IET Control Engineering Series 60

People in Control

Human factors in control room design

Edited by Jan Noyes
and Matthew Bransby

IET CONTROL ENGINEERING SERIES 60

Series Editors: Professor D.P. Atherton
Professor G.W. Irwin

People in Control
Human factors in control room design

Other volumes in this series:

Volume 2	**Elevator traffic analysis, design and control, 2nd edition** G.C. Barney and S.M. dos Santos
Volume 8	**A history of control engineering, 1800–1930** S. Bennett
Volume 14	**Optimal relay and saturating control system synthesis** E.P. Ryan
Volume 18	**Applied control theory, 2nd edition** J.R. Leigh
Volume 20	**Design of modern control systems** D.J. Bell, P.A. Cook and N. Munro (Editors)
Volume 28	**Robots and automated manufacture** J. Billingsley (Editor)
Volume 30	**Electromagnetic suspension: dynamics and control** P.K. Sinha
Volume 32	**Multivariable control for industrial applications** J. O'Reilly (Editor)
Volume 33	**Temperature measurement and control** J.R. Leigh
Volume 34	**Singular perturbation methodology in control systems** D.S. Naidu
Volume 35	**Implementation of self-tuning controllers** K. Warwick (Editor)
Volume 37	**Industrial digital control systems, 2nd edition** K. Warwick and D. Rees (Editors)
Volume 38	**Parallel processing in control** P.J. Fleming (Editor)
Volume 39	**Continuous time controller design** R. Balasubramanian
Volume 40	**Deterministic control of uncertain systems** A.S.I. Zinober (Editor)
Volume 41	**Computer control of real-time processes** S. Bennett and G.S. Virk (Editors)
Volume 42	**Digital signal processing: principles, devices and applications** N.B. Jones and J.D.McK. Watson (Editors)
Volume 43	**Trends in information technology** D.A. Linkens and R.I. Nicolson (Editors)
Volume 44	**Knowledge-based systems for industrial control** J. McGhee, M.J. Grimble and A. Mowforth (Editors)
Volume 47	**A history of control engineering, 1930–1956** S. Bennett
Volume 49	**Polynomial methods in optimal control and filtering** K.J. Hunt (Editor)
Volume 50	**Programming industrial control systems using IEC 1131-3** R.W. Lewis
Volume 51	**Advanced robotics and intelligent machines** J.O. Gray and D.G. Caldwell (Editors)
Volume 52	**Adaptive prediction and predictive control** P.P. Kanjilal
Volume 53	**Neural network applications in control** G.W. Irwin, K. Warwick and K.J. Hunt (Editors)
Volume 54	**Control engineering solutions: a practical approach** P. Albertos, R. Strietzel and N. Mort (Editors)
Volume 55	**Genetic algorithms in engineering systems** A.M.S. Zalzala and P.J. Fleming (Editors)
Volume 56	**Symbolic methods in control system analysis and design** N. Munro (Editor)
Volume 57	**Flight control systems** R.W. Pratt (Editor)
Volume 58	**Power-plant control and instrumentation** D. Lindsley
Volume 59	**Modelling control systems using IEC 61499** R. Lewis
Volume 60	**People in control: human factors in control room design** J. Noyes and M. Bransby (Editors)
Volume 61	**Nonlinear predictive control: theory and practice** B. Kouvaritakis and M. Cannon (Editors)
Volume 62	**Active sound and vibration control** M.O. Tokhi and S.M. Veres
Volume 63	**Stepping motors: a guide to theory and practice, 4th edition** P.P. Acarnley
Volume 64	**Control theory, 2nd edition** J. R. Leigh
Volume 65	**Modelling and parameter estimation of dynamic systems** J.R. Raol, G. Girija and J. Singh
Volume 66	**Variable structure systems: from principles to implementation** A. Sabanovic, L. Fridman and S. Spurgeon (Editors)
Volume 67	**Motion vision: design of compact motion sensing solution for autonomous systems** J. Kolodko and L. Vlacic
Volume 69	**Unmanned marine vehicles** G. Roberts and R. Sutton (Editors)
Volume 70	**Intelligent control systems using computational intelligence techniques** A. Ruano (Editor)

People in Control
Human factors in control room design

Edited by Jan Noyes
and Matthew Bransby

The Institution of Engineering and Technology

Published by The Institution of Engineering and Technology, London, United Kingdom

First edition © 2001 The Institution of Electrical Engineers

First published 2001

This publication is copyright under the Berne Convention and the Universal Copyright Convention. All rights reserved. Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted, in any form or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Inquiries concerning reproduction outside those terms should be sent to the publishers at the undermentioned address:

The Institution of Engineering and Technology
Michael Faraday House
Six Hills Way, Stevenage
Herts, SG1 2AY, United Kingdom

www.theiet.org

While the authors and the publishers believe that the information and guidance given in this work are correct, all parties must rely upon their own skill and judgement when making use of them. Neither the authors nor the publishers assume any liability to anyone for any loss or damage caused by any error or omission in the work, whether such error or omission is the result of negligence or any other cause. Any and all such liability is disclaimed.

The moral right of the authors to be identified as authors of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

British Library Cataloguing in Publication Data
People in control: human factors in control room design
(Control engineering series no. 60)
1. Control rooms – Design and construction 2. Human engineering
I. Noyes, Jan II. Institution of Electrical Engineers
620.8′2

ISBN (10 digit) 0 85296 978 3
ISBN (13 digit) 978-0-85296-978-6

Typeset by Tradespools Ltd, Frome, Somerset
Printed in the UK by MPG Books Ltd, Bodmin, Cornwall
Reprinted in the UK by Lightning Source UK Ltd, Milton Keynes

Contents

Lynne Collis and Felix Schmid

Iain MacLeod and Derek Smeall

Foreword

This book emanates from papers presented at the first *People In Control* conference held in Bath, 21–23 June 1999. The aim of the conference was to provide an opportunity for individuals to find out or update knowledge on the state-of-the-art of human–machine interactions within the context of the control room setting. The book echoes the aim of the PIC conferences. As industrial processes become more automated, there is increasing concern about the performance of the people who control these systems. Human error is increasingly cited as the cause of accidents across many sectors of industry, e.g. aircraft, railways, chemical plant, etc. It is also commonly recognised that human failings can significantly impact on the profitability of automated processes. Engineers are often responsible for designing the systems used by these 'people in control' and in setting up their management structures and processes. The objective of this book is to help engineers to design better systems and processes, and hence increase operational safety and efficiency. The chapters are grouped in three topic areas: human performance, methods and control room design. There is a particular emphasis on case studies and examples from the transport and process control industries, and the Navy are included.

