator in terms of BR indicators and assessment by himself/herself according to different performance criteria, e.g. productivity, quality, safety and workload), and objective measurement data in terms of performance criteria.
- Second, the data set after the barrier removal include the same type of data as the one before the removal. The unique difference between two types of data is, in the first one, people are required to operate the system by following the prescribed operational procedures/technical specifications without removing any prescribed barrier, even if he/she judges that some barriers may be removed.

By comparing and analysing two behaviour data set, on the one hand, the efficiency of barrier can be verified so as to accept or prohibit its removal; And on the other hand, the consistency between two data set may be used to analyse the objectivity of human operators' decision.

Based on the objective & subjective data on the barrier removal and the identified & verified barrier removal probability, the assessment of a constraint based similarity may be implemented, for instance, the Hierarchical Self-Organizing Maps (HSOM) (Zhang Z., Vanderhaegen F., 2002) may be applied. Then, if a barrier needs to be (re)designed or (re)configured, the anticipation of the possible consequences of its removal can be performed.

Several railway simulation experiments have been implemented at the University of Valenciennes (France), in collaboration with Delft University of Technology (Netherlands), to study the feasibility of BR analysis. In order to verify the comparative model, the study of data collected from 20 operators according to different criteria has been implemented too.

Case study - Application in railway experimentation

The case study is based on an experimental platform that simulates train movements from some depots to another ones crossing several exchange stations on which human operators convert products placed on the stopped train. In the process of the train traffic flow controlled by a human operator, several risks have been identified, e.g. derailment of trains: when a train is authorized to move, it may derail if the corresponding switching device is not operated correctly; Collision between trains such as a face-to-face or an overtaking collision, etc.

In order to limit relative risks and control the traffic flow, several barriers have been proposed. They are composed by the immaterial barriers and the materiel barriers. Immaterial barriers such as procedures that constrain the human controllers' behaviour: for example - to respect the timing knowing that it is better to be in advance.

Besides the immaterial barriers, there are material barriers such as signals with which human controllers have to interact:

- Signals to prevent traffic problems related to the inputs and the outputs of trains from depots.
- Signals to prevent traffic problem at the shunting device.
- Signals to control inputs and outputs at the transformation areas.
- Signals to inform the treatment of the content of a train is in course.

The experiment in which the proposed model was applied consists of two steps:

- First step of the experiment with all the designed barriers active, the data set without the removal of barrier include subjective data (qualitative judgment by the human operator in terms of benefit, cost and possible deficit and assessment by himself/herself according to different performance criteria, e.g. productivity, quality, safety and workload), and objective measurement data in terms of performance criteria.
- Second step of the experiment with only barriers that are selected by the human operator who controls the traffic, that means he/she may remove several barriers which were being judged as removable. The data set after the removal of barrier include the same data type as the data set in the first step.

After the first step, human operators have experienced all the prescribed barriers while after the second step where they could select the signals that they wanted to remove or maintain, they experienced the consequences of their decisions while performing scenarios with similar control difficulty. After both experiments, they answer a questionnaire on the evaluation of the interest of the removal of family of barriers in terms of benefit, cost and potential deficit. They have to take into account four performance criteria as follows:

- The quality related to the advancement of the planning.
- The production related to the percentage of product treated at the stations.
- The traffic safety in terms of collision, derailment and possible accident due to an incorrect synchronization of movement announcement message at transformation stations.
- The human workload related to the occupational rate.

By comparing the data before the barrier removal with the ones after their removal, i.e. the statistic comparison between two experiment data set, the corresponding results are shown in Table 1. In the table, "low" means the difference between the respective data is low, and "low (+)" indicates the average indicator value after the removal is increased compared with the average indicator value before the removal. They relate subjective data with 5 qualitative evaluation levels: very low, low, medium, high and very high. And each indicator which was used to be compared between each other and each element data on the perception of performance is the qualitative average of 20 human controllers' point of view.

Barrier family	Criteria	Difference on B	Difference on C	Difference on PD	Difference on PP	
Depot signals	Quality	0Low (+)	1Low (-)	2Low (-)	Quality	Low (+)
	Production	3Low (+)	4Low (-)	5Low (-)		
	Safety	6Low (+)	7Low (-)	8Low (+)		
	Workload	9Low (+)	10Low (-)	11Low (-)		
Shunting signals	Quality	12Very Low	13Low (+)	14Low (+)	Production	Low (+)
	Production	15Low (-)	16Very Low	17Low (-)		
	Safety	18Very Low	19Low (+)	20Low (+)		
	Workload	21Medium (+)	Low (+)	Medium (-)		
Flow signals at transformation area	Quality	22Low (+)	23Low (-)	24Low (-)	Safety	Low (+)
	Production	25Medium (+)	26Low (-)	27Very Low		
	Safety	28Medium (+)	29Low (-)	30Low (+)		
	Workload	31Low (+)	32Very Low	33Low (-)		
Stop signals at transformation areas	Quality	34Low (+)	35Low (-)	36Low (-)	Workload	Very Low
	Production	37Low (-)	38Low (-)	39Low (+)		
	Safety	40Medium (+)	41Very Low	42Low (+)		
	Workload	43Low (-)	44Low (-)	45Low (-)		

Table 1 - Differences between the data before the barrier removal and the ones after their removal (B=Benefit, C=Cost, PD=Possible Deficit, PP=Performance Perception of Barrier Removal)

Preliminary analysis of the data in the Table 1 tells us that there are some differences on the subjective judgments in terms of three BR attributes between the data before the removal (pre-removal) and the ones after removal (post-removal). Compared with the subjective judgment data before the removal, it is very likely that people often underestimate on the benefit of removal of a barrier (column of "benefit") as well as on the performance perception whereas they often overestimate on the cost (column of "cost"). It is very helpful feedback for human judgment when (s)he removes a barrier, and it may then be taken into account for the anticipation of a new barrier removal.

46Barrier family	47Average of removal probability
Depot signals	69.2 %
Shunting signals	31.3 %
Flow signals at transformation area	18.9 %
Stop signals at transformation area	0.6 %

Table 2 - Average of removal probability for different barrier families

By performing the statistic analysis of observed removal results (from the data set of second step experiment) for the four material barrier families and for the 20 operators, the removal probability for each barrier family can be calculated out (see Table 2).
Considering above-mentioned removal results as output data, and considering the corresponding indicator network data as input data, the assessment of a constraint based similarity between the objective & subjective data on the barrier removal and the identified & verified barrier removal probability may then be implemented, for instance, the HSOM method can be applied. Then, if a barrier needs to be changed or a new barrier needs to be designed, the anticipation of the possible results of its removal with the method can

be used to support his/her decision making on the removal of barrier through inputting the indicator data about the removal of this barrier in terms of different criteria.
In order to verify the feasibility of the application of the proposed model, Table 3 illustrates an example of anticipation results of the removal probability for a new barrier and actual observed results. Depot signals, Shunting signals and Stop signals at transformation area are considered as the existing barriers, and the 4th barrier - Flow signals at transformation area is supposed as a barrier which needs to be redesigned or a new barrier for this system. Along the experiment schedule, the 20 operators' data for previous three barriers and respective removal results were gathered to implement the assessment of constraint based similarity between the subjective data on the barrier removal and the identified & verified barrier removal probability. Once the assessment is completed, i.e. in this application, the method integrating SOM algorithm has been used to find the relation function between the input data set (BR indicators) and the output ones (corresponding removal results).

Serial no. of operator for a new barrier (Flow signals at transformation area)	Anticipation the removal probability of a new barrier with the method integrating SOM		Observation results	
Operator 1	*0Removal*	Perception, interpretation and identification of removal probability, then training the SOM with the previous 3 traffic signals data, finally, anticipation one by one of the removal results for the 4th barrier	Non removal	Anticipation results may be validated with the observed one, and the corresponding anticipation error is *10/20 – 8/20 =* ***10%***
Operator 2	Non removal		Non removal	
Operator 3	Removal		Non removal	
Operator 4	*1Non removal*		Non removal	
Operator 5	Removal		Removal	
Operator 6	*2Non removal*		Removal	
Operator 7	*3Removal*		Removal	
Operator 8	Removal		Non removal	
Operator 9	Non removal		Non removal	
Operator 10	Removal		Removal	
Operator 11	Removal		Non removal	
Operator 12	*4Removal*		Non removal	
Operator 13	Removal		Removal	
Operator 14	Non removal		Non removal	
Operator 15	Non removal		Non removal	
Operator 16	Non removal		Non removal	
Operator 17	Non removal		Removal	
Operator 18	Removal		Removal	
Operator 19	***5****Non removal*		Removal	
Operator 20	***6****Non removal*		Non removal	

Table 3 - Illustration of the application of the model to anticipate the removal probability of a new barrier - Flow signals at transformation area, an example in terms of "advance on the planning".

Since then, the data in terms of removal indicators for the new barrier - Flow signals at transformation area were input into the well trained SOM network, the removal anticipation result for this barrier is given one by one for 20 operators in the table. The column of observation is the result for all the barriers during the experiment. Anticipation results may be validated with the observed one, and the corresponding anticipation error is 10%.

Representing barriers as a constraint network can provide designers as tools to support the anticipation of removal of a new barrier. The designer may consider it as reference. Based on this predictive result, one can retrospectively reconsider the configuration pertinence of this barrier, final objective is to reduce the removal probability by making the benefit low, the cost high and the human operator's perception of the deficit high.

It should be noted that, in the column of observation, "removal" is judged only if one of barriers in the same barrier family is observed removed. Notice also that, in this illustration, the indicator data of a new barrier has been used to anticipate the relative removal results. In the case that a barrier is needed to be redesigned or reconfigured, the indictor data for this barrier before the redesign or reconfiguration can be gathered in the assessment of constraint based similarity (training of SOM in the example).

Conclusions

This paper has discussed a comparative model for the analysis of data on the BR so as to understand the mechanism better when human operator removes a barrier, and finally predict the consequences better on human actions for a new barrier. Indicator data before the removal of barrier and the ones after the removal have been studied and compared, preliminary analysis has been implemented. Based on this kind of feedback data analysis, representing BR indicator data and its corresponding removal results as a constraint network can provide designers as tools to support the anticipation of removal of a new barrier. Data from a railway simulator experimentation are used to illustrate its feasibility. However, only the qualitative (subjective) data on the BR have been studied in the paper.

The used approach works well, but problems may arise in the cases where less data are available. More numerous the BR data are, more realistic and objective the anticipation of BR probability will be.

In addition, it is known that human error is affected by a wide range of factors, for example task, psychological and organizational factors. These factors may vary within different situations at a given HMS system, between systems and between different peoples. Some subjective (qualitative) data judged by some human operators are affected by these factors. So, comparative study between the objective data before the removal and the ones after the removal which are measured during the experiment may be implemented in terms of performance criteria.

The perception error of risk should be considerably pointed out. During the experiment, human controller estimate by him/herself the qualitative levels of each indicator in terms of four performance criteria. Moreover, it was the first time for all the testers to participate in the experiments, the perception error of risk is therefore unavoidable. It may be partly corrected by comparing the corresponding objective performance data measured during the experiments.

In the near future, we are willing to spread out the static BR data analysis to the dynamic one with the data in the following stages of railway simulator experiments. These data will include not only the subjective data according to the BR indicators, but also the objective ones on the performance criteria. The results of experimental analysis by the proposed model will be published in another paper.

The simulator run environment was tried to make as realistic as possible, however, there is always the problem of making judgments about real life behaviour based on simulated situations. Therefore based on the simulator experiment and the experience feedback in some real industry fields, the further validation of HSOM could be performed, and the new refined method will be further applied for an European Urban Guided Transport Management System (*UGTMS*) (European Commission, 2000) project.

Acknowledgements

The authors thank Peter Wieringa and students from the Technical University of Delft and from the University of Valenciennes for their contribution to this project.

References

Basra, G., Dang V., Fabjan L., Derevinakine, A.A., Dušić, M., Furuta, T., Gibson, H., Hasegawa, T., Iqbal, M., Karsa, Z., Kim, J., Lasse, R., Mavko, B., Spitzer, C., Subramaniam, K., Tiippana, P., Venkatraman, R., Wreathall, J., Yoshimura, S., Yang, M., Zhang, Z. (1998). Collection and classification of human reliability data for use in probabilistic safety assessments, IAEA-TECDOC-1048, International Atomic Energy Agency (IAEA), Vienna, Austria.

De Gelder, P., Zhang, Z., Reiman, L., Laakso, K., Bareith, A., Subramaniam, K., Kim, J., Artemchuk, V., Mavko, B., Fabjan, L., Dang, V., Gibson, H., Wreathall, J. (1998). Collection and Classification of Human Reliability Data for use in Probabilistic Safety Assessments: Annex to the Final Report of the IAEA CRP, IAEA, J4-RC589.3B, Vienna, Austria, [COBISS-ID 13808167].

Dougherty, E. (1998). Human errors of commission revisited: an evaluation of the ATHEANA approach. Reliability Engineering and System Safety, (60)1.

European Commission (2000). GROWTH Work programme 2001-2002, Work programme for RTD actions in support of “Competitive and sustainable growth” 1998-2002, Edition December 2000.

Hollnagel E. (1999). Accident and barriers. 7th European Conference on Cognitive Science Approaches to Process Control, Villeneuve d’Ascq, France, pp.175-180.

Holmberg J, Hukki J, Norros L, Pulkkinen U, Pyy P. (1999). An integrated approach to human reliability analysis—decision analytic dynamic reliability model. Reliability Engineering and System Safety, (65):239-250.

Kaski, S., Kangas, J., and Kohonen, T. (1998). Bibliography of self-organizing map (SOM) papers: 1981-1997, Neural Computing Surveys, Vol. 1.

Kecklund L. J., Edland A., Wedin P., Svenson O. (1996). Safety barrier function analysis in a process industry: a nuclear power application. International Journal of Industrial Ergonomics, (17):275-284.

Kohonen T. (1998). Self-organizing map, Neurocomputing, (21)1:1-6.

Kohonen T. **(**2001). Self-Organizing Maps. Springer-Verlag, Third edition, Berlin, Heidelberg, Germany.

Polet, P., Vanderhaegen, F., Amalberti, R. (accepted in 2001), Modeling border-line tolerated conditions of use (BTCUs) and associated risks. Safety Science, to be published.

Polet P., Vanderhaegen F., Millot P., Wieringa P. (2001). Barriers and risk analysis. 8th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems, Kassel, Germany.

Polet P., Vanderhaegen F., Wieringa P.A. (accepted in 2002). Theory of safety-related violations of system barriers, Cognition, Technology and Work, to be published.

Pyy, P. (2000). An approach for assessing human decision reliability, Reliability Engineering and System Safety, (68):17–28.

Rousset P. (1999). Applications des algorithmes d'auto-organisation à la classification et à la prévision, PhD Thesis, University of Paris 1.

Simula O., Vasara P., Vesanto J. and Helminen R. (1999), The Self-Organizing Map in Industry Analysis, Chapter 4 in "Industrial Applications of Neural Networks", L.C. Jain and V.R. Vemuri, Editors, CRC Press, pp.87-112.

Svenson, O. (1998). Decision theoretic approach to an accident sequence: when feed water and auxiliary feed water fail in a nuclear power plant. Reliability Engineering and System Safety, (59)2:243–252.

Vanderhaegen, F., Polet, P. (2000). Human risk assessment method to control dynamic situations, 4th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes, SAFEPROCESS2000, Budapest, Hungary, June 14-16 2000, pp.774-779.

Vanderhaegen, F., Polet, P., Zhang, Z., Wieringa P. A. (2002). Barrier removal study in railway simulation, PSAM 6, Puerto Rico, USA, June 2002.

Zhang Z., Polet P., Vanderhaegen V., Millot P. (2002). Towards a method to analyze the problematic level of Barrier Crossing, lm13/Esrel2002, Lyon, France, pp.71-80.

Zhang Z., Vanderhaegen F. (2002). A method integrating Self-Organizing Maps to predict the probability of Barrier Removal, C. Warren Neel Conference on the New Frontiers of Statistical Data Mining and Knowledge Discovery, Knoxville, Tennessee, USA, June 22-25, 2002.

The Control Of Unpredictable Systems

Björn Johansson, Erik Hollnagel & Åsa Granlund

CSELAB, Department of Computer and Information Science
University of Linköping, Sweden
bjojo@ida.liu.se; eriho@ida.liu.se; asagr@ida.liu.se

Abstract. In this article we argue that success in controlling dynamic systems is equivalent to the successful detection of and recovery from unwanted performance variability. It is also argued that there are fundamental differences between the systems that can be targets of control, ranging from constructed dynamic systems of the process industry, where the system itself supports detection and recovery of deviations, to more *ad hoc* systems like fire-fighting or military situations. We propose an assessment framework to compare different kinds of joint cognitive systems in terms of the likelihood of detecting unwanted performance variability. It is further suggested that the efficacy of such detection is related to how well articulated and disseminated the desired target state is within a system. Two examples from actual systems show 1) what happens when a controlling system lacks an articulated steady state, and 2) how control is maintained in a system by joint negotiation of the desired steady state.

Introduction

In order effectively to control a system, it is necessary to be able to detect variability that may lead to unwanted consequences – in the following called deviations – and to provide adequate counteractions. Effective monitoring of system performance and detection of deviations from the desired state is therefore a pre-condition for efficient recovery. The monitoring/detection in turn depends on whether it is possible to define what the desired target state of the system is, and whether appropriate indicators of this state and criteria for deviations can be found. In this paper, we discuss some fundamental differences between control of constructed and natural dynamic systems. We suggest a simple framework for defining the pre-conditions for performance deviation detection by a Joint Cognitive System (JCS), both in terms of how well established the understanding of the desired target state is, and how well it is monitored and articulated by the JCS. A cognitive system (CS) is defined by its ability to modify its behaviour on the basis of past experience so as to achieve specific anti-entropic ends (i.e., being able locally to resist an increase in entropy). A JCS is defined as two or more natural or artificial cognitive systems that work together to achieve an overall goal. A JCS is thus defined by what it does (its function) rather than by what it is (its structure). The boundary between the JCS and the process is consequently defined relative to the chosen level of description and purpose of analysis, rather than in terms of structural considerations (architecture, organisation). Throughout this paper, we consider a JCS as a system concerning of several persons working with and through technology to achieve a goal, referred to as the "desired target state". Since the boundaries reflect the chosen level of description, the joint cognitive system could be expanded to comprise other functions in an organisation, with other and sometimes conflicting objectives. For reasons of clarity, we here focus on the JCS described above and its goal of maintaining a steady state through detection and counteraction.

Control

Brehmer & Allard (1991) described the problem of control as "the problem of finding a way to use one process to control another process". This definition is applicable to most every-day situations. When we want to cross a street, we press the switch for the traffic lights in order to change the signal and stop the traffic, thereby allowing us to get to the other side without being hit by cars. The number of processes that are affected by pressing the button is small if we only consider the inner workings of the traffic lights, but fairly large if we consider the functional relations between the traffic light, the cars and their drivers, the traffic flow, etc. Indeed, in several systems the traffic lights at the next intersection may also be affected.
Another description of control focuses on the effects by noting that "(t)he essence of control is that unwanted deviations from the desired or prescribed course of development are reduced to a minimum or do not occur at all" (Hollnagel, Cacciabue & Hoc, 1995, p. 7.) This means that control is lost if major deviations occur, such as hitting another car while driving. The notion of control is closely related to the Law of Requisite Variety (Ashby, 1956). According to this the regulator of a process must have at least as

much variety as the process. It follows from this that feedforward control is possible only if the regulator has perfect knowledge about all the possible disturbances and external influences that may occur. This is, however, never possible to achieve in practice; procedures and guidelines always need to be adjusted or filled in by the operator. Perfect feedforward control is also in conflict with the "fundamental regulator paradox". According to this, the task of a regulator is to eliminate variation, but this variation is the ultimate source of information about the quality of its work. Therefore, the better the job a regulator does the less information it gets about how to improve (Weinberg & Weinberg, 1979, p. 250) its performance.
System complexity and the need for feedback make it hard to envisage an operator trying to control a process without ever deviating from the desired way of doing it. This is especially true in the control of dynamic systems, since it is often hard to judge if the changes in a process are due to operator actions or if the process simply changed by itself (Brehmer, 1987). Control is then, to a large extent, an issue of understanding how to bring a target system into a certain, desired state and how to maintain it in that state once it has been reached.

Performance variability management

In light of the above considerations, it is easy to understand why the issue of detection and recovery from deviations has become a growing area of interest (e.g. Kanse & Van der Schaaf, 2000). By recovering from a deviation, a process is kept within a desired envelope of safe and efficient performance, and the control of the process can be seen as successful. A related area of interest is accident prevention, which – in contrast to the recovery function – mainly concerns creating barriers of different kinds physical, functional or symbolic where the latter often depend on training.

Deviations And The Steady State The articulation of what a steady system state is provides the foundation for detection of failures and deviations, since it effectively constitutes the norm to which everything is compared. A novice driver of a car will, for example, not notice small things like a quirking sound or an unsteady movement resulting from low tire pressure. The novice driver will usually be so occupied with handling the basic functions of the car that the threshold for detection is very high. An experienced driver, on the other hand, reacts to very subtle noises or changes in the car because they are meaningful – which effectively means that it possible to detect and recognise them.
When controlling constructed dynamic systems, such as industrial processes, the steady state is usually well understood by the JCS. In natural dynamic systems, the steady state depends on the operator's ability to understand the target system, and hence on the available articulated knowledge and experience. The understanding of the steady state may furthermore change according to the current conditions and fluctuate over time. For instance, in the case of a forest fire, a desired steady state may initially be that the fire has been extinguished and the appropriate action may therefore be to dump as much water as possible on it. If, however, the direction and force of the wind change, the desired steady state may be to ensure that no people are harmed, and the appropriate action will therefore be to evacuate people from an endangered area. This means that the steady state must be negotiated within the controlling system (the JCS). It also means that if there is no common articulation of the desired state, local interpretations will appear.
For many dynamic systems, the desired target state is defined in terms of a range of acceptable performance rather than as a uniquely defined state. For instance, the tuning of a process increases the level of control and the efficiency of the system iteratively. An accident is a deviation that exceeds the allowed range, and thus compromises safety or the fundamental systems functioning.
Hollnagel (2000) has proposed the principles of "performance variability management" as a way of viewing and handling these kinds of problems. The basic idea behind this is that all processes have variability and that the variability can be identified and monitored. In order to keep the process under control, the variability that may lead to negative or unwanted outcomes should be reduced while the variability that may lead to improvements and advances should be amplified. Performance variability management envisages four fundamental management functions called monitoring, detection, deflection and correction, cf. Table 5. Monitoring takes place when the system is in a steady state, for instance the normal production state of an industrial process such as a power plant or an oil refinery. Detection takes place when a deviation from the steady state occurs, which may mean that the process is in a pre-accident state – for instance the pre-crash state in traffic safety models. Deflection characterises the attempt to limit or contain the immediate consequences of an accident or malfunction. Finally, correction is concerned with the recovery from an action or malfunction.

The detection function is especially interesting since it presupposes effective monitoring, while monitoring in turn is based on assumptions about what can be monitored, thus in a circular fashion describing *what can be detected*. Since monitoring refers to the notion of the steady state of the system, it requires that we can define what this state really is, something that is far from trivial.

Table 5: Main functions of performance variability management (from Hollnagel, 2000)

Accident stage	Management function	Examples
Steady state performance	Monitor	Observe: system states and trends Confirm/Verify: responses, resources Manage: roles, permissions, schedules Record: data, unusual events
Pre-accident (build-up)	Detect	Identify: deviations, precursors, indicators Verify: functions, limits, barriers Comply with: criteria, rules, procedures Query: habits, assumptions
Accident	Deflect	Attenuate: retard rate of change Partition: separate into time and space Reduce: contact surface, duration Strengthen: resources, supports
Post-accident (recovery)	Correct	Replace: functions, resources Modify: rules, criteria, procedures Improve: structures, design staffing Query: explanations, demands, accepted wisdom ("givens")

The fundamental problem of monitoring and detecting of performance deviations is that in principle we can only detect and monitor what we know to be possible. In practice this is further narrowed down to what we expect. On a theoretical level this corresponds to what Neisser (1976) described as the perceptual circle, in which perception is based on schemata. Or going even further back it reflects Peirce's dictum that "every cognition is determined logically by previous cognitions" (Peirce, 1868). On a practical level it is demonstrated for instance by the problems in looking for possible latent (failure) conditions and Sneak Paths.

Constructed And Natural dynamic systems

Most systems studied in relation to control and error recovery clearly can be defined as dynamic, which means that the system states may change both as a result of the controller's actions and because of changes in the systems themselves (Brehmer, 1987). This makes it difficult to predict their development. While much of the research has dealt with constructed systems – i.e., process artefacts – it is equally important to consider systems that are naturally dynamic, and which are native (or spontaneous) in the sense that they are naturally occurring rather than the result of deliberate constructive efforts.

Constructed systems are designed to provide a specific function and usually include explicit control functions such as in a power plant or a factory. The system to be controlled –the target system – is in these cases clearly dynamic in the above sense, and is usually referred to as a process. Since the nature of the process is known it is possible to include means for monitoring and control as part of the system, and many efforts have gone into defining and designing an optimal interface between the human controllers and the process. For instance, Lind & Larsen (1995) have described how functional descriptions of designed systems (in this case a power plant) can be used to increase operator understanding of the system to be controlled.

By natural systems we refer to processes that are not explicitly designed and sometimes not even desired, like a forest fire. Despite this, such systems still need to be controlled in the sense defined by Hollnagel et al. (1995) above. The reference description of the behaviour of such a system – in common parlance, the model – is much harder to define and artefacts and procedures that can be use for control are consequently less well elaborated. In the case of a forest fire, for instance, the operator or decision maker has to find "sensors" and define control measures as well as possible within the constraints of the situation. There is normally high demand on the ability to predict performance in natural systems while at the same time, the feedback is more difficult to obtain and/or also of incomplete or of a lower quality. In natural systems, the desired target state depends on the operator or decision-maker's ability to understand the system as well as on the ability to articulate this in terms of appropriate control actions. The constructed and natural systems constitute two extremes. In reality, systems can be hybrids of constructed and natural systems. An example of this is systems for controlling radio nets (nets for cellular telephony traffic). These systems are constructed in the sense that the radio theory that forms the basis for the design is well understood and predictable. In reality, the controlled process is extremely complex and depends on such a large number of natural factors that the behaviour is often undesired, and difficult to both predict and control. For the clarity of argument, we will in the following concentrate on the two extremes, the constructed and the natural systems.

An important difference between constructed and natural systems is that there is limited time available for "getting to know" the process. The operators' understanding therefore greatly depends on what they already know – their existing "model" of the system – as well as on other articulated descriptions. The dissemination of adequate descriptions of the desired target state is therefore a crucial determinant of how well a system is able to control a process. Here, it is important to note that the JCS trying to control the natural system in it self might be *ad hoc* because natural systems, like forest fires, often demands larger and more complex control functions than the ones normally available. In such cases, the establishment of the JCS in it self becomes a task.

Characteristics of detection and recovery

Practice has shown that it is difficult to model detection and recovery for constructed and natural systems alike, although for different reasons. In order better to understand these difficulties, we have been working on a framework by which detection and recovery for the two types of systems can be compared, cf. Table 6. The purpose of such an approach is systematically to assess the possibilities for detection and correction for a given target system. As argued above, one of the fundamental aspects of process monitoring and detection of deviations is the ability to determine the steady state of a system or process. We also argued that the "steady state" of a system could be defined in several different ways depending on the nature of the system. In the case of constructed systems and explicitly designed control environments, the definition of the "steady state" is part of the design. In the case of natural systems, hence more *ad hoc* control tasks – such as a forest fire or a military situation – the "steady state" depends on the operator's ability to understand the system, hence on the available articulated knowledge and experience. Understanding how this knowledge and experience is obtained and maintained will be useful in developing support for the operators to monitor and interpret data from the process.

Table 6: Assessment of differences in monitoring and detection between natural and constructed dynamic.

	Natural system	Constructed system
Clearness of the steady state	Defined by the operator's understanding, based on assumptions and previous experiences. The articulateness of the model depends on how often the state occurs. For example, a fire brigade on a routine call to an apartment fire may know well what to do, while a fire in less common environments causes greater uncertainty.	Designed into the artefacts and procedures used by operators. Alarms, physical boundaries, rules etc reduces the likeliness of unforeseen deviations from the desired state.
Stability of the target system state	Depends on type of target system. A forest fire, for example changes depending on wind, humidity etc.	Probably high. Self-regulation often maintains a steady state, which only changes when normal process is disturbed.
Distribution of the steady state description within the JCS	Depends on possibility to do disseminate (communication channels, time available)	Normally high. Operators normally know how the system should perform within their own area of responsibility.
Information about the current state (Sensor coverage, feedback delays, signal validity, data reliability)	Depends on the organisation involved. A military organisation normally has an array of sensors available. Other *ad hoc* organisations may be less able to gather information.	Normally high since monitoring tools are part of the system design. Very unlikely events may be hard to monitor because the interface has not provisions for them.
Distribution of Information about current state (who gets it, how fast)	Depends on organisational structure and information system structure.	Depends on organisational structure and information system structure.
Surveillance of information about current state	Possibly high if sensors are available.	Likely to be high. Most factories and plants have employees working full-time with sensor monitoring.
Implicit and explicit understanding of target state in relation to current state	Depends on prior knowledge of task (education, experience, common ground) and how briefings and instructions have been carried out.	Implicit understanding likely to be high if personnel have been employed a longer period and know each other. Explicit understanding depends on how briefings and instructions have been carried out.
Vulnerability, susceptibility to disruptions	Often loosely coupled, hence more resistant to disruptions.	Usually reciprocal to the complexity of the control system, hence often high.
Stability of JCS (in terms of resources, structure, means of communication etc.)	In certain cases, the controlling system may be subject to dramatic changes, both in terms of structure and resources. Judging the state of the JCS itself might thus be a major task for the operator/operators.	Mostly stable, unless very unusual events cause major changes in the conditions of the controlling system, for instance total power failure, disease etc.

Discussion: The Steady State

In order effectively to control a system it is necessary to be able to detect deviations and to provide adequate counteractions. Efficient recovery therefore depends crucially on effective monitoring and detection. That in turns depends on whether it is possible to define what the desired target state of the system is, and whether appropriate indicators can be found.

The following examples from actual, natural dynamic systems show 1) what happens when a controlling system lacks an articulated steady state, and 2) how control is withheld in a system by joint negotiation of the desired steady state.

Example: failure to establish and disseminate steady state

In an example from the study of a large forest fire in Sweden (Johansson & Artman, 2000) the fire spread so rapidly that the commanders lost contact with their fire-brigades, causing very dangerous situations for the involved personal and also to some extent ineffective resource allocation. Without an articulated agreement on how the fire fighting should be co-ordinated, the fire fighters kept going on their own, making local assumptions about the situation. Since they did not have a good understanding of the development of the fire, situations arose where fire fighters found themselves surrounded by fire. Luckily, no one got seriously injured. Misunderstandings also occurred between professional fire fighters and volunteers because of unclear instructions and lack of understanding about the control process (forest fire-fighting) by the volunteers (Johansson & Artman, 2000). The combination of deficient communication and differences in the implicit understanding of the desired target state made it difficult for the commanders to articulate their intentions about how to reach that state.

Example: Negotiation and establishment of steady state

In a study of the operations and maintenance of a radio network run by a large and well established telecom operator (Granlund, 2002), it was found that the negotiation and establishment of the steady state not only increased efficiency of the tuning of the net and the handling of alarms, but was in fact a necessary and crucial activity for the management of the operations. An ongoing exchange of information between the different work teams (for instance, construction, fault management, tuning, planning and hardware) supported the establishment of knowledge about the net - the history of net behaviour and performed corrections - and provided an understanding of possible and desirable behaviour. Also, short term, this knowledge formed the basis for some decisions regarding tuning and alarm correcting actions.

General discussion

One important aspect in the definition of the desired target state is that there might be a difference between the articulated state and an objectively ideal state. It is important to recall that the articulation of the state is based on how the operators or commanders *understand* the target system and the control system, rather than what it is in an objective sense. In several *ad hoc* studies of disasters, operators have been criticised for making decisions that led to a worsening of the situation. This is clearly wrong, both because such studies represent *ex post facto* reasoning, and because they fail to recognise the complexity of the dynamic situation.

From the point of view of this paper, defining the “desired target state” is a task of understanding and articulation, in the sense that articulation must be allowed to take place within the JCS in a way that makes the situation understandable for all involved. It is no use to articulate an intention or target state if the persons it concerns cannot understand it. Builder, Banks & Nordin (1999) has developed a model for “command concepts” in the military area. A command concept is a “vision of a prospective military operation that informs the making of command decisions during that operation. If command concepts are the underlying basis for command, then they provide an important clue to the minimum essential information that should flow within the command and control systems” (Builder, Banks & Nordin, 1999, p 14). This argument makes an important contribution to the design of information systems for spontaneous organisations, but it fails to recognise the fact that the command concept/articulation rarely is known in advance, hence making it difficult to implement in a design. The pre-understanding of a spontaneous system can usually exist only on a general, non-contextual level.