One of the founder members and co-chair of the *People In Control* conference was Dr Matthew Bransby. Sadly, Matthew died in an accident in August 2000. He had been taking part with his son, Ben, in the North Face Thumbnail 2000 expedition at Agdlerussakasit on the southern tip of Greenland. The group had successfully reached the top and were abseiling down when an anchoring gave way, and Matthew fell 1,000 feet to his death. The Thumbnail, as they nicknamed it, is regarded as one of the most challenging rock climbs in the world. At the time of his death, Matthew was in the process of editing this book, and as a mark of respect for the work that he did in this field, he is included as a co-editor.

Eur Ing Dr Matthew Bransby FIEE

Matthew initially embarked upon an academic career after gaining an MA in Mechanical Sciences from Trinity College, Cambridge in 1968, followed by a PhD at Southampton University in 1971. For a number of years, he taught in the

Department of Control Engineering at Sheffield University. In 1975, he moved to industry working as a Research Officer for the then Central Electricity Generating Board where he worked on the on-line computer control of boiler pressure and generated load at Drax Power Station.

Matthew saw himself as a control engineer first and foremost, but his experiences in academia and industry had led him to appreciate the importance of the human element in the design of systems and the disasters that could occur if this aspect was neglected or overlooked. In this sense, he was a real advocate for the Human Factors/Ergonomics movement, and would describe himself as a control engineer 'wearing a thin coat of Human Factors psychology'. Below is an excerpt that demonstrates this, written by Matthew for the first *People In Control* conference:

> I have spent a good part of my professional life as a control engineer in industry. When I left academia to go into industry I was full of mathematical concepts about multivariable control, plant modelling and random processes. And I spent a few happy years trying to apply some of these ideas to live plant. It was great fun! Over the years my focus changed. I started to realise that the mistakes I had made with my control systems were not the equations, but the interfaces with the plant operators. They would say things like 'I couldn't understand what it was doing, so I switched it off', or 'What is that light meant to mean?' In my experience, plant operators learn to live with some pretty horrible systems designed by us control engineers.
>
> Recently, I carried out a survey of control room alarm systems. The results were quite clear. Most plant operators receive many nuisance alarms in normal operation, and in any serious plant upsets they are often completely overwhelmed. In one incident the operator was bombarded with over 150 alarms per minute for the first 15 minutes of the disturbance. To add to the stress, the audible warning was sounding continuously. In terms of hardware and software the alarm systems had been very well designed. And they had been thoroughly tested to ensure that they could cope with the worst possible alarm flood. However, they were virtually useless to the operator trying to manage the emergency. Again, we engineers were allowing the technical performance of the system to cloud the needs of the users.

Having reached 50, and then with National Power at Swindon, Matthew took early retirement, but continued working as an independent consultant for his business, Bransby Automation Ltd. His particular interest was alarm design – a subject about which he was undoubtedly one of the UK experts: he was responsible for authoring the much acclaimed *EEMUA Guide on Alarm Systems*. Due to the success of Bransby Automation in winning contracts, Matthew's planned retirement didn't really materialise and he spent a large part of the last few years 'staying over' in London in order to make an early start working on the new Jubilee underground line.

Matthew lived life to the full: he was always positive, enthusiastic and optimistic that the IEE projects we shared would work. And he was always right, because he had a unique talent to place himself in the position of others and work out what they would want from a colloquium, a tutorial or a conference. This, combined with his engineering background, his sensitivity for the psychological issues and his experiences in academia, industry and business, made him a very special person.

Jan Noyes
April 2001

Preface:
From HMS to HCI – and back again

Erik Hollnagel

For those of us who have been working for some time with the issues covered in this book, it is interesting to reflect on the changes in terminology that have taken place. In the 1970s, the range of problems was treated under the name of Man–machine systems (Meadow, 1970; Singleton, 1974), which later became human–machine systems (HMS) to comply with the movement to eradicate sexist language. (This linguistic malaise was largely confined to the English language, since other languages have separate words for human and male.) In the 1980s, under the onslaught of the personal computer (PC) revolution, the terminology changed once again to become human–computer interaction (HCI). This time, it was a change of both name and focus. In one sense, replacing 'machine' by 'computer' mainly acknowledged that a new technological archetype had arrived – although it also carried with it a strong metaphor of the mind that knocked psychology off its feet. In another sense, there was a real change, since the proliferation of computers and their wanton introduction into almost every field of human endeavour created a strong growth in problems with the interaction between humans and technology. The main problem was that people now had to use computers and the clunky interfaces that went with them to do things that hitherto had been accomplished so well with good old-fashioned technology. There was also a surge in the number of people with specialised educations who graduated from Computer Science and Psychology departments (as well as people who drifted into the area from artificial intelligence [AI] and other topics) and who considered this field as their prerogative. This was in marked difference to the Man–machine systems research of old, which had been carried out by a relatively small number of open-minded engineers and industrial psychologists who happened to share the same interests. The flood of experts on human–computer interaction meant that the main focus turned to problems directly related to the design of the interface and to problems that were not primarily connected to process control and classical human–machine systems research. A case in point was the emergence of usability engineering (Nielsen, 1993), which proposed a solution to a largely self-inflicted problem.

Similarly, even the field of industrial engineering became inundated with human–computer interaction problems because well-meaning engineers, managers and, in particular, vendors hoping to get a competitive advantage, saw fit to follow the general trend towards computerisation and the replacement of conventional interfaces (as they became known) with computerised ones. The new technology and the new interfaces removed the traditional limitations on the number of measurements and the amount of much information that could be brought to the control room and furthermore made it possible to present it all in a variety of forms and modes. This new diversity, which may have been appreciated by designers more than by operators, led to what became known as the data overload problem (Woods and Watts, 1997).

The full-scale attack from computerisation and digitalisation may have been tempered by experience in the field of process control, but has far from subsided. Indeed, other domains such as military command and control seem to follow the same old paths with renewed enthusiasm (e.g. Kallmeier *et al.*, 2000). Within the process industries, the computerisation of the human–machine interaction has, however, not only created more problems than it solved but also failed to solve the old problems completely. For an example of that one need look no further than alarm systems (Stanton, 1994). So while engineers initially had eagerly welcomed human–computer interaction as the long-sought solution to all human factors problems, they gradually realised that there was more to controlling a process than designing an impressive graphical interface. One consequence of that was yet another change in the terminology, this time to emphasise the notion of a system, leading to labels such as 'human interface systems engineering' or even a revival of the term 'human–machine system'.

The return, so to speak, from HCI to HMI does not mean that the years in between have been wasted. Even though the preoccupation with HCI in some sense was a distraction from the basic problems of how humans use technology, it also helped to broaden the perspective and provide valuable experiences. Thus, one essential difference between the study of human–machine systems and the study of HCI is that the former recognises the problems that come from coping with a dynamic process or a dynamic world. The tasks considered in the field of HCI have typically been user-paced rather than task-paced, hence focused on working with computers rather than through computers. In other words, the application has in all essence been residing in the computer rather than in a complex, dynamic and often unpredictable environment. There has therefore been little need to consider how users could handle multiple simultaneous tasks, and how they could manage to do so under serious time constraints. The tasks considered by human–machine systems, on the other hand, are typically complex and dynamic. The tasks are complex because the problems facing the operator usually are dependent and closely coupled, rather than independent and loosely coupled, cf. Perrow (1984). Therefore, they cannot be dealt with one by one, but must be considered as a whole

– even though the whole may be incompletely understood. The tasks are dynamic because the requirements and conditions, more commonly known as the context, change over time even if the operators do nothing at all. Dynamic processes are thus spread out in both time and space, and differences in process speeds may easily cover five orders of magnitude.

Another important difference is that the field of human–machine systems deals with processes and controlling systems that are spread out in time and space, rather than a simple human–computer dyad. Two characteristic examples are air traffic management and electrical power distribution. A general trend is that the spatial distribution of control increases rapidly due to advances in information technology and that the temporal distribution, particularly of functions such as planning and scheduling, also grows spurred by the need to optimise the use of resources. In order to control properly the processes, operators must have an adequate understanding of both time and space that matches the time/space characteristics of the physical process well enough for actions to be planned and carried out. The design of the interaction between people and processes must consequently encompass both time and space, and the interface must provide an adequate representation of the dynamics and topography of the process.

The title of this book also represents a change from a structural to a functional point of view. Just as cognitive systems engineering defined 'coping with complexity' as one of the main areas of concern (Rasmussen and Lind, 1981; Hollnagel and Woods, 1983) so does *People In Control* focus on problems that operators may have in managing the processes of which they are expected to be in control. In the structural view, the focus is on how a sub-system with one set of characteristics (such as a human) can almost efficiently interact with a sub-system with another set of characteristics (such as a computer). In the functional view, the issue is how the overall system – the human–machine ensemble – can achieve its goals, given that the environment is dynamic and partly unpredictable and given that many parts of the functionality become automated and thereby – intentionally or unintentionally – hidden from the operators. The almost paradigmatic condition is when people do not understand what is happening or what they are supposed to do. This condition unfortunately appears with increasing frequency, as clearly illustrated by the chapters that follow. However, it is not just a concern for the proper design and use of technology, but also for the softer issues such as training, procedures and co-operation. The solution, therefore, requires a cross-disciplinary collaboration and understanding that can lead to the development of new models and methods. The chapters of this book provide excellent examples both of the issues that must be addressed and of the diversity of solutions that have been suggested.

References

1 HOLLNAGEL, E., and WOODS, D. D.: 'Cognitive systems engineering: new wine in new bottles', *International Journal of Man-Machine Studies*, 1983, **18**, pp. 583–600

2 KALLMEIER, V. L., HENDERSON, S., McGUINNESS, B., TUSON, P., HARPER, R., PRICE, S. N., and STORR, J. P.: 'Towards better knowledge: a fusion of information, technology, and human aspects of command and control'. Paper presented at Symposia at Shrivenham on *People in digitized command and control*, December 12–14, 2000
3 MEADOW, C. T.: 'Man–machine communication' (John Wiley & Sons, New York, 1970)
4 NIELSON, J.: 'Usability engineering' (Academic Press, Boston, 1993)
5 PERROW, C.: 'Normal accidents: living with high-risk technologies' (Basic Books, New York, 1984)
6 RASMUSSEN, J., and LIND, M.: 'Coping with complexity' (Risø-M-2293) (Risø National Laboratory, Roskilde, 1981)
7 SINGLETON, W. T.: 'Man–machine systems' (Penguin Education, Harmondsworth, 1974)
8 STANTON, N. (Ed.): 'Human factors of alarm design' (Taylor & Francis, London, 1994)
9 WOODS, D. D., and WATTS, J. C.: 'How not to have to navigate through too many displays', *in* HELANDER, M. G., LANDAUER, T. K., and PRABHU, P. (Eds.): 'Handbook of Human–Computer Interaction, 2nd edition' (Elsevier Science, Amsterdam, 1997)

Contributors

Les Ainsworth
Synergy
Royal Oak Barn
The Square
Cartmel
Cumbria
LA11 6QB

Arthur Aune
Norwegian University of Science and Technology
Department of Engineering Cybernetics
N-7194
Trondheim
Norway

Matthew Bransby

Lynne Collis
Halcrow Group
44 Brook Green
Hammersmith
London
W6 7BY

Malcolm Cook
University of Abertay Dundee
158 Marketgait
Dundee
DD1 1NJ

Lucy Cordiner
Air Traffic Management Development Centre
National Air Traffic Services
Bournemouth Airport
Christchurch
Dorset
BH23 6DF

Chris Dicken
Innogy plc
Operations & Engineering
Drax Business Park
PO Box 3
Selby
North Yorkshire
YO8 8PQ

Craig Donald
Leaderware
PO Box 783254
Sandton 2146
South Africa

Christian Heath
Work, Interaction and Technology Research Group
Management Centre
King's College
The Franklin-Wilkins Building
150 Stamford Street
London
SE1 8WA

Simon Henderson
Centre for Human Sciences
Building S7
QinetiQ
Fort Halstead
Sevenoaks
Kent
TN14 7BP

Erik Hollnagel
CSELAB
Department of Computer &
Information
Science
University of Linköping
S-581 83 Linköping
Sweden

Andrew Lichnowski
Innogy plc
Operations & Engineering
Windmill Hill Business Park
Whitehill Way
Swindon
SN5 6PB

Paul Luff
Work, Interaction and Technology
Research Group
Management Centre
King's College
The Franklin-Wilkins Building
150 Stamford Street
London
SE1 8WA

Wendy Macdonald
Centre for Ergonomics & Human
Factors
School of Human Biosciences
La Trobe University
3086
Victoria
Australia

Iain MacLeod
PMS Group
Type 45 PCO
5th Floor Platform
Design Centre
South Street
Scotstoun
Glasgow
G14 0XN

Matthew Mills
Software Engineering Centre
Building S11
QinetiQ
Fort Halstead
Sevenoaks
Kent
TN14 7BP

Neville Moray
Department of Psychology
University of Surrey
Guildford
Surrey
GU2 7XH

Sarah Nichols
Institute for Occupational Ergonomics
University of Nottingham
University Park
Nottingham
NG7 2RD

Jan Noyes
Department of Experimental Psychology
University of Bristol
8 Woodland Road
Bristol
BS8 1TN

Raphael Pascual
Centre for Human Sciences
Building S7
QinetiQ
Fort Halstead
Sevenoaks
Kent
TN14 7BP

Carl Sandom
Thales Defence Information Systems
Wookey Hole Road
Wells
Somerset
BA5 1AA