Another important aspect of the steady state is that it concerns both the target system and the controlling system. In most cases the controlling system is well defined in term of resources and competence. Spontaneous systems, however, often lead to *ad hoc* JCS as in the example of the forest fire above. In that case the JCS was subject to major changes due to different numbers of volunteers from more or less stable

organisations (national guards, sports clubs, locals etc). This caused great uncertainty about resources, the right to allocate them in different situations and their competence level. Being able to judge the state of the controlling system itself (the JCS) was a great burden for the commanders in charge.
Clearly, these questions have great implications for the view on feedforward in information system design. Today, much of the discourse and research concerning this focuses on feedback and different ways to gain faster, more accurate and better presented feedback. Solutions to the problem of detection of errors and deviation seem to be more feedback at a higher rate rather than trying to establish *what kind* of feedback that is important. If instead focus was on how to support articulation of steady states and the dissemination of this, we would gain insights in how to build systems for appropriate feedback.

References

Ashby, W.R. (1956) *An introduction to cybernetics.* London: Chapman and Hall.

Brehmer, B. (1987) Development of Mental Models for Decision in Technological Systems. In (eds.) Rasmussen, J., Duncan, K. & Leplat, J., *New Technology and Human Error*. Suffolk: John Wiley & Sons Ltd.

Brehmer, B & Allard, R. (1991) Real-time dynamic decision making. Effects task complexity and feedback delays. In J. Rasmussen, B. Brehmer & J. Leplat (eds.), *Distributed decision making: Cognitive models for cooperative work.* Chichester: Wiley.

Builder, Banks & Nordin (1999) Command Concepts – A Theory Derived From the Practice of Command and Control. RAND Report MR-775-OSD.

Granlund, Å. (In progress) Trust in Automation - Field studies in radio network control Proceedings of *NES 2002*. Nordiska Ergonomisällskapets 34:e årliga konferens, Norrköping, Sweden, 1-3 October 2002.

Hollnagel, E. (2000) Performance Variability Management. Proceedings of *People In Digitized Command And Control Symposium*. Royal Military College of Science at Shrivenham, UK. 12-14 December 2000.

Hollnagel, E. Cacciabue, P.C. & Hoc, J. 1995 Work with Technology: Some Fundamental Issues. In (eds.) Hollnagel, E. Cacciabue, P.C. & Hoc, J., *Expertise and Technology – Cognition & Human-Computer Cooperation*. Lawrence Erlbaum Associates. Hillsdale, New Jersey.

Johansson, B. & Artman, H. (2000) Interkulturell Kommunikation vid Nödsituationer – En Förstudie. *Forskningsrapporter från ÖCB.* Överstyrelsen för Civil Beredskap: Stockholm.

Kanse, L. & Van der Schaaf, T. W. (2000). Recovery from failures – Understanding the positive role of human operators during incidents. In: D. De Waard, C. Weikert, J. Hoonhout, & J. Ramaekers (Eds.): *Proceedings Human Factors and Ergonomics Society Europe Chapter Annual Meeting 2000*, Maastricht, Netherlands, November 1-3, 2000, p. 367-379.

Lind, M. & Larsen, N.L. (1995) Planning Support and Intentionality of Dynamic Environments. In (eds.) Hollnagel, E. Cacciabue, P.C. & Hoc, J., *Expertise and Technology – Cognition & Human-Computer Cooperation*. Lawrence Erlbaum Associates. Hillsdale, New Jersey.

Neisser, U. (1976) *Cognition and Reality*. San Fransisco: Freeman.

Peirce, C. S. (1868). Some consequences of four incapacities. *Journal of Speculative Philosophy*, *2*, 140-157. (Reprinted in P. P. Wiener (Ed.), Charles S. Peirce – Selected Writings. New York: Dover, 1958.)

Weinberg, G. M. & Weinberg, D. (1979). *On the design of stable systems*. New York: Wiley.

Finding Order in the Machine

Mark Hartswood[1], Rob Procter[1], Roger Slack[1], Mark Rouncefield[2]

[1]ICCS/HCRC, Division of Informatics, University of Edinburgh
mjh|rnp|rslack@cogsci.ed.ac.uk
[2]Department of Computing, University of Lancaster
m.rouncefield@lancaster.ac.uk

Abstract: This paper examines the use of a computer-aided detection system for mammography and argues that it is important to appreciate work practice in order to understand the ways that the system will be used outwith the context of clinical trials. As an illustration of the lacuna, the 'missing what' in clinical trials, we show how order is found in the lived work of using the system and how readers make sense of its behaviour in the work setting. The paper concludes with a call for the use of ethnography in trials as a means of explicating the uses of technologies in real-world situations.

Keywords: mammography; ethnography; reading practice; clinical trials.

Introduction

The aim of this paper is to show how clinical trials elide the 'lived work' (Livingston, 1986) of, for example, doing reading mammograms. We show how such trials presumptions leave out the wider interactional contexts of doing work in real world settings; what we would call the sociality of work. The paper argues for the introduction of context-sensitive methodologies such as ethnography to explicate this lived work. Such a technique would consider technologies at the point of use in the work setting as opposed to the trial context. It is only through such considerations that we can appreciate concepts such as performance impact, usability and utility in their fullest sense.

Technology in Action

Breast cancer accounts for one fifth of all deaths from cancer among women in the UK. Established in 1988, the goal of the UK breast screening programme is to achieve a reliable cancer detection rate. Clinicians or readers[22] of mammograms, i.e., radiological images of the breast, are required to find what might be small and faint features in complex visual environments and to ensure that they detect as many cancers as possible (true positives) while keeping the number of women recalled unnecessarily (false positives) as low as possible. A number of computer-aided detection (CAD) systems have been developed which analyse the mammograms and 'prompt' readers to look at suspicious features; this paper details the trial of one such system.

The CAD system on trial consists of two components – the mammogram scanning and analysing unit and the mammogram viewing box with built-in displays for visual prompts. Mammograms on the viewing box are scrolled up and down. The prompts are synchronised with the mammograms, but the timing of their presentation is controlled by the reader. The system prompts for two types of features that are early indicators of breast cancer: micro-calcification clusters -- small deposits of calcium visible as tiny bright specks; ill-defined and stellate lesions -- areas of radiographically-dense tissue appearing as a bright patch that might indicate a developing tumour. The former are marked by a shaded triangle, the latter by an asterisk and a circle is drawn around either prompt type if the system's confidence is high.

Readers were observed doing the various trial sets and then asked about their experiences of using the prompts. Readers were also taken back to cases identified in the trial set where they had appeared to have had difficulty or spent a long time making their decision, and asked to talk through any problems or issues to do with the prompts and their decisions. Although there were variations in how readers approached a reading and the trial, the fieldwork extract below gives some idea of the process observed:
Case 1: Gets blank mammogram to mask area of the mammogram (so I can concentrate on it ... these are set up differently from the way I usually look at them ... so I have to train my eye each time."). Using

[22] Most (but not all) readers are qualified radiologists. We will use the more general term of reader.

magnifying glass. Marking on booklet. Looking from booklet to scan. Homing in on an area -- "I'd say it's benign."
Case 2: Using blank mammogram. Takes mammogram off roller and realigns. Magnifying glass. Looking from booklet to mammogram. "I'd not recall ... what the computer has picked up is benign ... it may even be talcum powder."
Case 10: Looking at mammogram - using blank mammogram to mask area. Magnifying glass. Looking at booklet prompts - looking back at mammogram. "This is a case where without the prompt I'd probably let it go ... but seeing the prompt I'll probably recall ... it doesn't look like a mass but she's got quite difficult dense breasts ... I'd probably recall."

As with everyday reading so with the trial, readers used a repertoire of manipulations to make certain features 'more visible'. A magnifying glass may be used to assess the shape, texture and arrangement of calcifications or, where the breast is dense, the mammogram may be removed and taken to a separate light box. Where a reader wished to attend to a particular segment of the mammogram, another mammogram may be used to blank off a part of it. In cases where a suspicious feature was seen on one view, readers used their fingers or an object such as a pen for measurement and calculation. These repertoires of manipulations are an integral part of the embodied practice of reading mammograms.

Strengths of the CAD system in supporting this kind of work lay in picking up subtle signs that some readers felt they might have missed and stimulating interaction between reader and the available technology by prompting them to re-examine the mammogram. Of course, this does not carry within it a decision as to what to do with the prompted feature – that requires decisions to be made by the readers. Readers also frequently express the opinion that they are better at 'spotting' some cancers -- as having skills or deficiencies in noticing particular types of feature within mammograms. This was another area where the CAD prompts were seen as useful, as both compensating in some (consistent) way for any individual weaknesses of the reader and as a reminder of good practice.

Two sources of uncertainty for readers can be found in deciding whether a mammogram is not recallable: first, a detected feature warrants a recall (if the feature is 'sufficiently suspicious' or regarded as artefactual and so on) and, second, satisfaction of search (when does one give up looking for features?). The aim of the CAD system is to deal with the second dimension as opposed to the first. Our previous studies have shown how readers reflexively adapt their work practices in order to build and sustain their 'professional vision' (Goodwin, 1994), and that this, in turn, contributes to the management of individual and collective performance. Readers have evolved an 'ecology of practice' for performance management that is deployed as part of the routine of the work. (Hartswood, Procter, Rouncefield and Slack, 2002). Through artful use of the public character of the screening reporting form and annotation work, readers use double reading to make their work observable-reportable to manage the uncertainties mentioned above -- when it is arguably most salient -- as they do it. Our interest here is to examine the role of the CAD system in the management of uncertainty: does the system manage uncertainty or create yet more areas of uncertainty for readers?

The CAD system should not be taken to make things less uncertain – decisions still have to be made and these fall to the readers (c.f. case 10 above). The prompts are, so to speak, docile in that their character is simply to prompt, as opposed to say what should be done. In the above fieldwork, we see that readers attempt to ascertain what a prompted feature is. That a prompt occurs is a meaningful thing, but what to do about it is still a readers' matter. It seems to us that the missing elements in these trials – the readers' sense making is not taken into account. There is still a deal of sense-making to be done in order to account for what the system is showing as accountably this or that feature warranting this or that decision. In other words, the system still requires the professional vision of the reader to remedy prompts as what they accountably are.

The question is how readers make sense of the CAD system. Following Schütz, we might argue that readers render mammograms intelligible using a mosaic of 'recipe knowledge': "a kind of organisation by habits, rules and principles which we regularly apply with success." (Schütz, 1972:73). While the common experiences and rules embodied in the mosaic are always open to potential revision they are, nevertheless, generally relied upon for all practical purposes as furnishing criterion by which adequate sense may be assembled and practical activities realised. Unlike everyday interaction, the CAD system cannot repair what it 'means', and difficulties can arise as readers rush to premature and often mistaken conclusions about what has happened, what is happening, what the system 'meant', what it 'is thinking', and so on (Hartswood and Procter, 2000). The reader using the system for the first time is confronted with docile prompts and what the meaning of these becomes apparent only when one looks at the natural history of what the system's prompts have come to (obviously this can only be done retrospectively).

Everyday Reading Work and the Problem of Clinical Trials

Our initial evaluation of the CAD system also raises a number of questions concerning the appropriateness of quasi-clinical trials for technological innovations. Divorced from the lived reality of everyday reading work and the various affordances of the work setting (such as the annotated record, previous mammograms or a medical biography), the value of such a trial for the deployment of the technology in a (very different) real setting is in some doubt. Following Berg (1997), we argue that new technologies must become part of local work practices and practical reasoning – they are used in context and not in some abstract ideal-typical setting, such as are supposed in clinical trials. The record of achievement in the field of clinical support systems is patchy: many systems perform well in laboratory trials but are found wanting in use. The design rationale for clinical support systems, for example, often assumes generic difficulties, whereas clinicians' needs may be highly specific. The clinicians' problem becomes re-formulated in terms of what the technology can do, rather than their actual needs.

In that they are closely linked to existing technology affordances, the work setting practical actions and practical reasoning of clinicians raises important questions for the design and use of new technologies. They also raise questions as to whether the changes in practice changes that new technologies are intended to support are actually achievable. We might, indeed, question why one seeks to replace current practice as opposed to supporting it: the use of new technologies seems too often to reconfigure as opposed to supporting existing practice. This image of medical technology as panacea seems to be borne out by the logic and conduct of trials: in such contexts it is possible to eliminate all 'worldly' contingencies such as work practices and thereby to suggest a clear perspective on the impact of the tool. We argue that things are not as clear-cut in that trials take place in a rarefied world – what happens out in the 'messy' 'real world' is a different matter.

Conclusions

Reader training programmes for CAD systems may need to be designed to provide not only a resource for initial familiarisation, but also to support the continued learning of users and evolving of practices. The issue of change over time also raises some wider issues of evaluation in that many of the benefits of information are unlikely to be evident until several years or more after its introduction and adaptation to the particular circumstances of use. Yet, inevitably, evaluations are set up to investigate evidence of immediate benefits. The experimental context in which these trials are undertaken elide the social character of the work involved and thereby erase some of the most crucial dimensions of readers' work – we would argue that these need to be put back in order to have technologies that afford work in real settings as opposed to clinical trials. We do not advocate scrapping clinical trials, rather the point is to put back some of the context and to explicate tool use in context: such an appreciation would provide a more robust investigation of what tools actually do in practical work settings.

References

Berg, M. (1997). *Rationalising Medical Work: Decision Support techniques and Medical Practices*. Cambridge: MIT Press.
Goodwin, C. (1994). *Professional Vision*. American Anthropologist. 96; 606-633.
Hartswood, M. and Procter, R. (2000). Computer-aided mammography: a case study of coping with fallibility in a skilled decision-making task. *Journal of Topics in Health Information Management*, vol. 20(4), May, p. 38-54.
Hartswood, M, Procter, R., Rouncefield, M. and Slack, R. (2002). *Performance Management in Breast Screening: A Case Study of Professional Vision and Ecologies of Practice.* Journal of Cognition, Technology and Work, vol. 4(2), p. 91-100.
Livingston, E. (1986) *The Ethnomethodological Foundations of Mathematics* London: RKP.
Schütz, A. (1972) *The Phenomenology of the Social World* London: Heinemann Educational.
.

Accomplishing 'Just-in-Time' Production

Alexander Voß[1], Rob Procter[1], Roger Slack[1], Mark Hartswood[1], Robin Williams[2], Mark Rouncefield[3]

[1]ICCS, Division of Informatics, University of Edinburgh
[2]Research Centre for Social Sciences, University of Edinburgh
[3]Computing Department, Lancaster University
av|rnp|rslack|mjh@cogsci.ed.ac.uk, r.williams@ed.ac.uk, m.rouncefield@lancaster.ac.uk

Abstract: We present an ethnographic study of work in the control room of a manufacturing plant. While work in the plant is oriented towards a very strict production orthodoxy and is to large degrees automated, we find that the overall dependability of the plant is not so much the outcome of careful planning that has gone into the design of the production management system. Rather, it is a situated accomplishment resulting from the work of competent members going about their everyday activities, which are oriented towards and made accountable through the production orthodoxy as opposed to being determined by it.
Keywords: production management, situated accomplishment

ENGINECO: The Boundaries of Planning and Control

We present findings from an ethnomethodologically-informed ethnographic study (Hughes et al 1995) of work in ENGINECO, a manufacturing plant producing mass-customised diesel engines. We illustrate some of the working practices of Control Room workers as they attend to the contingencies of production management and control. The findings support a contingent view of production planning and scheduling, leading us to argue that the implementation of production plans is always a practical and situated activity, the character of which emerges in action. The contingent view emphasises the incompleteness of knowledge and the set of circumstances - more or less intended, arbitrary, uncontrolled or unanticipated - that affect action (Dant and Francis 1998). In contrast to the rationalist view, the implementation of a production plan is a production worker's formulation, produced in response to issues concerning the 'local logics' of day-to-day production management. This points to the dynamic yet situated nature of knowledge and plans, the "minor actions, minor decisions and minor changes", upon which the organization rides (Boden 1994.). That is to say, local logics attend to the incompleteness of knowledge on both organizational and spatial-temporal levels - that which is an acceptable solution *just here and just now* with these circumstances and in this organisational context is a basis for proceeding *right now*. Decisions are made in the fabric of both real space and time which as Boden notes "is (...) an accountable matter (...) open to appraisal, revision, condemnation (and) repetition", (*op cit*, p. 192). This stands in marked contrast to the rationalist view of planning where plans stand as directives - 'scripts for action' - for future actions produced out of a systematic analysis of the possible options and constraints on their application.

Normal, Ordinary Troubles of Production

The production environment at ENGINECO is shaped according to a just-in-time (JIT) production orthodoxy centred around the notion of the 'buildability' of an order which needs to be guaranteed before production can start. The management of production should not, however, be read as the straightforward instantiation of a plan but rather it is a situated accomplishment that invokes the spirit rather than the letter of 'buildability' (Voß, Procter, Williams 2000). While one might not want to say that the concept of JIT has been abandoned, it has certainly been appropriated to local contingencies such as scarce material or inadequate performance of the logistics system. Control room workers have taken over some responsibilities from assembly planning regarding the management of orders and material and they may well schedule orders for production although material is still on its way to the plant. So, while originally buildability was a relationship between the order with its bill of material and the inventory, now it involves a judgement made by control room workers about how 'things will work out'. The above discussion points to some of the worldly contingencies that Control Room workers routinely deal with as a part of their planning and scheduling work. More precisely, all plans are contingent on "situated actions" (Suchman 1987). We found a series of expectable or ordinary troubles whose solution is readily available to members in and as a part of their working practices. That is, such problems do not normally occasion recourse to anything other than the usual solutions - where a problem contains within it a candidate (used-before-and-seen-to-work) solution. These problems and their solutions are normal and natural in and as a part of everyday work invoking a search through a series of seen-to-work-before repertoire of candidate solutions.

This does not mean that any old behaviour will fit as a candidate solution. The worker facing a problem draws upon a repertoire of candidate solutions and thereby makes themselves accountable as having done this or that (this might also be a way of making later actions (such as placing the problem in the complaint book or calling in an engineer) accountable - having tried all the usual solutions one has the 'right' to call on others to attempt to remedy the problem and one is accountable in this manner "I have tried this and that but they did not work, therefore I have done this".

Dealing with Unexpected Troubles

Problems not susceptible to these remedies also demand a solution – workers cannot remain indifferent to their presence – but by definition that solution is not a normal or usual one. In order to keep production running, workers have to find and evaluate possible solutions quickly, taking into consideration the present situation, the resources presently available, as well as, ideally, any (possibly long-term and remote) consequences their activities might have:

From fieldwork notes: *A material storage tower went offline. Material could be moved out of the tower to the line but no messages to the Assembly Control Host were generated when boxes were emptied. Control Room workers solved this problem by marking material in the tower 'faulty' which resulted in new material being ordered from the logistics provider. This material was supplied to the line using forklift trucks. […] A material requirements planner called to ask why so many parts were suddenly 'faulty'.*

Such situated problem-solving results in workarounds that are initially specific to the situation at hand but may become part of the repertoire of used-before-and-seen-to-work candidate solutions. They may be further generalised through processes of social learning as workers share the various 'local logics' with colleagues or they might in fact get factored into the larger socio-material assemblage that makes up the working environment. This process of problem solution and local logics, however, is critically dependent on members' orientation to the larger context, their making the solution accountable to colleagues and their ability to judge the consequences. The following fieldwork material illustrates how problem solutions can get factored into ongoing development of the production management system as well as how they can adversely affect the success of the system:

From an interview with one of the system developers: *[Such a complex system] will always have flaws somewhere but if the user has to work with the system and there's a problem he will find a work-around himself and the whole system works. [...] The whole works, of course, only if the user really wants to work with it. If he says: "Look, I have to move this box from here to there and it doesn't work. Crap system! I'll let a forklift do this, I will not use your bloody system" then all is lost. Then our location information is wrong […] then it will never fly. [If they come to us and say] that something's not working, we will say "oh! we'll quickly have to create a bugfix and, for the moment, I'll do this manually without the system", then it works, the system moves on, everything stays correct, the whole plant works and if the next day we can introduce a bugfix the whole thing moves on smoothly.*

This bears on the possibility of offering a fully automated solution to planning and production management. It is difficult to see how one could solve unexpected problems in an automated manner. Human intervention (and resourcefulness) is required to find and implement a solution to the problem. But the boundaries between the types of problem are semi-permeable. Members will view different problems in a variety of ways and this may lead to the resolution for the problem through the ability to recognize some kind of similarities inherent in this and a previous problem through the phenomenon of organizational memory (Randall et al 1996). As in other collaborative work (see e.g., Hartswood and Procter 2000), members are aware of, and orient to, the work of their colleagues. This is supported by the affordances of their socio-material working environment as the following example illustrates:

From a video recording of Control Room work: *Oil pipes are missing at the assembly line and Jim calls workers outside the Control Room to ask if they "have them lying around". This is overheard by Mark who says that: "Chris has them". He subsequently calls Chris to confirm this: "Chris, did you take all the oil pipes that were at the line?" Having confirmed that Chris has the oil pipes he explains why he thought that Chris had them: "I have seen the boxes standing there".*

Here, the visibility of situations and events on the shop floor leads to Mark being aware of where the parts in question are. The problem that the location of the parts was not accurately recorded in the information system was immediately compensated by his knowledge of the shopfloor situation. Likewise, Jim's knowledge of working practices leads him to call specific people who are likely to have the parts. Mark's observation provides him with a shortcut, making further telephone calls unnecessary.

(continued) *Since it was first established that parts were missing, production has moved on and there is the question what to do with the engines that are missing oil pipes. Jim and Mark discuss if the material structure of the engine allows them to be assembled in "stationary assembly".*

Workers in the plant are aware of the material properties of the engines produced and are thus able to relate the material artefact presented to them to the process of its construction. In the example above, Mark and Jim discuss this relationship in order to find out if the problem of missing oil pipes can be dealt with in stationary assembly, i.e., after the engines have left the assembly line. They have to attend to such issues as the proper order in which parts can be assembled as well as, for example, the physical orientation of the engine as some parts can only be assembled when the engine is positioned accordingly. Turning engines can only be done with heavy equipment available at the assembly line. The knowledge of the material properties of engines also allows members to detect troubles, i.e. the product itself affords checking of its proper progress through production (cf. Hartswood and Procter 2000). In her discussion of decisions and decision-making Boden (1994) suggests that classical theoretical treatments often confound our understanding of these organizational phenomena, suggesting instead that decision-making is located in fine-grained, sequential organisational activities. Of particular relevance is the notion of 'local logics': *"As they sift through locally relevant possibilities (…) social actors use their own agendas and understandings to produce 'answers' that are then fitted to 'questions'* (Boden 1994).

Conclusions

Our study shows how local logics are deployed to provide 'routine' but nevertheless skillful responses to both expected and unexpected 'normal natural troubles'. Underlying mainstream work on production planning is the notion of uniform and predictable prescriptions of activity. In contrast we document working practices that attend to the 'worldly contingencies' of production, the 'normal, natural' troubles whose 'usual' solution is readily available in and as part of everyday working practice. We document such problem solving 'from within' detailing the production of workarounds that may themselves become part of the repertoire of candidate solutions.

References

Boden, D. (1994) *The business of talk: organizations in action*. Cambridge : Polity Press, 1994

Dant T. and Francis, D. (1998) Planning in organisations: Rational control or contingent activity? *Sociological Research Online*, 3(2), 1998.

Hartswood, M. and Procter, R. (2000) Design guidelines for dealing with breakdowns and repairs in collaborative work settings. *International Journal of Human-Computer Studies*, 53:91–120, 2000.

Hughes, J., King, V., Rodden, T. and Andersen, H. (1995) The role of ethnography in interactive systems design. *Interactions*, pages 56–65, April 1995.

Randall, D., Hughes, J. A., O'Brien, J. and Rouncefield, M. (1996) Organisational memory and CSCW: supporting the 'Mavis' phenomenon. *Proceedings of OzCHI*, 1996.

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

Voß, A, Procter, R. and Williams, R. (2000) Innovation in use: Interleaving day-to-day operation and systems development. In Cherkasky, Greenbaum, Mambrey, Pors (eds.), *Proceedings of the Participatory Design Conference*, pages 192–201, 2000.

Acknowledgement: The research reported here is funded by the UK Engineering and Physical Sciences Research Council (award numbers 00304580 and GR/N 13999). We would like to thank the staff at the case study organisation for their help.

Modelling Collaborative Work in UML

Rachid Hourizi, Peter Johnson, Anne Bruseberg, Iya Solodilova,

Dept. of Computing Science, University of Bath, Claverton Down, Bath BA2 7AY, UK.
http://www.cs.bath.ac.uk/~flightdk

Abstract: Though the investigation and understanding of aviation accidents is a well-researched topic, the influence of that knowledge on current design tools has been less well explored. In this paper, we take a market standard design tool, the Universal Modelling Language, and examine the extent to which it can be used to describe the complexities of complex, collaborative work on the modern, glass cockpit flightdeck. We use a well-reported aviation accident as an example scenario for our study. We find that the model of collaboration and communication implicit in the UML is insufficiently detailed to allow a designer using the language, even to model many of the goal structures and communication problems, which would need to be incorporated in the design rationale of a safer system. We conclude, therefore, that a discrepancy exists between the requirements of the design community and the tools currently available to support them in their work.

Keywords: User Models, UML, Collaborative Systems, System Failure, Safety Critical Design.

Introduction

A great deal of effort has been put into the investigation of complex socio-technical system failures, particularly in the aviation domain. The resulting publications include numerous accident and incident reports (e.g. ATSB, 2001) alongside a large amount of research, which describes, categorises and ranks the importance of the causal elements observed during such failures (e.g. FAA, 1996, Hutchins, Holder & Hayward (1999)). As has been noted elsewhere (Shappel & Wiegemann, 2000) an increasing percentage of the failures investigated in this way are affected by human activity rather than mechanical failure (so called "human error") and in addition to the literature mentioned above, a number of frameworks have been developed, which start to describe the elements involved in both successful and undesirable human behaviour in these multi-agent, collaborative environments. In this paper, we will present our preliminary exploration of the extent to which this understanding of dynamic collaborative interaction has been explicitly supported in a popular design tool – the Unified Modelling Language (UML). We acknowledge that other tools are used (e.g. HAZOPS or THEA [Pocock, Harrison, Wright & Johnson (2001)]) and could, therefore have been discussed in the place of UML, but have chosen the market leader, not only for its widespread acceptance, but also for the notion that it can be used with any design process, in any domain (Fowler, M, Scott, K, (2000)). Our central research question is, therefore; to what extent has this increased understanding of the collaborative nature of this (and other similar) domains influenced the tools, like the UML, available to the designers charged with the creation of effective artefacts within them?

Case Study: Qantas QF1, 23rd September 1999, Bangkok.

In order to present an authentic set of domain problems within this paper, we have chosen a well reported accident which contains a wide range of collaborative relationships and communication breakdowns. Though limited space precludes a full description of our chosen accident, it can be summarised as follows: On 23 September 1999, in a heavy thunderstorm, a Qantas Boeing 747-438 aircraft, (scheduled flight QF1) came in to land at Bangkok International Airport, Thailand, with moderate flap settings, no reverse thrust planned and the autobrakes set to engage upon contact with the ground. Boeing manuals for the aircraft type showed the increased importance of reverse thrust on a wet runway, but the Qantas company standards had not incorporated this information, and the crew were unaware of it. To make matters worse, neither the Pilot Flying (First Officer) nor the Captain were aware of the changing severity of the airport weather conditions (Air Traffic Control failed to pass an updated weather report), despite the fact that another Qantas aircraft had aborted its landing there just moments before and the Second Officer had overheard the information on a company radio channel, but failed to pass the information to his colleagues. At the point of landing, the Pilot Flying (PF) had brought the plane in a little too high and a little too fast and the captain (the Pilot Not Flying or PNF) decided to order a go-around (i.e. to abort the landing attempt). As he gave this instruction, he reconsidered his decision. As the PF started to apply thrust (in order to execute the go-around), the Captain changed his mind and decided to land the plane.. In his haste to change the situation, the Captain (previously PNF) simply placed his hand on the (shared) thrust levers, over that of the PF and pulled backwards (reducing thrust). The PF later testified that this had caused some uncertainty as to who was actually in control of the plane. Unfortunately, the captain had not successfully reduced the thrust in all four engines. He had, inadvertently, pulled back on only three of the four thrust levers, leaving one engine fully engaged. This combination of contact with the ground and an engaged engine was read by the automated systems as a take off configuration – rendering the automatic application of the brakes inappropriate. Consequently, the auto-brakes disengaged. The plane was now aquaplaning down a wet runway with a moderate flap setting, no brakes engaged and forward, rather than reverse thrust! By now, the outcome of the landing was all but fixed. The crew did apply the brakes manually, but were unable to stop the plane travelling off the end of the runway and crashing into an antenna 217 meters beyond (FAA, 1996).

Modelling

From our description, we can see that the causal structure of the accident contains failed communication and collaboration between both individuals (e.g. PF, PNF, Second Officer) teams (e.g. QF1 crew, ATC) and organisations (e.g. Qantas, Boeing). Furthermore, we could argue that, beyond the human agents, the semi-autonomous automated systems also played an *active* role in these breakdowns (e.g. the autobrake levers, changed the state of the system, without explicit instruction from the crew, leaving the PF and PNF unaware of the change for a brief, but important, period of time). Thus far, the UML suite is sufficient to the task of modelling our *collaborative* behaviour, since the notion of actors contained within it, includes descriptions of individuals, groups, people and automation.

Beyond this point, however, a number of omissions begin to appear. The various collaborating actors, for example, do not enjoy a single, unified work relationship. In fact, their goals or objectives (generally modeled by the UML Use-Case Diagram) were linked in a number of different ways: Some goal pairings were of mutual benefit e.g. the *shared* goal of "land plane", held by the PF and PNF, the *complementary* goals of the ATC (manage airspace) and flight crew (fly plane), without which the higher level goal (land plane) could not have been achieved and the *dependant* goals of the PNF lowering the landing gear and the PF putting the aircraft on the ground. Equally, however, some goal combinations observed during the flight were mutually harmful e.g. the *conflicting* goals of the crew, trying to bring the plane to a halt and the automated system disengaging the autobrake setting and the *mutually exclusive* goals of the Captain trying to "land" and the FO trying to "go-around". An understanding of these, more complex relationships is vital if the designer is to support authentic flight deck interaction, but is not supported by the appropriate part of the standard modelling tool (UML Use Case Diagrams).

If we dig deeper into the varied relationships observed, we also find communications which are more complex than the simplified model inherent in the next set of UML models, (Sequence or Collaboration diagrams). In particular, the communication observed (roughly equivalent to UML messages) varies along a number of dimensions. For example, we can identify alternative message *sources* - some single (e.g. the Captain or First Officer), some involving multiple actors (e.g. the landing gear sensors and the engines were both needed to "inform" the automated systems that the aircraft was in take off configuration) and, arguably, some with no specific source at all (e.g. the "message" to the FO that the Captain had taken control had not been intentionally sent by the latter but rather inferred by the former). In the same way, we can identify alternative *destination* categories – some single, some multiple (e.g. the Captain, in pulling back on the thrust levers sent messages to both the automated systems and the FO) and some undefined (e.g. the autobrake sent a message generally into the cockpit, but did not identify a specific recipient). Lastly, we find various categories of message *failure* – some messages are *sent but not received* (e.g. the message from autobrakes to crew), some are *received but not sent* and some (such as the Captain's message to change from "go-around" to "land") are *altered in transit* (e.g. the First Officer received an instruction to change PF and the autobrakes received a message that the plane was in take off configuration). Standard UML contains only the notion of a single successful message passing from one origin to one destination, though it has been extended to include the notion of multiple, simultaneous messages originating from a single source. The rich complexity of inter-agent communication and the range of failures observed cannot be described by the UML models commonly used early stages of the design process. As argued in the introduction, above , we believe these omissions to be detrimental to the chances of optimal *collaborative artifacts* being created by those relying on the models concerned.

Conclusion

In this preliminary work we have shown, that, despite a claim of universal applicability, the UML contains, embedded within itself, limited notions of goal relationships, communication and failure. When we try to use the tool to model a complex, collaborative scenario, such as the air incident discussed, we find these limited notions to be insufficient for the task of describing the interwoven objectives and activity observed and, by extension the full causal structure of the breakdowns inherent in the example. In fact, we have shown that many of those breakdowns which seem to be central to the failure scenario described cannot currently be modelled within the language. We believe this omission in the tool to be a serious impediment to the production of practical, efficient collaborative systems and are concerned that the combination of its wide acceptance; and the increasing introduction of computerised agents into many complex domains (commerce, network management, aviation etc) make the need for either an extension to existing tools or the production of new ones all the more pressing.