Felix Schmid
Department of Mechanical Engineering
University of Sheffield
Western Bank
Sheffield
S10 2TN

Andrew Shepherd
3 Outwoods Drive
Loughborough
Leicestershire
LE11 3LR

Charlotte Skourup
ABB Corporate Research AS
Bergerveien 12
PO Box 90
N-1375 Billingstad
Norway

Derek Smeall
Department of Electrical & Electronic Engineering
Cumbernauld College
12 Campsie View
Kildrum
Cumbernauld
G67 2AU

John Wilson
Institute for Occupational Ergonomics
University of Nottingham
University Park
Nottingham
NG7 2RD

John Wood
CCD Design & Ergonomics Ltd
Bank House
81 St. Judes Road
Englefield Green
Egham
Surrey
TW20 0DF

List of illustrations

List of tables

Part One

Human performance

Chapter 1

Human error

Jan Noyes

1.1 Introduction

One of the features of being human is making errors. Human errors are ubiquitous and frequent and they lead to a range of outcomes. We might make errors that are of no consequence, e.g. a linguistic error where you call someone by the wrong name, or we may make errors that have more serious results. Some errors will lead to incidents, e.g. pulling out to overtake a car whilst on the brow of a hill and narrowly missing colliding with oncoming traffic, and some will lead to being the primary cause of accidents, e.g. a fatal collision with an oncoming vehicle whilst overtaking. Although errors generally have negative connotations, i.e. we perceive their occurrence detrimentally, there are a group of errors that have positive outcomes. These are the errors that lead to the 'near miss' situation where the catastrophe does not occur but we learn from the situation and thus prevent a future accident. In the example above, the driver who attempts to overtake on the brow of a hill may have recently passed their driving test and not be aware of the potential hazard of overtaking in this particular road situation. However, the scare of narrowly missing a vehicle coming over the top of the hill will be sufficient to make the driver never overtake in these circumstances again. Thus, s/he has learnt something from the original error that prevents a road accident in the future. The concept of the 'near miss' is recognised in safety critical industries where a chain of events look as if they are leading to an accident but interventions manage to avert this happening, and much is gained from the situation in terms of it becoming a learning exercise. It can thus be seen that errors are of primary interest in the design of control rooms and associated operations because (i) they are an intrinsic feature of human behaviour, and (ii) they can lead to incidents and accidents. Thus, the need to understand human error is important because of its safety implications. This is particularly true in safety-critical situations such as those found in the process control industries, transport and medicine, where human error could seriously hamper the smooth running of procedures and place lives at risk. The aims of this chapter are to consider human error within the context of system design and to put forward some suggestions concerning its management in workplace applications.

1.2 What is human error?

The answer to this question may seem rather obvious, especially given the familiarity that we all have with making errors. However, despite the fact that the topic of 'human error' has been well researched, we are still some way from having any real, in-depth understanding of it (Amalberti and Wioland, 1997). Even establishing a precise and accurate definition for human error continues to prove a difficult and elusive goal (Senders and Moray, 1991). There are many reasons for this. First, there is an extreme view held by some individuals that human error does not exist at all, since it is a cognitive process that is neither observable nor measurable. Consequently, we can only deal with the outcomes of making errors and have to make inferences about the nature of the error. Second, there are a number of difficulties relating to how we classify errors (assuming we take the stance that human error does exist), for example when is an error defined as an error? An action may not be an error in itself, but leads to errors being made: a decision made in good faith today could in a couple of weeks time prove to be a wrong decision and a grave error. A further problem concerns establishing who decides the action is an error – should this be the individual who made the error, their line manager or other senior members in the organisation, or maybe the authorities, e.g. the police, the legal system, *etc*.? Moreover, some individuals may deliberately violate laid-down procedures but not view this as making errors, and indeed, this *modus operandi* may have become standard practice in the workplace. Errors also have a temporal aspect. A situation may arise where an error may be unrecognised for a period of time. If it has no consequences and if it remains undetected, the question arises of whether it should be counted as an error. A final difficulty relates to intent. Since humans make frequent errors, one of the issues is whether to class unintended actions of no consequence as errors. These may arise, for example, from moments of absent-mindedness. One study of errors on the civil flight deck indicated that during one flight of 11 hours, the crew made 162 raw errors (Amalberti, 1997). However, 157 of these 162 errors were detected and recovered from within 15 seconds. It is perhaps worth noting that the definition of error being applied here was to include all errors however basic, as well as the error having to be recognised by the crew when they were being debriefed as being an error. The latter was to safeguard against violation of regulations by the crew, and could be viewed as specific to this application's area. Therefore, it can be seen that crucial to the definition of error is 'intent' and whether actions that are unintentional should be included under the umbrella of error.

It can thus be seen that errors can arise because of a person's judgement and actions in a specific situation. To return to the overtaking example, it may be that on 99 out of 100 occasions, overtaking on approaching the brow of the hill could safely take place, and the driver would be totally unaware of any errors being committed. Hence, errors cannot be defined objectively by considering the performance of humans alone; they are caused not only by changes in performance with respect to the norm, but also by changes in the criteria of judgements. Successful completion of higher cognitive tasks often requires optimisation on several criteria, as well as being subject to time-variant characteristics. In conclusion, there is often no right or wrong decision, only the best decision for any given set of characteristics at a specific point in time. The outcome in terms of successful decision-making and avoidance of errors is often not known until later. To return to the question 'what is human error' a useful working definition might be to consider error as occurring when 'any member of a set of human actions exceeds some limit of acceptability'. Therefore, human error could be viewed as an out-of-tolerance action, where the limits are defined by the particular situation and/or system being used. In addition, there are a number of conditions, e.g. the error is not intended by the person, it is not desired either by a set of rules or an external observer, and it led the task or system outside of its acceptable limits (Senders and Moray, 1991).

1.3 How are errors classified?

One of the earliest reported studies of human error was carried out by Kollarits (1937). He compiled a database of errors committed by himself, his wife and his colleagues and from this generated a phenomenological four-way classification. Errors were classified as being in one of the following categories: substitution, omission, repetition, insertion. However, Kollarits noted that it was actually quite difficult to assign errors unequivocally to one category. Phenomenological classifications like this focus on categorising errors according to their consequences. This is in contrast to classifications based on the underlying causes (see Prabhu and Prabhu, 1997). However, this approach is also beset with problems. For example, some work we did at Bristol into classifying errors made by library database users resulted in a number of problems (Sheard *et al.*, 2000). It was possible to know what error had been made, e.g. confusing two function keys, but not the reason why the error had occurred. Hence, was function key F4 confused with F11 because of memory failure, or attention lapse. When classifying errors, there have been two main approaches taken: one, to focus on the consequences of the error, or two, the underlying causes.

When considering defining different errors, it soon becomes apparent that the word 'error' is very general, and in fact, encompasses all of the types of error. One commonly used classification is to divide errors into three basic types according to Rasmussen's skill, rule and knowledge-based model of human performance (Rasmussen, 1986). From this model, three basic error types emerge. These are skill-based slips (and lapses), rule-based mistakes and knowledge-based mistakes. Each is discussed here in turn.

1.3.1 Slips and lapses

These occur when actions deviate from the intention, e.g. forgetting to boil a kettle and making a cup of tea with cold water, or attempting to fill a teapot with hot water when it has already been filled. Reason and Mycielska (1982) categorised these types of failures as slips and lapses. Whereas slips and lapses are both unintentional actions, they differ in terms of whether or not they can be observed. A slip is an error that is observable, e.g. a slip-of-the-tongue. In contrast, lapses tend to be covert, e.g. forgetting to carry out an action and, consequently, they may be apparent only to the person. It can be seen that slips and lapses are execution failures. The intention was correct for the situation, but the task failed to be carried out. With reference to Rasmussen's model, slips and lapses arise when behaviour is skill-based, i.e. the processes of carrying out the task have become automatic. Riding a bicycle would be an example of skilled-based behaviour.

1.3.2 Mistakes

These are errors that arise from failures in planning (Norman, 1983). The actions may be correctly executed but the original plans for achieving the goal are incorrect, e.g. driving the wrong way down a one-way street. Again, referring to Rasmussen's model, mistakes would be divided into two types: those that are rule-based and those that are knowledge-based. Rule-based mistakes are those errors that arise when a situation or an action has associated rules that need to be applied and this has not happened, e.g. grammatical mistakes in written English. Knowledge-based mistakes, on the other hand, arise when an individual has to apply their knowledge and understanding to a situation, e.g. when having to make decisions perhaps under time pressure and in a safety-critical situation.

Implicit in this discussion of error types is whether or not the action was intended, i.e. whether the individual had deliberately tried to make an error. When the error is intentional, it is often referred to as a violation. Violations have been defined as 'any deliberate deviations from rules, procedures, instructions and regulations' (Health and Safety Executive, 1999: 16). In the workplace, violations are important because of their implications in terms of compromising safety. The Health and Safety Executive (*ibid.*) have divided violations into three categories: routine, situational and exceptional. Routine violations occur when the rules in the workplace have been broken to such an extent that the way of working is now common practice and new workers may not even be aware that they are violations, e.g. fixing machinery when it breaks down rather than calling maintenance. These routine violations may occur because they provide short cuts and save time and effort, or because workers feel the rules are inappropriate and so ignore them. Situational violations occur when workers break the rules because of pressures arising from a particular job, e.g. safety equipment may not be available and workers may be under time pressure to complete a task, so they take risks by going ahead without it. The Health and Safety Executive (*ibid.*) gave the example of an unharnessed steel worker who fell 20 metres and was killed. It transpired that harnesses were available, but there was no facility for fixing them and there were no other safeguards. The third category of exceptional violations is much rarer. Exceptional violations occur when workers have honourable intentions but indulge in behaviour that places them at high risk. For example, the control room operators at the Chernobyl nuclear power plant continued to carry out pre-planned tests rather than attend to a situation that was rapidly becoming critical and which was placing them and the environs at great risk. It can thus be seen that the third category of exceptional violations is particularly hard to manage.

In summary, studies that have defined and categorised errors help us to understand more about the nature of the errors and their causes. They are also useful in that they inform our comprehension and knowledge about how people behave, especially in the workplace. This is particularly important in the case of violations. This greater understanding can then be used to good effect when considering the design of control rooms and training programmes for the management of human error.

1.4 What are the causes of errors?

Classifications of different types of errors help provide an indication of their causes. The common factor is always the involvement of humans, because of the major role we play in the design, management and maintenance of systems. Therefore, it is reasonable to conclude that at some point, errors are occurring because of the fallibility of humans. As already stated, humans make frequent errors leading to a variety of consequences. They are therefore a function of human information processing and arise as a result of a difference between what the human is trying to achieve and what they actually do. Certainly, Rasmussen's model worked on the premise of errors arising from failures in planning and execution. Hence, one common view is that human errors arise because of a mismatch between the human and the task. If the human is operating a piece of equipment, frequent misfits are usually thought to be due to errors in the design of the system, while occasional misfits are more likely to be due to human errors (Rasmussen, 1987). This indicates that errors are either caused by factors external to the person, e.g. poor design of the workplace and/or equipment they are using, or by the individual themselves.

There is a school of thought that states if errors have causes, they are lawful and therefore, predictable. Hawkins (1993) highlighted this point in his classification of errors where he divided them into random, systematic and sporadic. He illustrated this graphically by considering rifle shots at a target. Random shots, as the term suggests, pepper the target. This type of error would be particularly difficult to predict in everyday life, e.g. the car driver who makes a number of unrelated mistakes without any recognisable pattern. Systematic shots are the ones that would group together to one side of the target, say, to the left of the bull's eye. In everyday life, this would the error that we repeatedly make, e.g. the car driver who scratches the car on the same gate post on a number of occasions. Sporadic shots, on the other hand, combine both the features of random and systematic shots. These shots will be found across the target similar to the random shots but with some degree of grouping akin to the systematic shots. Hawkins suggested that in everyday life, the sporadic errors make up the group that are the hardest to predict and, because of this, they will be the most difficult to manage. In the car driver example, this would be the person who normally reverses into the garage without any problems, but then inexplicably makes an error by driving too close to the wall and in doing so, marks the bodywork of the car. It can thus be seen that of the three error types, the systematic errors could be predicted given the consistency of the person in making this type of error.

1.5 Do errors lead to accidents?

Some of the negative consequences of errors include incidents and accidents. Senders and Moray (1991: 104) defined an accident as sometimes being 'a manifestation of the consequence of an expression of an error'. However, just as errors do not always lead to accidents, accidents do not always arise from errors. One of the difficulties of addressing whether errors lead to accidents is associated with the methodological and practical problems of studying the causes of accidents. The study of human error is very different from researching human behaviour based on simple, rule-based decisions as often investigated in the field of experimental psychology. However, a number of approaches are possible: accident data and post-hoc accident analyses, simulation of the accident, laboratory studies on accident models and errors, and self-report techniques, e.g. where people are part of a 'no blame' reporting system. Despite this plethora of methods, none is entirely satisfactory in providing quality evidence and definitive information about the relationship between errors and accidents. Take, for example, post-hoc analyses of aircraft accidents. In a fatal crash, a number of the key witnesses may have died, or vital information about the cause of the accident may have been lost in the damaged hull. Despite these shortcomings, analyses of accident data in conjunction with observation and self-report techniques have been reported as the most frequently used methods of studying errors (Nagel, 1988). For example, one such study analysing 338 underground mining accidents found that 35 per cent were attributable to human error and none was solely the result of errors made by humans (Sanders and Shaw, 1988).