In conclusion, therefore, the market leading modelling tool (the UML) is insufficient to describe our domain of interest. In this sense, we would argue that the increasing understanding that much of the failed interaction between people and automated systems can be described as unsuccessful collaboration has not yet filtered through to the popular design tools with which such systems are created. In the future, therefore, we need not only to refine our understanding of the collaborative process, but equally to extend and improve the tool set available to designers such that they are able accurately to model, understand and improve the both systems and collaboration in question.

Acknowledgements: This work is being funded by the EPSRC (grant number GR/R40739/01) and supported by QinetiQ and GKN Westland Helicopters.

References:

1. ATSB (2001). Investigation Report 1999045538. Boeing 747-438, VH-OJH, Bangkok, Thailand, 23 Sep 1999.
2. FAA (1996). Federal Aviation Administration Human Factors Team Report on: The Interfaces between Flightcrews and Modern Flight Deck Systems. Washington, DC: FAA.
3. Fowler, M, Scott, K, (2000) UML Distilled: A Brief Guide to the Standard Object Modelling Language, Addison Wesley, Boston, US.
4. Hutchins, E., Holder, B., Hayward, M. (1999). Pilot Attitudes towards Automation. URL: http://hci.ucsd.edu/hutchins/attitudes/attitudes.pdf.
5. Pocock, S., Harrison, M., Wright, P & Johnson, P. (2001) THEA: a technique for human error assessment early in design. Proceedings of Interact (2001) 247-254.
6. Sarter, N. B., Woods, D. D. (1995). How In The World Did We Ever Get Into That Mode? Mode Error and Awareness in Supervisory Control. Human Factors (37) 5-19.
7. Shappell S.A., Wiegmann D.A. (2000). Human Factors Analysis and Classification System—HFACS. Office of Aviation Medicine, FAA, Washington, DC; 2000.
8. Stevens, P, Pooley, R.J., (1999), Using UML: Software Engineering with Objects and Components, Addison Wesley, Boston, US.

Organisational Improvisation: A Field Study At a Swedish NPP during a Productive-Outage

Vincent Gauthereau[1] & Erik Hollnagel[2]

Quality management, Department of Industrial Engineering, University of Linköping, Sweden.
vinga@ikp.liu.se
CSELAB, Department of Computer and Information Science, University of Linköping, Sweden.
eriho@ida.liu.se

Abstract: This article relates early findings part of an ethnographic study at a Swedish Nuclear Power Plant during a plant shut down for a short non-productive outage. While improvisation usually is seen as a positive quality, it may look less attractive in a high hazard industry. Improvisation may nevertheless be one way to achieve resilience. Instead of seeing improvisation as an occasional characteristic, the set of events reported illustrates the improvisational nature of organising. It shows that there is no clear-cut distinction between what is improvisation and what is not. Understanding this side of organising is essential if we are to fully understand how the "conflict" between centralisation and decentralisation is resolved by high reliability organisations.

Keywords: Improvisation, planning, maintenance, nuclear power plant

Introduction

Studies of work practices have often highlighted the improvisational character of actions at the "sharp end" where physical and temporal constraints force individuals to depart from prescribed procedures (e.g. Keller & Keller, 1993; Leplat, 1989). While there is no dispute about the improvisational nature of individual actions, the issue is more contentious when it comes to organisations. The focus is here often on the way in which an organisation structures and regulates work; indeed, the word "organisation" is itself often equated with structure. Some theorists have insisted on changing the discourse of organisational studies from "organisation" to "organising" (e.g. Weick, 1979), only later to focus their attention on the improvisational characters of organisations (Weick, 1993; Weick, 1998). Recently, several studies have tried to understand how organisations improvise (Crossan, 1998; Miner et al., 2001; Moorman & Miner, 1998; Pasmore, 1998), especially in relation to product development activities. Improvisation in these contexts is often understood as a positive quality, although it does not directly correlate with concepts such as innovation or creativity (Moorman & Miner, 1998).

When it comes to high hazard industries improvisation may look less attractive, since safety must be planned rather than left to serendipitous actions! Improvisation is often caused by uncertainty, which is clearly an unwelcome aspect of safety. On the other hand, resilience is a highly praised characteristic and a good organisation should be able to adapt to unexpected variability (Gauthereau et al., 2001; Roberts, 1993; Weick, 1993). Theorists working with High Reliability organisations (HRO) have emphasised that adaptation to a changing environment is a major quality of a HRO (e.g. La Porte & Consolini, 1991; Rochlin, 1989; Rochlin et al., 1987). However, these studies have too often focused on exceptional circumstances, on "cosmological events" as Karl Weick puts it, where a danger is 'identifiable' (Hutchins, 1995, chap 8; Weick, 1993). Just as the naturalistic approach to the study of work show the importance of improvisation in everyday situations, this paper argues that the nature of organising is improvisational as well.

Empirical Material

The findings reported in this paper come from a larger study of operational readiness verification in a Swedish Nuclear Power Plant (NPP), i.e., the procedures for testing before restart after a period of maintenance (Hollnagel & Gauthereau, 2001). The events were observed at one of the three units of the NPP. This study mixed participant observations with informal and formal interviews. The present analysis is based on a subset of the collected data, which covered a broader range of issues important for operational readiness verification.

The NPP under study undergoes a so-called productive outage (PO) four times a year. Technically, the safety systems of the unit are divided into four independent arrays, which can be made inoperative one by one thereby allowing maintenance to take place while still producing (thus the name "productive outage").

The first day of this PO, important traces of contamination in the primary cooling system were found which suggested an “urgent” need to replace the damaged fuel, hence a need for a non-productive short-outage (SO). The data analysed in this work concern the planning and completion of this SO.

Discussion

Three Weeks / Three Phases: This three-week period in the life of the unit showed three quite different types of organisational behaviour. The first week could be described as chaotic. The work in itself is, of course, not chaotic: work is performed in an orderly manner, with method, and the planning of the work to be performed during the SO was done as seriously as possible. However, the level of uncertainty seemed to worry the employees. While the employees readily acknowledge that only approximately one third of the jobs to be performed during the SO can be planned long in advance and while they usually can handle such a level of uncertainty, the lack of clarity of the situation to come (or at least the lack of communication about it) created a rather noticeable concern. The end of this first phase set the rough structure for the SO to come and while the content was still unclear, boundaries were defined. The second week differed from the first in the sense that the employees focused their attention on planning and on preparing the upcoming events in more detail. Preparing this rather exceptional event in the life of the plant had become quite of a routine work. People fell back on routines learned (and mastered?) during previous outage planning. Even the major documents (work orders) were borrowed from earlier outages. The third week was about “following” the plan; more concretely, it was about adapting it to the circumstances.

The Meaning of Improvisation: In the literature, the construct of improvisation is often used in a narrow sense. For instance, Miner et al. (2001) propose that improvisation presents four different features: material convergence of design and execution, temporal convergence, novelty, and that finally the whole process should be deliberate. The different degrees of improvisation proposed by Weick (1998, p. 544-546) seem to confirm the rarity of improvisational acts: improvisation is seen on a continuum that ranges from simple “interpretation” to “improvisation” through “embellishment” and “variation”. It seems that the concept of improvisation has been over-specified, and although the definitions found in the literature might be relevant to the contexts in questions (e.g. product development), they do not seem relevant for the study at hand. In fact, most of the definitions would restrain our focus to the third week at the unit (that is, when the SO actually took place) by the requirement that improvisation is the convergence of design and execution. Since using such a focus would hide important parts of the process it is preferable not to be overly constrained by terminological disputes. Rather than trying to fit the observed events into a previously defined frame, we shall further analyse them and explain why we understand the concept of improvisation as suitable for describing them.

Planning And Improvisation: The essence of planning is prediction. A planned action is one where control is carried out by feedforward: the anticipation of the outcome directs the action. In our case, although the staff’s experience with SO was limited, their experience from similar events, especially from refuelling outages, was relevant for the planning. However, in order to build on this experience the staff needed to recognise the situation that was to be dealt with. Without a proper frame of reference, the goal of performing a SO is too under-specified to allow the staff to plan. Defining the start and the end of the SO provided the staff with a structure for their future actions. This need for a structure has also been found by studies of improvisation in jazz (e.g. Barrett, 1998; Mirvis, 1998): the basic structure of the song permits each player to anticipate the actions of others, not in a precise manner but in a way that allows coordination. While the guiding structures should allow some flexibility, they should also be non-negotiable (Barrett, 1998, p611) in the sense that they cannot be changed at will. These basic structures are a prerequisite for improvisation as they allow “everyone knows where everyone else is supposed to be” (Ibid, p. 612). The lack of a minimal structure, as observed during the first week, consequently clearly left the staff in a rather uncomfortable situation.

Despite that, we observed individuals performing their tasks. Once again the jazz metaphor seems useful: while preparing themselves for a session, musicians do not need to know which songs will be performed in order to perform certain task. Instruments need to be tuned quite independently of what is going to be played. Moreover, for the persons directly concerned with planning activities, preliminary schedules could be seen as design prototypes or as a minimal structure that “*provided imaginative boundaries around which they could explore options*” (Ibid, p. 612). Once this minimal structure has been defined, improvisation can theoretically begin. However, in our case the organisation had one more week to prepare itself for the event, which enabled more detailed planning. An interesting observation during this week was that some persons were eager to start the SO as soon as possible. One person even stated that an unplanned SO would be easier. In fact, everybody seemed to acknowledge that no matter how carefully an SO was planned there would always be new tasks to be performed, which would only be discovered when the plant was in a non-productive state. The preparations during this second week were more about further defining the structure, than careful planning of the content.

During the third week, i.e., during the SO itself, there were no indications that employees purposefully deviated from the plan. On the contrary, the plan was followed as closely as possible. When deviations occurred, they were goal-oriented adjustments forced by external factors, which did not change the structure.

Managing The Centralisation-Decentralisation Conflict: Since Perrow's (1984) Normal Accident Theory, the literature concerned with safety in organisation has often looked for solutions that manage both centralisation (as a solution to the problem of tight-coupling) and decentralisation (as a solution to the problem of complexity). A vital characteristic of HROs is their capability to maintain both (Rochlin, 1989; Weick, 1987). Yet it has as also been noticed that people in such organisations usually do not concern themselves with this centralisation-decentralisation conflict (Rochlin, 1999). The findings in the present study could be interpreted as a successful management of the two opposite: a central planning on the one hand providing coordination between individuals, and a local improvisation of individuals actions on the other hand. Yet for the people involved the situation was not one of carefully balancing the two opposites, but rather one of going through a quite unproblematic process of planning and improvisation. Improvisation is often defined by the degree to which planning and execution converge in time (Miner et al., 2001; Moorman & Miner, 1998). This brings to mind Schützenberger's (1954) distinction between strategy and tactics. According to this the difference is that tactics does not take into account the whole of the situation, but proceeds according to a criterion of local optimality. What was observed at the plant could thus be defined as successful tactical planning.

What needs to be studied is how people decide when to improvise and when not to, i.e., how they trade off the simple efficiency of following a plan with the improved thoroughness of making local adjustments. It is this, more than anything else that in practice determines whether the organisation will be robust and resilient.

References

Barrett, F. J. (1998). Creativity and Improvisation in jazz and Organizations: Implications for Organizational Learning. *Organization Science* **9**(5): 605-622.

Crossan, M. M. (1998). Improvisation in Action. *Organization Science* **9**(5): 593-599.

Gauthereau, V., E. Hollnagel & B. Bergman (2001). Managing Variability - A strategy for creating safety in organisations? *Arbete Människa Miljö & Nordisk Ergonomi* **2001**(1): 25-40.

Hollnagel, E. & V. Gauthereau (2001). Orperational Readiness Verification: A study of Safety During Outage and Restart of Nuclear Power Plants. Stockholm, Statens Kärnkrafts Inspektionen (SKI).

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, Mass., MIT Press.

Keller, C. & J. D. Keller (1993). Thinking and acting with iron. *Understanding practice: Perspectives on activity and context*. S. Chaiklin and J. Lave. Cambridge, UK, Cambridge University Press**:** 125-143.

La Porte, T. R. & P. M. Consolini (1991). Working in Practice but Not in Theory: Theoretical Challenges of 'High-Reliability Organizations'. *Journal of Public Administration Research and Theory* **January**: 19-47.

Leplat, J. (1989). Error Analysis, instrument and object of task analysis. *Ergonomics* **32**(7): 813-822.

Miner, A. S., P. Bassoff & C. Moorman (2001). Organizational Improvisation and Learning: A Field Study. *Administrative Science Quarterly* **46**: 304-337.

Mirvis, P. H. (1998). Practice Improvisation. *Organization Science* **9**(5): 586-592.

Moorman, C. & A. S. Miner (1998). Organizational Improvisation and Organizational Memory. *Academy of Management Review* **23**(4): 698-723.

Pasmore, W. A. (1998). Organizing for Jazz. *Organization Science* **9**(5): 562-564.

Perrow, C. (1984). *Normal Accidents: Living With High-Risk Technologies*. New-York, USA, Basic Books, Inc.

Roberts, K. H., Ed. (1993). *New challenges to understanding organizations*. New York, Maxwell Macmillan International.

Rochlin, G. I. (1989). Informal organizational networking as a crisis-avoidance strategy: US naval flight operations as a case study. *Industrial Crisis Quarterly* **3**(2): 159-176.

Rochlin, G. I. (1999). Safe operation as a social construct. *Ergonomics* **42**(00).

Rochlin, G. I., T. La Porte, R , & K. H. Roberts (1987). The Self-Designing High-Reliability Organization: Aircraft Carrier Flight Operations at Sea. *Naval War College Review* **Autumn**: 76-90.

Schützenberger, M. P. (1954). A tentative classification of goal-seeking behaviours. *Journal of mental science*, 100, p. 97-102

Weick, K. E. (1979). *The Social Psychology of Organizing*, McGraw-Hill, Inc.

Weick, K. E. (1987). Organizational Culture as a Source of High Reliability. *California Management Review* **29**(2): 112-127.

Weick, K. E. (1993). The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster. *Administrative Science Quarterly* **38**(December): 628-652.

Weick, K. E. (1998). Improvisation as a Mindset for Organizational Analysis. *Organization Science* **9**(5): 543-555.

Centralised vs. Distributed Alarm Handling

Kenneth Gulbrandsøy and Magnus Reistad

Dep. of Engeneering Cybernetics, Norwegian University og Science and Technology,
Odd Bragstads plass 2D, 7491 Trondheim, NORWAY
kennetgu@ itk.ntnu.no, magnus.reistad@itk.ntnu.no

Abstract: This paper is a short version of a full paper with same name (see [7]) and discuss the possible method of distributed alarm handling (DAH) in contrast to centralized alarm-handling (CAH), that is preformed to day. It is shown that DAH does not reduce the system reliability compared to CAH. Depending on the design of the DAH system it is possible to improve system reliability with at least 5% (very conservative design). A general routing algorithm is presented and some problems regarding possible implementations are discussed. The conclusion is that DAH systems are feasible, but some big issues must be solved before the large potential can be utilized.

Keywords: alarm distribution, alarm handling.

Introduction

With increasing focus on hand-held devices (PDA's) and portable computers in industry applications, alarm handling is an area that could benefit for this. Alarm handling today is in general done from a central control room where one or more operators monitor and make corrective actions to signalled alarms. The aim for this paper is to evaluate an alternative organisation of the alarm handling system: distribute alarm handling to control room operator and operators in the plant. Many problems connected to this method must be addressed and solved. Most of them are connected to human factors and the technical solutions available today. Interaction between humans and hand-held devices is limited to screen size, methods for data input, data connectivity, ability to function under extreme environmental conditions and so on. This implies that special considerations must be taken before individual tasks, e.g. alarm handling, can be preformed on these devices. A central question is whether distributed alarm-handling increases or decreases the system reliability. This must, for obvious reasons, be evaluated before further work on system design is carried out.

Operator Properties

When working with assignments of tasks to operators, an important property must be declared: availability. Operator availability depends on several other operator properties, like tasks the operator is responsible for, work location in the plant, workspace and channel of communication. Number of tasks and task properties results in different levels of cognitive load, it influences the operators ability to multitask and to be attentive. Some tasks can only be performed at certain locations in the plant. Combined with work location (desk, terminal etc.) it may result in some operators becoming more available than others. Which communication channels alarms are assigned through is also important. There are roughly three practical channels: Sight, hearing and touch. Screens, light and sound are often used to alert the operator of a new alarm. In some noisy environments where screens are unavailable, vibration is a good way to alert the operator. Which channels to use, depends on the environment the operator operate in. Operator properties must be accounted for when assigning alarms to operators. For any given alarm one or more operators in the plant are going to be more available to handle the alarm than other operators. In systems where alarms can be assigned to a number of different operators the system must be operator-aware. This means that the system must know about each operator's status at any given time.