Analyses of major accidents have allowed some interesting findings about errors to be uncovered, e.g. the fact that it is very unusual for one error to lead to an accident, but rather several factors come together to create a causal chain. This can be demonstrated by considering the following two examples of accidents. The first, Three Mile Island, occurred over 20 years ago but since that time, has been extensively researched and documented. The second concerns a recent rail crash that occurred at Hatfield in the UK.

Three Mile Island was the location of a nuclear power plant in Pennsylvania on the east coast of America, where in 1979, a major incident occurred. One of the relief valves that should have operated automatically had not opened, and the flow of coolant water to the main pumps was interrupted. For over two hours, the control room operators fought to isolate the problem and then, to solve it. Their concern was a meltdown of the reactor or at very least, an escape of radiation into the environment. The emergency continued for over 16 hours, but the operators managed to prevent loss of life, although a small amount of radioactive material was released into the atmosphere. The cost to the operating companies and insurers was estimated to be in the region of one billion dollars. The subsequent Nuclear Regulatory Commission enquiry found that although the operators did make a number of errors, there were other problems associated with the design of the control room and the operating procedures (Malone *et al.*, 1980). For example, when the emergency first occurred, the operators were overloaded with information with a large number of displays flashing and auditory alarms sounding. The control panels were poorly designed since the alarms were not organised in a logical manner, so the operators had to scan 1600 windows and gauges. Several of these displays had gone 'off scale' so were not providing any useful information other than indicating the seriousness of the situation. The emergency was exacerbated by a printer failure which meant data was lost that would have helped indicate the problem, and when the printer was operational again, it was running more than two hours behind events as they were happening.

At Hatfield a broken piece of rail track was thought to be the primary reason for derailing a high-speed passenger train on October 17, 2000, resulting in the deaths of four people. However, it was not clear whether the track had broken before the accident or was a result of the crash (Perry and Morris, 2000). Subsequently, it was found that the railway company had been aware of the broken track for at least 10 months before the accident (Harper, 2001). This railway accident was just one of several that the UK rail industry had experienced over the last few years. On October 5, 1999, an express train crashed into a commuter train just two miles west of Paddington station, London. It was thought the driver of the commuter train had passed through a red warning signal. Twenty-eight people died and around 150 were injured and required hospital treatment (Harper, 1999).

In both these examples, it becomes apparent that there were a number of facets of the situation that led to the accident. Some of these were already present when the accident occurred and some were created on the day. This is not unusual and extensive reviews of other accidents, e.g. Bhopal, Challenger, Chernobyl, King's Cross and the Herald of Free Enterprise have indicated similar findings (see Reason, 1990; O'Hare, 2001). A medical analogy has been used to refer to the problems that already exist in the system as 'resident pathogens'. For most of the time, these pathogens are kept in check by other parts of the system – similar to the immune system protecting the body from viruses. However, there will be times when the system breaks down, i.e. the person succumbs to illness. In a complex, technological system, these resident pathogens are referred to as latent failures. They contrast with the 'active failures' that occur at the time of the accident or incident.

One of the main conclusions from accident analyses of major events is that they are usually created by the coming together of several causal chains, where the elimination of any one would have deflected the sequence of events leading to the incident. Reason (1990) described this in his 'dynamics of accident causation' model that is presented in modified form in Figure 1.1. It can be seen that a number of latent and active failures emanating from a number of causes come together to produce the 'impossible' accident.

1.5.1 Three Mile Island

- **Organisation** – grossly inadequate operating procedures, insufficient training.
- **Workplace** – poor layout of controls; those needed were positioned away from the centre of primary activity.
- **Human–machine interactions** – poorly designed control panels, information overload, printer problems.
- **Other personnel** – maintenance tags had been left covering some of the relevant controls.

1.5.2 Hatfield

- **Organisation** – despite expressing concerns, the company responsible for track maintenance had not reduced the number of broken rails nor had they introduced the standard rail used in Europe – the UIC60; this is a stronger rail and requires less maintenance. More specifically, two recommendations to replace a suspect section of track where the derailment occurred had been ignored.

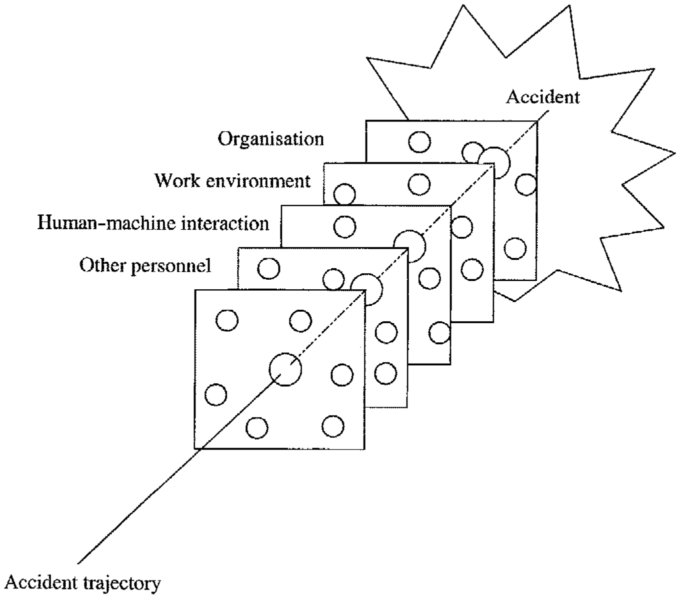


Figure 1.1 Model of causes of accidents.

- **Work environment** – new locomotives were believed to be the reason for an increase in 'corner gauge cracking' but management had not taken any actions concerning this.
- **Traffic** – In addition, the growth in rail traffic, particularly heavy freight, had had a major impact on track deterioration.
- **Human–machine interactions** – the track at Hatfield was known to be poor, but no speed restrictions had been introduced. Following the accident, speed restrictions of up to one-third of the permitted speed were imposed on bends for high-speed lines throughout the UK.
- **Other personnel** – the track at Hatfield was due to be inspected by Health & Safety on the day of the accident.

Despite the association of happenstance with accidents, it can be seen from analyses of Three Mile Island, Hatfield and other disasters that it is quite unusual for an accident to occur as the result of a single action (e.g. one error). Although there will be active failures at the time of the accident, there will also be a myriad of factors leading up to the accident. These may result from latent failures resulting from poor design, insufficient training, inadequate procedures, system design (hardware and software), regulatory, management and operational failures, as well as communication difficulties and maintenance activities. A further difficulty is deciding how far back in the time frame to go, since there may be resident pathogens present in the system for many years that are of no consequence until they become combined with other failures. For example, the problems with the quality of the rail track at Hatfield may in fact result partly from the lack of investment in the UK railway system that has occurred over many decades.