System Reliability

Alarm handling systems primary goal is to prevent damage to people, plant and financial loss. This means that DAH systems must have advantages that CAH systems today do not have, if they are to be implemented. DAH systems must also have the same or better reliability then CAH systems have today. A good alarm handling system must be designed in such a way that it compensates for human weakness and utilize human's strong points. A fault-tree analysis (FTA) is used to estimate system reliability for CAH and DAH. FTA has several weak points; dynamic behaviors and common errors are difficult to model.

FTA is still often used because it can model redundancy in a simple way. The main component in FTA is events. An event may either happen or not happen. This means that a probability can be attached to each event in the tree. In the FTA analysis each event is defined as an error. The probability of a particular event occurs is equivalent to the probability for a particular error. If e is the probability for an error, then reliability is defined as r = 1 - e.

System definition: For both CAH and DAH the top-action is defined as "System fail to handle alarm correct". Reliability is then defined as "System handle alarm correct". The different alarm categories and number of utilized operators make the reliability dynamic as a function of alarm category and number of operators. There are many more possible variables that influence on the reliability. Some of these are task priorities vs. alarm priority, constraints in communication between operators. It is possible to find many more, but that is left out for later and deeper studies. CAH and DAH fault-trees (see [7]) are based on the human and system factors that influence the system reliability.

Qualitative analysis: Qualitative analysis' main goal is to identify minimal spanning trees in the primary fault-tree. If the top-event in a minimal spanning tree occurs, the top-event in the fault-tree will also occur. This means that the number of minimal spanning trees and system reliability are closely connected (increasing number, decreasing reliability). CAH fault-tree fewer minimal spanning trees than DAH fault-tree. This doesn't automatically mean that DAH is more unreliable then CAH. Some of the additional error-events in the DAH fault-tree get its effect reduced with increased numbers of redundant operators. An operator becomes redundant when he can handle the same alarms that other operators do. The significance of this redundancy is dependent on the number of redundant operators. A complete picture cannot be presented before a quantitative analysis is preformed.

Quantitative analysis: The quantitative analysis is based on [3] and uses theory that depends on reliability diagrams. Reliability diagrams are equivalent with fault-trees. Before the calculation of reliability can start, assumptions must be taken (see [6]). These assumptions will reduce the results usability, but it is impossible to do such an analysis without some assumptions. The degree of cognitive load will affect the system reliability (large load means stressed operators). Calculation of reliability is therefore divided into four alarm categories that correspond to four different levels of cognitive load: location-based, skill-based, rule-based and knowledge-based. The biggest challenge is to find good and applicable human error probability data and data on probability distributions related to operator overload as function of alarm category and number of active tasks. Most of the data is derived from [4]. Since the goal is to identify the difference between CAH and DAH conservative estimates of probabilities connected to the different events are enough. The difference will in any case be preserved. The quantitative analysis has shown that DAH systems can improve system reliability with 5 - 50% depending on number of task each operator has, task priorities (larger or smaller than the priority of alarm handling), number of redundant operators and on other assumptions made. It also shows that DAH system degrades to a CAH system when number of redundant operators is less than 2. A central difference between CAH and DAH is that the DAH system is able to handle larger number of alarms with same degree of cognitive load on each operator then CAH system can. A CAH system has only one (or two) operators. DAHs have *n* operators. If the DAH system balance cognitive loads evenly, and the same alarms are delivered to CAH and DAH system, then each operator in the DAH system will have 1 / *n* of the load the single operator in the CAH system has. Results also show that configuration of the DAH system is essential for efficient use of operators. Large number of redundant operators (handles the same alarms) will reduce throughput of alarms and therefore overall system efficiency.

System Design

The reliability analysis has shown that DAH does not reduce system reliability compared to CAH. It is now reasonable to develop a general system design for distributed alarm handling. The main task in a DAH system is to find an optimal allocation of alarms to operators. This implies that operator availability must be evaluated in some way. For the time being, let this be a conceptual function. When a big process disturbance occurs in a DAH system must probably handle a large number of alarms. Alarms with time constraints on when an alarm must be handled, imply demands on the maximal alarm throughput the system can produce. The DAH system routing algorithm must therefore be designed for both speed and reliability. There are many ways to achieve high throughput. In addition to efficient algorithm design, another way is to reduce the amount of alarms to evaluate for assignment to operator. This can be done with techniques already used today, like alarm suppression or shelving, or by introducing a new concept. Instead of routing every raised alarm by evaluating operator availability and picking the most available operator, it is possible to route the underlying problem. Each problem is a conceptual container for the group of alarms that it raises directly or indirectly. Since the problem is a conceptual entity, the designer is free to choose any problem definition with corresponding alarm group. The benefit with this method is that when a problem is assigned to a operator (problem is active), new alarms belonging to the routed problem can be routed directly without finding the optimal operator (which is an expensive operation).

It is indicated in [6] that DAH system reliability is not worse than CAH system reliability. This makes the study of DAH system reasonable. A DAH system is possible to design an implement, but several constraints today are reducing its feasibility. Some economical aspects, like number of employees needed, also reduce the build-in-power of DAH systems because existing operators must be used to prevent increased staffing costs. This means that the DAH system efficiency is reduced. Before highly efficient DAH system can be developed several new developments with regard to information presentation (HCI) and system diagnostic (detection problems) must done.

Acknowledgements
This work is supported by the Norwegian Research Council and is part of the research programme, KIKS (Wearable Information and Communication System) at Department of Engineering Cybernetics, NTNU. Support to this research is also given from The PetroHammlab project at Institute for energy technology in Halden.

References

[1] Jenny Preece, Yvonne Rogers, Hellen Sharp et al (1994). Human - Computer Interaction, Addison – Westley, ISBN 0-201-62769-8.

[2] C. R Timms (1999). A Methodology For Alarm Classification And Prioritization, Shell U.K. Exploration and Production.

[3] Terje Aven (1994). Reliability and risk analysis, 2. edition, Universitetsforlaget, Norway.

[4] John Charles Cluley (1993). Reliability in Instrumentation and Control, Butterworth-Heinemann.

[5] David I. Gertman (1994). Human Reliability & Safety Analysis Data Handbook.

[6] Kenneth Gulbrandsøy (2002). Wearable control room: Test and design of distributed alarm handling. Available at http://www.itk.ntnu.no/cgi-bin/databaser/hovedoppgaver/arkiv.cgi after September 2002. Before this, contact author at kennetgu@itk.ntnu.no.

[7] Kenneth Gulbrandsøy (2002). Distributed vs. Centralized Alarm Handling (full paper) http://www.dcs.gla.ac.uk/~johnson/eam2002/

Is Overcoming of Fixation Possible?

Machteld Van der Vlugt, Peter A. Wieringa,

Delft University of Technology, Dept. of Design, Engineering & Production, Human-machine Systems,
Mekelweg 2, 2628 CD Delft, The Netherlands
m.vandervlugt@wbmt.tudelft.nl

Abstract Fixation is described as the tendency of people to stick to an initial plan whereas an alternative plan should be taken because of the changed situation. Strategies proposed, like improved training methods and better systems design, aim at preventing the recurrence of fixation. However, the possibility of the occurrence of fixation is not ruled out by following these strategies. Interventions that aim at overcoming fixation in real time seem a welcome addition to the existing strategies. However, to intervene in fixation a theory about fixation is necessary. So far the leading theory, which explains fixation by the limited cognitive processes of humans does not satisfy. Watzlawick's theory could give new opportunities for overcoming fixation and therefore, this theory is brought to the attention of the human-machine systems community.

Difficulties Overcoming Fixation

Several studies (among which Moray & Rotenberg (1989), De Keyser & Woods (1990), Xiao & Mackenzie (1995)) noted a "difficult to influence" type of behavior, which is characterized as fixation (similar terms that refer to fixation are given in Table 7). In accordance with Cook and Woods (1994), fixation is defined as failing to revise plans or diagnoses in response to new evidence indicating these should be changed. Thus, Cook and Woods only speak of fixation when advice, alarms or other signals fail to catch someone's attention and change his plans accordingly. De Keyser and Woods (1990) distinguished three characteristic behavior patterns for fixation, which are given in Table 8.

Table 7: Synonyms of fixation found in examined studies

Synonyms of fixation	Literature
Cognitive lockup	Moray & Rotenberg (1989); Cook & Woods (1994); Xiao & Mackenzie (1995)
Error/Target Fascination	Boer (1998)
Functional Fixedness	Anderson (1980)
Loss of Situation Awareness	Cook & Woods (1994); Endsley (1995)
Opportunistic Control	Hollnagel (1993)
Preoccupation	NTSB accident reports; Boer (1998)
Task Forgetting	Boer (1998); Kerstholt & Passenier (1995)
Tunnel Vision	Cook & Woods (1994); Moray & Rotenberg (1989), Kerstholt & Passenier (1995)

Table 8: Classification of fixation patterns with their characteristics (from: De Keyser & Woods (1990))

	Behavior pattern	Characteristics
1	"This and nothing else"	Uncoordinated jumping to and fro between several actions
2	"Everything but this"	Repeating the same actions
3	"Everything is OK"	Taking no actions at all

Fixation is considered a serious phenomenon in the human-machine area, because of the possible severe consequences; in some aviation accidents the cause was traced back to fixation of the pilot (see for example the NTSB-AAR-73-14, 79-7 report). De Keyser and Woods (1990) found that overcoming fixation was hard to realize within the available possibilities. The fixated operators, they observed, rejected advice and low alarm signals (De Keyser & Woods (1990)).

Instead of overcoming fixation, the proposed strategies, like training programs and alarm systems, aim to prevent the occurrence of fixation. The training programs prepare the crew by anticipating situations in which humans are prone to fixation. The existing alarm systems, indicating a problem in a system state, aim at getting the operator's (or the pilot's) attention; they do not aim at overcoming the operator's fixation.

The objective of this study is to find ways to intervene in fixation as a supplemental safety net. Therefore, complementary to the alarm systems that annunciate problematic system states, a "system" capable of indicating and (ideally) overcoming the operator's fixation is necessary. After all, when a fixated pilot is facing a serious situation of which the outcome could be severe, preventing the occurrence of fixation for a future flight is not an option at that moment. New insights for overcoming fixation in real time call for a theory about behavior that includes fixation, giving starting points for intervening successfully in fixation in real time.

Why Has "Overcoming Fixation" Not Considered So Far?

The examined studies (Table 7) about fixation deal with the situation in which fixation may occur instead of with overcoming fixation. All strategies aim to prevent the recurrence of fixation by anticipating situations that are likely to provoke fixation, but when fixation does occur the means available do not succeed in overcoming fixation. Why?

The prevailing theory about fixation explains fixation by the limited cognitive processes of humans (among which Moray (1989), Cook & Woods (1994) and Endsley (1995)). This explanation stems from the generally accepted theory that sees human decision making as a part of information processing. The examined studies that accepted this theory, determined a strong correlation between the occurrence of fixation and the complexity of dynamic situations. The more complex and demanding the situation the more likely fixation was to occur. This has led to the conclusion that in complex dynamic situations limited cognitive processing 'causes' fixation.

Because the cognitive processes of humans are limited, factors influencing the complexity of the situation (see Table 9) are used to prevent a recurrence of fixation. Yet, these factors are largely uncontrollable due to their nature and, when fixation has occurred already, reconditioning some of these factors, if possible at all, does not guarantee overcoming fixation. Moreover, the theory of limited cognitive processes does not explain why fixation might occur anyway when the situation is not complex (see Boer (1998) and Anderson (1980)). Complex dynamic situations are not decisive for the occurrence of fixation.

Table 9: Correlations found in several studies between fixation and situation

Reference	Factors Influencing Complexity Situation
Moray & Rotenberg (1989)	Uncertainty information, mental model
De Keyser & Woods (1990)	Time pressure
Hollnagel (1993)	Number of goals/tasks and time pressure
Cook & Woods (1994)	Complexity system, time pressure
Kerstholt & Passenier (1995)	Extent of experience
Xiao & Mackenzie (1995)	Complexity system, uncertainty information

Fixation Is Context Dependent

Adding 'context' as a connecting pattern to the existing views about fixation may lead to opportunities to overcome fixation in real time. To determine fixation, the current approaches use the three behavior patterns given in Table 8. However, these patterns are possibly a necessary but not sufficient condition for determining fixation. Whether behavior is judged as fixation or not, is decided by the observer's context, an excellent strategy in one context can be the worst in another (De Keyser & Woods, 1990), although the behavior pattern is the same. When, for example, the situation is normal and someone shows "everything is OK" behavior, this behavior will not be identified as fixation. On the other hand, if the situation is serious and actions are required this same behavior will then be identified as fixation. Thus, for the determination of fixation, the observer's context is decisive. Just as the observer's context is decisive for his identification or judgment of the subject's behavior, the actions of the subject are decided by the subject's context. Yet, the context of the subject differs from the observer's context.

A System Theory About Behavior, Influence, and Meaning: To overcome fixation, the system theory (as developed by Watzlawick and colleagues (1967)) propagates that the observer must understand the

subject's context, although this context could be seen as undesirable to the observer. The starting point for understanding the subject's context is that the subject's behavior (in our case fixation) is interpreted as meaningful by the observer. This understanding is achieved via discovering the subject's point of view.
To make fixation understandable for the moment, it is convenient to explain fixation in terms of three of the pillars of Watzlawick's system theory. The first one is based on the axiom "one cannot not behave" of Watzlawick (1967). Within the system theory behavior is defined as only existing in relation with others. Behavior implies influence, one's comings and goings always affect other people, or following Watzlawick's axiom "one cannot not influence". The second important pillar is that one cannot back out of influence. Influence is a never-ending loop, there is no begin and there is no end. The effect of influence is that humans always attach meaning (intended and unintended) to everything and can be aware of doing so. The context influences the attached meaning, which manifests itself in behavior. The third and most important pillar is, that people who do not realize the effect of their behavior and on their behavior cannot change their behavior or in other words, only when people realize the effect of influence they can change behavior.

Analysis of Fixation from a System Theoretical Point of View
The effect of influence is reflected in behavior, and thus in fixation, and depends on the point of view of the person affected. From the subject's point of view his fixation is meaningful, whereas this usually is not the case from the observer's point of view. What makes the difference? A system theoretical hypothesis is that the fixated subject does not realize the effect of his behavior (= influence). The observer, on the other hand, does realize the effect of the subject's behavior. In case of fixation, the subject is guided (= influenced) by his context in how to act in this situation; from the subject's point of view he acts as he ought to do. However, the subject interprets the effect of his behavior as if it has no effect. The subject does not realize the effects, and therefore cannot change his behavior.
Watzlawick's point of view about fixation stemming from "human-human systems" could offer a new lead for overcoming fixation. Several adaptations have to be made though, before this theory can successfully be applied within the area of human-machine systems. There are two important differences between human-machine systems and the human-human systems. The first difference is that the intervention methods are based on human-human interaction whereas in human-machine systems the interaction is mediated by technical systems. The second important difference is the time span in which interventions take place. Within human-machine systems the time is much shorter (from minutes to hours depending on the system) than in human-human systems (from days to weeks and even years). Despite these differences, the basis of the theory gives good starting-points for a new approach to find ways for overcoming fixation.

Reference List

Anderson, J.R. (1980) Cognitive Psychology and its Implications. Worth Publishers, New York.

Boer, L.C. (1998) Cockpit Management: Poor Task Situation Awareness despite Moderate Workload. pp.1-9. Vienna. Proceedings of the 23rd Conference of the EAAP.

Cook, R.I. and Woods, D.D. (1994) Operating at the Sharp End: the Complexity of Human Error. In: Human Error in Medicine, pp. 255-310. Ed.: S.M. Bogner. Lawrence Erlbaum Associates.

De Keyser, V. and Woods, D.D. (1990) Fixation Errors: Failures to Revise Situation Assessment in Dynamic and Risky Systems. In: Systems Reliability Assessment, pp. 231-252. Eds: A.G. Colombo & A.S. d. Bustamante. Kluwer Academic Publishers, Dordrecht.

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

Hollnagel, E. (1993) Human Reliability Analysis: Context and Control. Academic Press, London.

Kerstholt, J.H. and Passenier, P.O. (1995) Can Cognitive Lockup in Ship Operation be Overcome by an Integrated Training? TNO-TM 1995 C-50. TNO, Soesterberg.

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

NTSB (1979). Aircraft Accident Report –United Airlines Inc., Mcdonnell-Douglas DC-8-61, N8082U, Portland. NTSB-AAR-79-7, Washington-DC, National Transportation Safety Board.

Watzlawick, P., Beavin, J.H., and Jackson, D.D. (1967) Pragmatics of Human Communications. Norton & Company, Inc, New York.

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

Supporting Distributed Planning in a Dynamic Environment: An Observational Study in Operating Room Management

Jos de Visser, Peter A. Wieringa (1), Jacqueline Moss, Yan Xiao (2)

(1) Man-Machine Systems, Delft University of Technology, Delft, The Netherlands.
j.devisser@wbmt.tudelft.nl, p.a.wieringa@wbmt.tudelft.nl
(2) Human Factors Research Group, University of Maryland, Baltimore, USA
jmoss001@umaryland.edu, yxiao@umaryland.edu

Abstract: Planning in collaborative work is usually distributed among workers, with access to different types of information that make decisions through negotiations. In dynamic work settings, unique changes arise for distributed planning. In this paper, we present a study of activities in coordinating surgical schedules in an operating room suite affiliated with a busy trauma center. Implications for an improved design of a support system for distributed scheduling tasks are discussed.

Keywords: coordination, planning board, distributed planning, observational study.

Introduction

Distributed planning is a widely occurring phenomenon in collaborating activities, where plans of each collaborator are shared and negotiated (Roth et al., 1999). A coordinator is often responsible for the smooth running of these activities and monitors and informs the collaborators. Coordination tools, like large planning boards, help distributed planning (Berndtsson, 1999). Our study focuses on how a planning board was used in managing changing plans in an event-driven, constantly changing work environment. Our goal was to establish a basis for building supporting tools, with visual aids for distributed workers, to help improve efficiency and reliability of distributed planning.

Setting

The study was conducted in the six operating room (OR) suite of a trauma center. On average 22 surgical cases were performed in any weekday. Certain operating rooms were dedicated to specific types of surgery, such as orthopaedic and emergency surgery. Each afternoon, a hardcopy of lists of surgical cases was delivered to the coordinator of the OR suite. Although every case had an expected duration, only the first case of an OR was assigned to a specific starting time, the rest for that OR was "to follow". The actual case sequence, case duration, and used OR was usually different from the delivered schedule and was determined by the combination of patient condition, progress of other cases, order of request and surgeon preference. In addition to the cases posted the day before, urgent cases were requested on the day of surgery. In comparison to other types of OR's, are here many of the decisions on case scheduling made on the day of surgery and more dynamic.

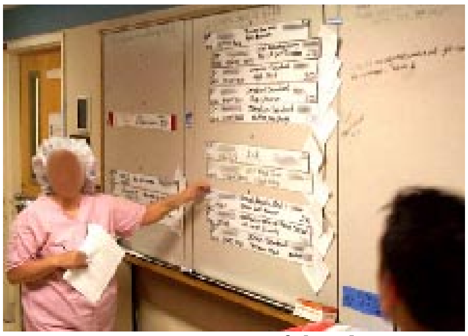


Figure 1 – Overview of the planning board.

A large whiteboard (12 x 4 ft, see figure 1) functioning as a planning board was placed adjacent to the suite where many collaborators came. The coordinator, a nurse, was in charge of updating the board. This planning board showed sequences and progress of the cases for each room. A detailed study of this planning board was reported in (Xiao et al., 2001).

Observations and data analysis

A research observer shadowed three different coordinators for nine days. The collected data include the communications concerning determination of surgery start times between the coordinator and the collaborators. To get better understanding of these communications, semi-structured interviews were conducted. The data contained also the representation of surgical cases on the planning board, including all

the changes occurring during the day, and the expected start time and duration of the cases and the actual starting time and ending time of the cases.
The first step in the data analysis was to construct a sequence of events and activities to reflect as much as possible the observed data. Diagrams were designed and produced to reflect changes in schedules and timing of actual case starts and endings.

Results
The observation of scheduling activities is used here to illustrate the graphical representation in figure 2. The activities associated with two (out of six) OR's are represented in this figure (OR4 and OR6). In this chart, the upper lines represent the actual start- and end time of a surgery. The two bars below the lines (strip #1 and strip #2) represent the schedules for the first two cases planned for the corresponding OR's. The non-planned bar (bottom of the figure) represents a strip on the planning board that is posted but not scheduled, and is put aside until there is more information about the patient, surgeon or timeslot in the OR's. The patient names (aliases) are inside the bars or above the lines. The actual durations are above the lines while the expected durations are inside the bars. The case of Roy in OR 4, for instance, was scheduled for 2 ½ hours, but actually it took only 1 hour and 9 minutes.

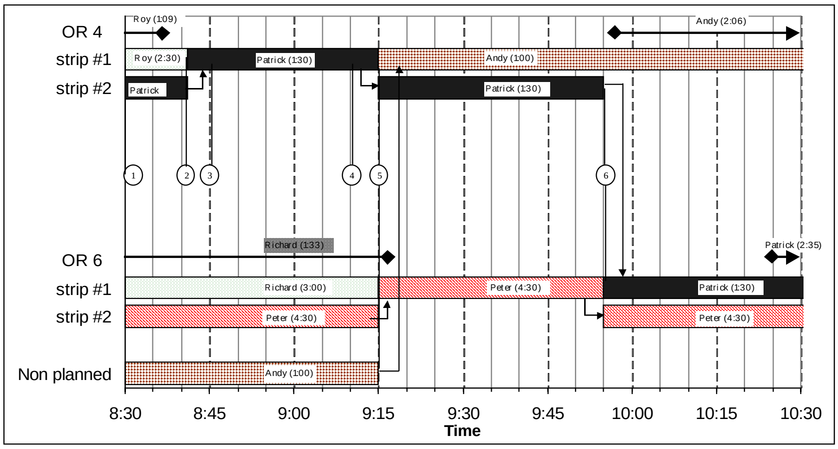


Figure 2 - A graphical representation of the schedule during two hours with the start and end time of cases.

Activity time line: The list below shows the activities of the coordinator during re-scheduling of surgical cases. The numbers correspond with figure 2.

1- 8:30 Patient Roy was in OR4; case Patrick was to follow.
Patient Richard was in OR6 till 9:17; case Peter was to follow.

2- 8:41 An OR nurse **informed** the coordinator **in front of the planning board** that case Roy in OR4 was finished; 1 hour earlier than expected. The coordinator **removed the strip from the board**.

3- 8:45 The coordinator made a new plan to forward the next surgery **by overseeing the planning board**. She called and paged the surgeon of patient Patrick from 8:45 till 9:15 to tell him that the operation could start earlier, but he did not answer.

4- 9:10 The coordinator generated another plan. She wanted to let the non-planned case Andy go before case Patrick, because she knew that this surgeon was available for almost the whole day. The coordinator **asked** the surgeon of case Andy **in front of the planning board**: "Are you available to do the surgery for your patient Andy when OR 4 is cleaned?" He confirms that he is available.

5- 9:15 The coordinator finally reached the surgeon of case Patrick and said that his case is delayed because case Andy goes first. After that the coordinator **changed the schedule on the planning board** and asked the patient's availability.

6- 9:55 The surgical fellow of case Patrick, who assists the attending surgeon, **requested the coordinator in front of the planning board** if her case could start earlier. The coordinator changed case Patrick to OR6 after agreement with the surgeons of patients Peter and Patrick.

From this example and data from other days, a number of distributed planning functions are derived that were associated with the use of the planning board. Part of the activities by the coordinator was to disseminate plans for their execution. The coordinator disseminated plans in part by updating information on the planning board for the other participants. The collaborators would discuss the planning while pointing to the board. An example of using the board was observed when a case was cancelled. Through the changes on the board, the nurse assigned for that case knew that they did not have to prepare for it. The coordinator knew that the nurse was aware of the new situation without talking to her. The collaborators (surgeons, anaesthesiologists and nurses) planned their activities with the overview of the planning board and sometimes requested changes while pointing at it.

We found that the planning board supports the following planning functions: 1. Disseminating individual tasks, 2. Orientating in a wider context about the previous, current and future situation in the OR's, 3. Establishing shared awareness, such as when observing people in front of the board and therefore assuming that they are aware of the last updated schedule, 4. Establishing common basis during negotiating, 5. Displaying opportunities for new plans and the consequence problems can cause.

Discussion

Our results have several implications for designing computerised artefacts to support cognitive activities in distributed planning. We will discuss three areas: distributed access of information, input of status information and visual aids. In general, these implications are to improve situation awareness as well by the coordinator as by dispersed collaborators, both in terms of accuracy of situation awareness and in terms of speed of gaining situation awareness.

The use of a planning board without distributed access requires many (long) phone calls. Telephone communications are a disruptive form of communication. Through a planning board with distributed access, the coordinator can distribute status and plan information asynchronously. Conversely, people remote to the coordinator can enter status information at their desk with low workload impact. For example, a nurse can enter the important landmarks of a surgery case to allow better monitoring by the coordinator. Wireless devices, such as personal digital assistants (PDA), can potentially be connected with the coordinator to lower the cost of communication, especially through asynchronous means. Short messages could be sent to relevant collaborators. In the case of OR management, a message can be "Your surgery is delayed for at least one hour" or "This case is cancelled, stop preparing the equipment." Other access methods are also as possible, such as those tested in (Berkowitz et al., 1999).

Electronic planning boards are commercially available, for example Navicare® Systems. They make it possible to deploy visual aids to the coordinator. The overview we developed for capturing events (figure 2) can be used for a graphical representation of the current status and events. The visual aids can reduce cognitive workloads by comprehending the current situation and projection of future status. The visual aids can potentially also help analyse and show opportunities and bottlenecks for new plans. This early analysis allows the coordinator to act more proactively. If this overview of current and future status can be distributed fast to collaborators, it can serve as a basis for negotiating new plans.

In short, this article shows that the planning board supports situation awareness. We recommend more research on usability of distributed planning tools, especially for visual aids and wireless personal digital assistants.

Acknoledgement

The research was in part supported by a grant from the National Science Foundation (#9900406). The opinions presented here are those of the authors.

References

Xiao, Y., Lasome, C., Moss, J., et al. (2001). Cognitive Properties of a Whiteboard: A Case Study in a Trauma Centre. In *Proc. of the 7th Eur.Conference on ECSCW* (pp. 259-278). Bonn, Germany

Roth, E. M., Malsch, N., Multer, J., et al. (1999). Understanding how train dispatchers manage and control trains: A cognitive task analysis of a distributed team planning task. In *Proc. of Human Factors and Ergonomics Society Annual Meeting (pp. 218-222).*

Berndtsson, J., Normark, M. (1999). The coordinative functions of flight strips: air traffic control work revisited. In *Proc. of ACM SIGGROUP conference on Supporting Group Work* (pp. 101-110). New York, NY: ACM Press.

Berkowitz, D. A., Barnett, G. O., et al. (1999). eWhiteBoard: A real time clinical scheduler. *American Medical Informatics Association,* Annual Symposium 1999, Washington, DC, AMIA.

Virtual Reality as Enabling Technology for Data Collection of Second-Generation Human Reliability Methods

S. Colombo*

Politecnico di Milano, Dipartimento di Chimica, Materiali e Ingegneria Chimica "G. Natta", P.zza Leonardo da Vinci, 32, 20133 Milano, Italy

Objective of the proposal

As Fig. 1 synthesise, the overall objective of this paper is to propose an alternative perspective to the aged issues of data collection and expert judgement using Virtual Reality. Moreover the paper poses the question on whether the exploitation of Virtual Reality as enabling technology will modify the inherent structure of second-generation at the point that the new resulting structure can be considered as a new evolutionary third-generation HRA methods.

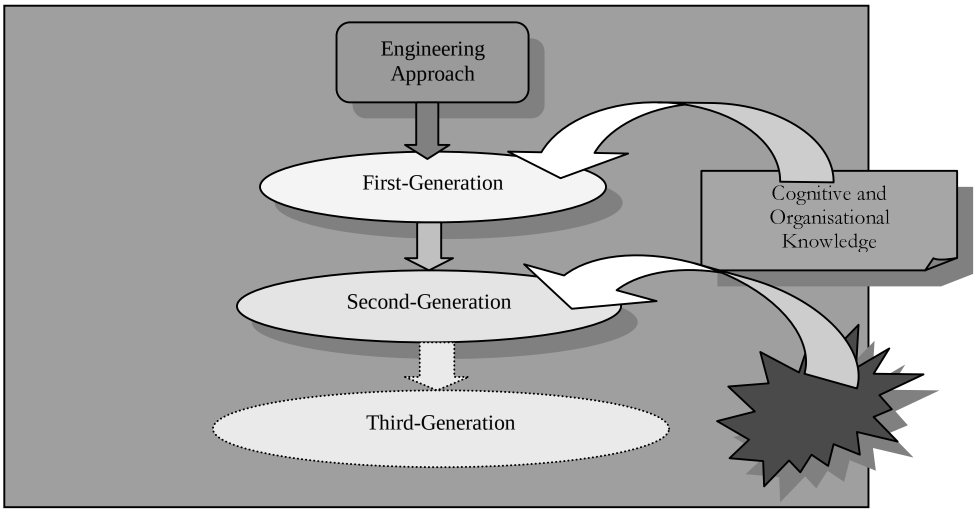


Fig. 1 Third-Generation dilemma

Data collection

The importance of data and the difficulties of their retrieval: As for any scientific method used for predicting future occurrences even for HRA methods the role of data is crucial both for validating the method itself and for carrying out the analysis. As Fig. 1[1] shows, current and historical data are both necessary whether a prediction on human performance and human reliability is sought.

Moreover data for being of any help in the assessment must have two fundamental characteristics, namely:

- Be reliable, in the sense they must be collected within contexts and situations that are representative of the real working context, and;
- Be consistent with the methodology's requirements.

At the operational stage, the data collection task is limited by some constraints that have their root on both socio-cultural and methodological aspects, namely:

- ✓ Retrieval systems are frequently instituted whereby disciplinary actions are to be taken against those who commit "errors" (erroneous actions) or even inadvertently (i.e. non-deliberately) violate internal norms and rules;
- ✓ Retrieval systems receive, on average, reports on those events that have caused the process to take a dangerous course and that has clearly been visible or electronically recorded by the control system, i.e. they are very often closer to incidents/accidents than near misses;

* Tel: 0039.02.23.99-3262; Fax: 0039.02.70.63.81.73; e-mail: simone.colombo@polimi.it

- ✓Retrieval systems are usually conceived for collecting actions, interventions and system variation that alter the technical specifications neglecting those that can cause latent failures;
- ✓Retrieval systems, probably due to a mix of resources constraints and the above reasons, do not normally have a sufficient number of informations (events) stored to be considered statistically significant.

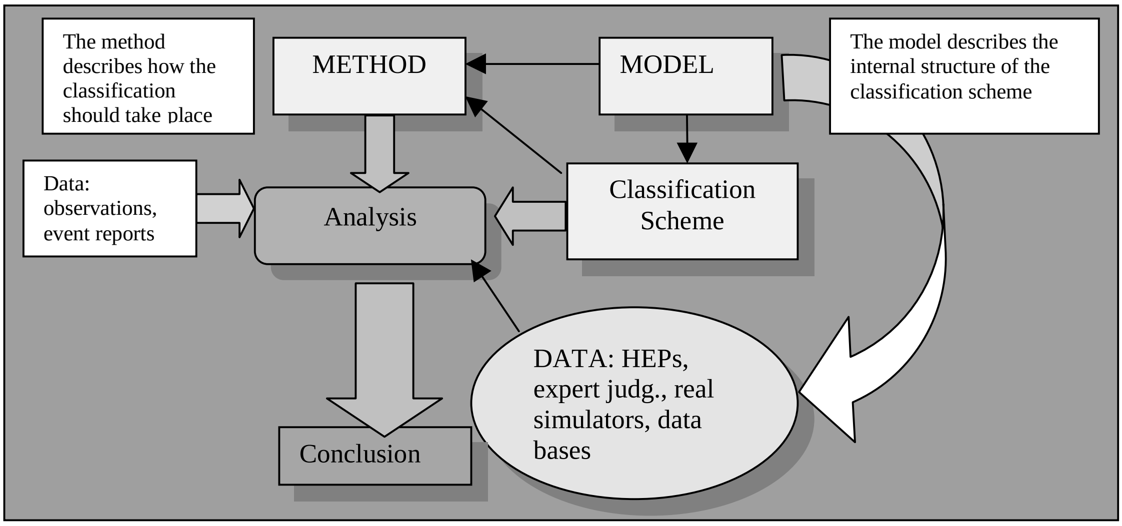


Fig. 2 The MCM framework for performance prediction (from Hollnagel [1])

At the design stage the difficulties of data collection simply relate to costs: data on human factors are not normally collected since building up a ***real*** simulator of any "approved" plant design for assessing the associated human factors it would not be affordable. The ergonomic aspect is addressed by qualitative analysis, like HAZOP, FMEA and the like, expertly weighted up with the experience gained in prior similar projects. Of course being able to collect predictive human factors data associated with any commissioned work it would enable to make suitable modification of project layouts substantially improving the inherent safety before the plant is put in place. In this regards second-generation HRA methods seems to be promising to provide a qualitative indication of the goodness of Human Technology Interfaces even at the design stage [2][3].