The fact that many factors lead to accidents is important since this multi-causal aspect is often overlooked in a culture where the tendency is to blame an individual. In the subsequent NRC (Nuclear Regulatory Commission) report on Three Mile Island, some blame was placed on the operators. It is also likely that with Hatfield, blame will be placed on the person key to the accident when it happened, i.e. the driver. This was certainly the case with the Herald of Free Enterprise, a car and passenger ferry that capsized outside of Zeebrugge in March 1987, with significant loss of life. The ferry worker who failed to close the bow doors quickly enough was identified as the person responsible for the accident whereas it could be argued that the resident pathogens were already in the system. For example:

(i) the basic design of the 'roll-on/roll-off' ferries had originally been a cause of concern since it only needed a small amount of water across the deck and the whole structure would capsize;
(ii) the poor safety culture on the ferries – another deckhand had been asked to close the doors but had responded that it was not his job;
(iii) the fact that management had recently rejected a proposal (on the grounds of cost) to have warning lights to indicate when the bow doors were open;

(iv) in the Ship's Standing Orders, there was no reference to closing the doors before sailing, i.e. there was no company policy concerning the doors;
(v) the undermanning of the ferries resulting in workers on double-shifts – the person whose job it was to close the doors was actually asleep at the time having already worked a shift;
(vi) the time pressures on the ferry crew to keep to schedule so that they were leaving port with the doors still open;
(vii) the poor weather conditions combined with high spring tides on the actual night of the accident.

The concern with blaming the individual who was the last to have contact with the system is the belief that the removal of that person will actually make the system safe again. It can be seen from the accident model in Figure 1.1. that this is not the case. The holes, i.e. the resident pathogens/latent failures are still there, and if they are not addressed, the 'impossible accident' will occur again. However, it should also be noted that after many catastrophic events, the situation becomes far safer. After Three Mile Island, the design and procedures of nuclear power plants were extensively reviewed, with the result that nuclear power in the western world is now far safer today than it was at the time of the accident. Major events such as the Herald of Free Enterprise and Hatfield tend to become headline news and are usually followed by national enquiries into their causes in order to make recommendations to avoid similar incidents happening in the future.

1.6 How can we reduce error?

As already mentioned earlier in the chapter, there are two main ways of addressing the reduction of error in the workplace: one, by considering external factors relating to the design of 'the system', and two, by focusing on internal aspects relating to the human. The former would involve the study of external performance-shaping factors such as equipment and workplace design, human–machine interactions, operating procedures, environmental and organisational considerations and job aids, e.g. manuals and supporting documentation. The latter relates to the human, e.g. individual differences such as levels of motivation, skills and experience. These internal performance-shaping factors can be addressed via selection and recruitment procedures, and training programmes. As an example, the workers should be trained to admit the possibility of making an error and to acknowledge it when they do. Open-reporting systems such as this allow the management to find out where the errors are being made and to address the situation before an accident or incident occurs. An extension of this is the no blame reporting system as found in the avionics environment where 'no blame' is attached to an admitted mistake in contrast to one that is subsequently found out. However, it is generally thought that internal traits are not as influential as external factors in contributing to human errors; it is therefore vital that systems are designed to optimise human interactions and to reduce opportunities for making errors.

In the post-war years, technological developments boomed. The philosophy of this time was to automate systems as much as possible and to replace as many human activities as possible by machines (Birmingham and Taylor, 1954). Part of the underlying philosophy was to prevent humans making errors by ensuring that the machines carried out as many activities as possible. This resulted in a situation in the 1960s and 1970s where machines carried out all the jobs apart from those that were too difficult to automate; this was referred to as the 'left-over' principle by Hollnagel (1999). Consequently, opportunities for human error still existed. An advance on this approach was to consider the attributes of humans and machines, and to allocate function according to the strengths and weaknesses of each. This became known as the 'compensatory' principle (Moray, 1999). More recently, in the 1980s and 1990s this simple comparison approach has been replaced by a more integrative one that takes into account the overall goals of the system and the tasks needed to achieve them (Singleton, 1989). This more recent approach has been referred to as 'complementary' allocation. The increase in complex systems has not only had a number of implications for automation but also with regard to managing human error. There has been a shift from designing systems that attempted to prevent humans from making errors to those which take into account this aspect of human fallibility. Hence, the current approach is to design 'error-tolerant' systems (Billings, 1997). There are a number of ways that systems can be designed to be error-tolerant; for example, redundancy can be built into critical operations. A simple example of this is the confirmation question that appears in response to a request to delete files in a word processing application. A further design technique is to allow graceful degradation of systems, so if errors have been made and/or systems malfunction, they very slowly decline and thus allow the human operator to be kept 'in the loop' about the current status of operations. This is in comparison to systems that just shut down and then return control to the human. A third example concerns the implementation of a reversion technique into the system, so if errors are made, the system always has some means of backup in order to ensure the goals of the operation can be maintained. A number of guidelines for system design exist in the human factors engineering literature; for example, Smith and Mosier (1986) reported 944 such recommendations and many of these facilitate the development of error-tolerance. These three examples given here represent just a tiny fraction of the volume of information that is currently available for designing systems according to recognised ergonomic principles. Although there are a number of tried and tested ways of reducing opportunities for errors in the workplace, errors (and errors that lead to accidents) will still inevitably occur. It is impossible to prevent individuals from making errors and, unfortunately, some will have detrimental effects.

1.7 Conclusions

Making errors is a natural part of human behaviour and it is debatable whether we would want to prevent humans from making errors, because most are inconsequential and allow us to learn from the situation. This can prove to be of benefit since it prevents us in the future from making the error that has the more serious consequences. It is also evident in the change in approach that has taken place over the decades in system design. In the 1970s, systems were designed to suppress errors, i.e. operators were not allowed to make errors, in contrast to the error-tolerant systems of the 1990s that recognise that humans make errors and there is a need to accommodate this. However, despite advances in comprehending human performance, it is probably reasonable to conclude that psychologists still have a limited understanding of human error. Certainly, the link between errors and accidents is not straightforward. This is partly because accidents (and incidents) are nearly always multi-causal, with active and latent failures coming together at a particular point in time. When considering the reduction of human error in the workplace, there are two main approaches. These are either to address the external personal shaping characteristics, e.g. the design of systems, human–machine interactions, workplace and environmental design, or the internal aspects of the human, e.g. through selection, recruitment, training and the engendering of a strong safety culture. In conclusion, managing human error warrants a holistic approach that considers an array of aspects relating to the workplace and the individual.

Acknowledgement

Parts of this chapter are based on an earlier paper by the author: 'Managing error' in the *Proceedings of the UKACC International Conference 'CONTROL '98'* (pp. 578–583) published by the IEE, London.