Current ways of collecting data

At present data on human factors are gathered following four different ways, namely [4]: Observation at laboratory and simulators; Observation at work places; Interviews and questionnaires; Mandatory and confidential reports. Data artificially collected through laboratory-reproduced situations and within ***real*** simulators can have the inconvenient of not being representative since human beings, differently to machines, are context-sensitive. Assessing and predicting the performance of a pump under different socio-cultural contexts, e.g. laboratories instead of real working environments, it does not make any difference for the pump once the operational variables (like temperature, pressure, degree of humidity and the like) are the same or within the expected range fixed by designer. For human beings this way does not work properly since they are more or less consciously context-sensitive. This means that trying to deduce the probability of occurrence of undesired specific erroneous actions under given stimuli within an aseptic context (i.e. a laboratory or a real simulators), could result enormously misleading. This is probably worthy for testing hearing, sight, feeling, sense of smell, and reflexes but a bit wonky for checking cognition responses.

Working context reproductions for being helpful and reliable means for assessing human proneness to unsafe (erroneous) behaviours, should be extremely realistic, vivid, and include, at least, "the essential" of the real working environment, that is all those elements (hardware and software) that make the operator under simulation perceiving the same feelings (human factors) as the ones daily perceived in its real micro world (e.g. the same psychological pressure to get the production done, the same influence by supervisor, the same disturbances like noises, smells, lights, temperature, and the like). The difficulty is clearly to define what of the actual working environment is “the essential” and what it is not in order to reliably assess the cognitive behaviour through its reproduction.

VR as the way out for data collection

It is widely acknowledged that data collection it is not an easy and straightforward task; that’s why, very often, data are adapted and translated from one domain, typically the nuclear domain, to the one of interest. Yet following this way their representativeness is verisimilarly compromised, and with it even the outcomes of the analysis since the model through which data should translate from one domain into the others it is opaque. In any case, even if the translation model would be uncovered and well defined, the underlying premises would be arguable. In fact when translating data from one domain to another is implicitly assumed that the same stimulus can give rise to a similar cognitive “reaction” and then to a similar visible manifestation (action), independently to the context.

Data must then be collected within each specific domain of application and, moreover, the optimum would be within each specific micro-environment in which the human being daily operate.Virtual Reality in this sense could provide a generous help thanks to its inherent flexibility of vividly reproducing the more varied working-context situations.

The perspective of “virtual” data: The work carried out up to now within PRISM Network [5] has allowed to envisage that the perspective given by data collected through ***virtual*** simulators (Virtual Environments) seems to be more reliable than the one given by ***real*** simulators since within synthetic environments one could navigate and explore one’s own virtual micro-world perceiving almost the same feelings (human factors) as those daily perceived when undertaking tasks in the real world. For this reason data collected under synthetic scenarios could be more reliable than those collected in real simulators and probably ready to be used as such without any further manipulation or extrapolation. Of course the possibility that operators could fail when asked to carry out the same actions at the operational level (in reality) is still present but its investigation goes beyond the aim of this paper since has to do with the validation process of an idea that has to be further investigated and developed.

Although only conjectures can presently be made about reliability of data collectible using Virtual Reality technology, the comforting aspect is that, thanks to the flexibility of virtual reproductions, data will fulfil the methodology’s requirements since they can be collected under synthetic environments that can be modelled and adjusted since they reflect the real micro-worlds characteristics.

*Positive aspect of using VR for collecting data:*The first and most important added value that none but virtual simulators can provide is the possibility of “living” an accidental sequence and its effects without experience the real consequences. Any action/intervention made upon the productive system and its consequence, modelled and virtually reproduced through VR devices, can be seen and, say, *virtually experienced*. The most varied scenarios, old, new and the ones under which somebody else in the same domain has gone through, can be tested, leaving traces on people mind.

Another added value that will probably make virtual simulators more attractive than real ones is the economic character. At present even though there are not effective costs comparison, what can be assumed as verisimilar is that a VR suite, and the associated cost of modelling, seems to be less expensive than a real simulator (mock-up).

Conclusions

The coming of second-generation has allowed, on the one hand, to tackle human reliability in a more holistic way providing outcomes that are more representative than the ones achievable with the first-generation. On the other hand this evolutionary change has not allowed to enshrine the quantitative feature “subtracting” a necessary characteristic to all of those who have to dealing with the Quantitative Risk Assessment (QRA) process.

This contribution has proposed a new perspective of collecting data of second-generation HRA methods for allowing the quantification of human reliability to be newly restored.

Despite the general uncertainties linked with the usage of VR and IV for predicting human behaviour, such as the failure at the operational level, are present even for the present idea, the progresses made so far have allowed to focus the attention on how to integrate VR tools and techniques within the theoretical framework leaving behind the doubt on whether this integration would have been feasible.

Acknowledgements
The author wishes to thank Prof. G. Biardi and Dr. P.C. Cacciabue which many essential contributions have allowed to make this proposal presentable.

References

[1] Hollnagel E. (1996). "*Reliability Analysis and Operator Modelling*". Reliability Engineering and System Safety", Vol. 52, pp. 327-337.
[2] Kim I. S. (2001). "*Human Reliability Analysis in the Man-Machine Interface Design Review*". Annals of Nuclear Energy, Vol. 28, pp. 1069-1081.
[3] Piccini M., Carpignano A. (2002). "*Human Factors in the Design Process of Advanced Supervisory and Control Systems*". Proceedings of the European Conference on System Dependability and Safety. Lion
[4] Cacciabue P. C. (2000). "*Human Factors Impact on Risk Analysis of Complex Systems*". Journal of Hazardous Materials, Vol. 71, pp. 101-116.
[5] European PRISM Network at http://www.prism-network.org.

Learning and Failure in Human Organisations

Darren Dalcher,

Software Forensics Centre, Middlesex University, London, N14 4YZ, UK.
d.dalcher@mdx.ac.uk

Abstract: This paper looks at the role of learning in organisational behaviour in terms of response to identified problems viewed from the perspective of action research. If normal action entails making decisions, obstacles stopping decision makers provide a learning opportunity. The reflection that results from identifying a problem facilitates the learning. Different learning strategies invoke the difference between simple and complex feedback systems with implications at the personal, as well as, organisational levels. Development of new systems therefore requires consideration of how, and where, learning is likely to take place so that adaptation in the light of problems becomes possible.

Keywords: learning, single-loop feedback, double-loop feedback, reflection.

Learning in Action

The continuous accumulation of knowledge and skills is conducted through the process of learning; a relatively permanent change in attitudes and behaviour that results from practice (Atkinson, 1996). Action researchers focus on the results of action, which is viewed as decisions and learning. Learning is the process of detecting and correcting error, where error is defined as any feature of knowledge or knowing, that makes action ineffective (Argyris, 1976). Dealing with errors results in learning, as action implies a problem with previous perceptions. Learning is therefore the act of repunctuating continuous experience (Weick, 1999). Errors are thus crucial to theories of action and the understanding of learning is fundamentally associated with detection, diagnosis and correction of errors.

Kolb (1986) rationalised that learning often starts with the experience of an event or stimulus which the individual reflects upon in trying to make sense of it. Reflection enables practitioners to deal with troublesome divergent situations of practice that do not conform to normal expectations and procedures. Learning takes place when a mistake or mismatch is acknowledged, its producers are identified, and it is corrected (Argyris, 1976; Ackoff, 1995). Ackoff (1995) therefore argued that it was better to do the right thing wrong than the wrong thing right as the former led to learning while the latter simply reinforced an error. Detection and correction of errors equates with learning and provides the core activity of any organisation or system (Argyris, 1980). Individuals engaged in that activity in an organisational capacity, become agents of organisational action and learning (Sage, 1995).

Comparing Learning Strategies

Organisations do not produce the behaviour that leads to learning as this is done by **individuals** acting as agents of the organisation (Argyris, 1988). Surprise or unfulfilled expectations lead to interruption of on-going activities as part of the need to find an explanation (Louis, 1980). The move from error detection to error correction entails learning, as the sources of error must be discovered prior to action.

When the process allows an organisation to maintain its current policies, the organisation employs a basic, thermostat-like single-loop learning procedure (Argyris, 1988). Single-loop feedback is essential for focusing on operational effectiveness, norms and performance issues. Simple learning becomes concentrated on the adjustment of parameters to correct performance without examining the assumptions and theories that underlie performance deviations. The approach thus, relies on the assumption of rationality as the model strives for the most satisfactory solution. The four basic values shared by people operating in this mode (Argyris, 1982; Argyris, 1988) are to: achieve their purposes through controlling the environment; maximise winning and minimise losing; utilise defensive norms in order to minimise negative feelings; and emphasise intellectual rationality and objectivity and minimise emotions.

The value of single-loop feedback is in the immediate response that enables the system to maintain unadjusted performance levels and optimise their performance in the short-term through progression from the present towards an optimised (and fixed) goal. Computing technology is particularly amenable to implementing this kind of single-loop simplification which aims to offer a *satisficing* alternative to radical change. Learning is thus limited to *satisficing* and the replication of previous successes and trends as control parameters.

A more comprehensive approach to learning would entail a double-loop procedure that enables an organisation to question the underlying goals, objectives and policies (see Fig. 1). Double-loop learning is dynamic and recognises the need to alter performance norms rather than purely focus on maximising them. It enables utilisation of new ideas, exploitation of emerging opportunities and reflection about past performance. The values in this approach (Argyris, 1977; Argyris, 1982) focus on: helping to produce valid information as basis for action; making free and informed choices; and, combining commitment with constant monitoring of the implementation and preparedness to change.

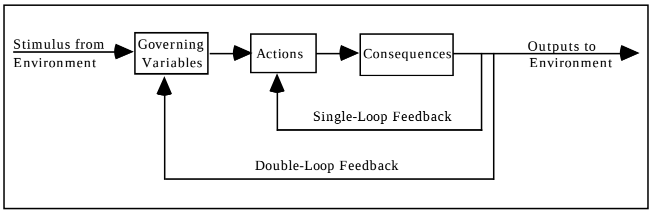


Figure 1. Interpretation of Double-Loop Feedback and Learning

Double-loop learning enables the evaluation of organizational assumptions in order to improve capabilities. Acceptance of the inevitability of change leads to realisation that goals are not stable and strategies need to be continuously invented, shaped and modified. The key emphasis is on responsiveness. Monitoring environmental feedback thus contributes to the determination of the need to redefine rules and norms in order to cope with change and bring about adaptation and self-organisation that are critical to the survival of any open system. This facilitates a bi-directional exploration capable of spotting opportunities, and actively monitoring the gap between dynamic objectives and a fast altering present Such true learning occurs not when a problem is identified or a solution proposed, but when the solution is implemented and performs against the standard (Argyris, 1988). It requires the active interaction of problem, solution and expected results — a far greater challenge for developers.

Which one to Use?

Argyris (Argyris, 1977) noted that many organisations apply single-loop learning procedures by focusing on the *status-quo* and correcting serious deviations. Application of double-loop learning under these conditions will only occur as a result of: an external crisis in the environment; a revolution from within; or, an internal crisis precipitated by management Consequentally, organisations in single-loop modes seem to be driven by crises. The delay in establishing organisational response frames in response to crisis situation serves to escalate the crisis. Even following prolonged success is likely to give rise to the 'stuckness problem' (Miller, 1990), where success leads to failure as organisations get trapped by the patterns that created success in the past. Similarly, Argyris (1977), identified a paradox in single-loop organisational response modes. Management are faced with multiple problems they need to address which enables them to ignore and suppress certain dilemmas. But, as Argyris noted at some point when the system becomes better established, ignored problems will catch up with participants.

Discussion and Implications for Computer Systems

Computers typically rely on single-loop feedback. Computational techniques may be used to supplant human decision makers when the contextual system is relatively closed. Relatively closed systems are good at ignoring the impact of the numerous environmental factors and avoiding the focus on human involvement within the system. However, the potential inabilities of humans to disaggregate situations into components and to analyse them places severe limitations on the application of computational techniques to open systems. Open systems, such as complex ambulance despatch systems or air traffic control systems, with their inherent imperfections and unknown factors, need to rely on more judgmental approaches and hence the process cannot be programmed explicitly.

Moreover, rule based analytical approaches cannot deal as well as an experienced operator with the small minority of difficult cases – i.e. the exact situations that are likely to generate reflection in humans. Such approaches wrongly reduce the influence of the operator. Overreliance on technology often results in ignoring the need for double-loop learning and the ability to intervene. Design, especially in systems involving reflection and experience, should come from the individual people outwards. Human ability and limitations thus need to be understood and designed into the system as part of the learning process. This offers the opportunity to work to the strengths of current resources, using the enabling technology to magnify these assets while taking care not to magnify the limitations so as to cripple the system.

Information Technology must be seen as part of the whole and adjusted to. With very little time for feedback, learning, and correcting, the arrangement of information needs to be accompanied by training and experience in dealing with feedback and recovering from disaster. As safety critical, mission critical and security critical systems become more common, reflection and learning considerations are likely to become more important. Implementing single-loop procedures will simply not suffice!

The following implications should therefore be addressed whenever new systems are designed:
The Need for learning arises from the requirement to adjust knowledge in light of observed deviations. This should be built into systems to allow for double feedback loops.
Moreover, the process that is utilised in building new systems should likewise allow for learning to take place prior to implementation so that users and operators are not forced to conduct all the learning in 'real-time' while working on the new system
Learning often comes down to whether the willingness to learn and to dedicate resources to addressing a mismatch exists. From an economic perspective, it makes sense to facilitate learning early on and to allocate resources to the task to reduce the need for 'emergency learning'
Imposing a new working environment is likely to lead to a variety of adjustment problems. Involvement can alleviate some of the tensions while providing an early opportunity to experience some of the implications.
Organisational ability to learn results from individual learning. Operators and users need training in how to reflect and learn effectively (rather than encouraged to hide and disconfirm and conform)
The culture and perceived openness of an organisation dictate the type of defensive routines that are likely to be adopted. In order to facilitate learning, an organisational attitude that is more open towards mismatches and challenges is required.
Reflection is the key to learning in practice. The ability to reflect is a key skill that may enable professionals to deal with challenges and improve, while simultaneously enhancing their value to the organisation, as well as the value of the organisation.

References:

Ackoff, R. L. (1995). 'Whole-ing' the Parts and 'Righting' the Wrongs. Systems Research **12**(1): 43-46.
Argyris, C. (1976). "Single and Double Loop Models in Research on Decision Making." Administrative Science Quarterly **21**: 363-375.
Argyris, C. (1977). "Double Loop Learning in Organisations." Harvard Business Review: 115-125.
Argyris, C. (1980). Some Inner Contradictions in Management Information Systems. The Information Systems Environment. W. Lucas and e. al., North Holland**:** 99-111.
Argyris, C. (1982). "The Executive Mind and Double-Loop Learning." Organisational Dynamics: 5-22.
Argyris, C. (1988). Problems in Producing Usable Knowledge for Implementing Liberating Alternatives. Decision Making: Descriptive, Normative and Prescriptive Interactions. D. E. Bell, H. Raiffa and A. Tversky. Cambridge, Cambridge University Press**:** 540-561.
Atkinson, R. L. and e. al. (1996). Hilgard's Introduction to Psychology,. Orlando, Harcourt Brace.
Kolb, D. A. (1986). Experiential Learning. Englewood Cliffs, NJ, Prentice-Hall.
Louis, M. (1980). "Surprise and Sensemaking: What Newcomers Experience in Entering Unfamiliar Organisational Settings." Administrative Science Quarterly **25**: 226-251.
Sage, A. P. (1995). Systems Management for Information Technology and Software Engineering. New York, John Wiley.
Weick, K. E. and F. Westerley (1999). Organisational Learning: Affirming an Oxymoron. Managing Organisations-Current Issues. S. R. Clegg, C. Hardy and W. R. Nord. London, Sage Publications**:** 190-208.