References

1 AMALBERTI, R.: Unpublished internal report, 1997
2 AMALBERTI, R., and WIOLAND, L.: 'Human error in aviation', *in* SOEKKHA, H. M. (Ed.): 'Proceedings of International Aviation Safety Conference '97' (VSP BV, Rotterdam, Netherlands, 1997) pp. 91–108
3 BILLINGS, C. E.: 'Aviation automation: the search for a human-centred approach' (LEA, New Jersey, 1997)
4 BIRMINGHAM, H. P., and TAYLOR, F. V.: 'A human engineering approach to the design of man-operated continuous control systems'. Report No. 433, US Naval Laboratory, Washington DC, 1954
5 HARPER, K.: 'Inquiry to focus on driver error', *The Guardian*, October 6, 1999, p. 1
6 HARPER, K.: 'Hatfield: The warnings that Railtrack ignored', *The Guardian*, January 16, 2001, p. 1
7 HAWKINS, F. H.: 'Human factors in flight' (Ashgate, Aldershot, 1993, 2nd edn.)
8 Health and Safety Executive: 'Reducing error and influencing behaviour' (HSE Books, Norwich, 1999)

9 HOLLNAGEL, E.: 'Control versus dependence: striking the balance in function allocation', *CSERIAC Gateway*, 1999, **IX** (4), pp. 12–13

10 KOLLARITS, J.: 'Beobachtungen über Dyspraxien (Fehlhandlungen)', *Arch. für die Gesamte Psychologie*, 1937, **99**, pp. 305–99

11 MALONE, T. B., KIRKPATRICK, M., MALLORY, K., EIKE, D., JOHNSON, J. H., and WALKER, R. W.: 'Human factors evaluation of control room design and operator performance at Three Mile Island–2'. Report prepared for the Nuclear Regulatory Commission by the Essex Corporation, Fairfax, VA, 1980

12 MORAY, N.: 'Human operators and automation'. Proceedings of *People in Control* Conference, Publication No. 463 (Institution of Electrical Engineers, London, 1999), p. 8

13 NAGEL, D. C.: 'Human errors in aviation operations', *in* WIENER, E. L., and NAGEL, D. C. (Eds.): 'Human factors in aviation' (Academic Press, San Diego, CA, 1988)

14 NORMAN, D. A.: 'Position paper on human error'. Proceedings of *NATO Advanced Research Workshop on Human Error*, Bellagio, Italy, 1983

15 O'HARE, D.: 'The "Wheel of Misfortune": a taxonomic approach to human factors in accident investigation and analysis in aviation and other complex systems', *Ergonomics*, 2001, **43**, (12), pp. 2001–19

16 PERRY, K., and MORRIS, S.: 'Inspection of site was due on day of crash', *The Guardian*, October 19, 2000, p. 6

17 PRABHU, P. V., and PRABHU, G. V.: 'Human error and user-interface design', *in* HELANDER, M., LANDAUER, T.K., and PRABHU, P. (Eds.): 'Handbook of human-computer interaction' (Elsevier Science BV, Amsterdam, 1997, 2nd edn.) pp. 489–501

18 RASMUSSEN, J.: 'Information processing and human-machine interaction' (North-Holland, Amsterdam, 1986)

19 RASMUSSEN, J.: 'The definition of human error and a taxonomy for technical system design', *in* RASMUSSEN, J., DUNCAN, K., and LEPLAT, J. (Eds.): 'New technology and human error' (Wiley, Chichester, 1987)

20 REASON, J. T., and MYCIELSKA, K.: 'Absent-minded? The psychology of mental lapses and everyday errors' (Prentice-Hall, Englewood Cliffs, NJ, 1982)

21 REASON, J. T.: 'Human error' (Cambridge University Press, Cambridge, 1990)

22 SANDERS, M. S., and SHAW, B.: 'Research to determine the contribution of system factors in the occurrence of underground injury accidents' (Bureau of Mines, Pittsburgh, PA, 1998)

23 SENDERS, J. W., and MORAY, N. P.: 'Human error: cause, prediction, and reduction' (LEA, New York, 1991)

24 SHEARD, M., NOYES, J. M., and PERFECT, T. J.: 'Older adults' use of public technology', *in* MCCABE, P. T., HANSON, M. A., and ROBERTSON, S. A. (Eds.): 'Contemporary ergonomics 2000' (Taylor and Francis, London, 2000) pp. 161–65

25 SINGLETON, W. T.: 'The mind at work: psychological ergonomics' (Cambridge University Press, Cambridge, 1989)

26 SMITH, S. L., and MOSIER, J. N.: 'Design guidelines for user-system interface software'. Report MTR-9420, Mitre Corporation, Bedford, MA, 1986

Chapter 2
Memory and complex skills

Malcolm Cook

2.1 Introduction

Supervisory control and command systems are knowledge-rich environments where large bodies of information must be processed to successfully function and these can only be interpreted within frameworks relating to the task environment, the organisation operating the system, the team managing the task and the specific tasks themselves. When people learn to use complex systems they develop experience with the system in a number of domains and in a number of ways. Initially the knowledge acquired is largely concerned with facts related to the system and gradually as experience develops this is associated with procedural knowledge, describing how to do certain things. The operator's task is to learn as much about the system as possible and create a mental model of system operation in long-term memory. Gradually the operator develops an appreciation of what is normal and what is exceptional in respect of system performance, with the appropriate cues and patterns of information stored in long-term memory.

Many changes take place as the operator learns to use a new system and some of the changes that occur relate to the information processing processes the operator is expected to carry out. With experience the operator is more able to attend selectively to the important signals and events that are critical for successful or safe system operation. By developing 'an eye' for the likely problems the operator reduces their workload in normal operation and increases their ability to manage abnormal states. The trend towards automaticity is not specific to control tasks but applies to all aspects of skilled behaviour. As experience with system operation continues to increase many operators find that they can remember greater amounts of information and this reflects a shift away from using their short-term memory to directly accessing information, concerning system function, from long-term memory. The changes in attentional strategy and memory function are interrelated because attention may be driven by expectations derived from past experience. Memory retrieval can equally be triggered by salient events that attract attention.

There are a number of models currently in use in human factors that make direct and indirect reference to memory functions, e.g. Mica Endsley's (1995; 1996; 2000) model of Situation Awareness, Jens Rasmussen's model of Skill Acquisition (1983; 1986), and Gary Klein's (1997a) model of Naturalistic Decision-making. Analysis of these models reveals some potentially hidden assumptions in their application and explicit recognition of these assumptions may generate critical hypotheses about the nature of human skilled performance in the operation of complex supervisory control systems that can be assessed by experimental analysis. This chapter aims to discuss the way these models of memory can be applied to a wide range of issues in human factors, and control rooms, in particular. It is assumed that experts remember task-relevant information better, have better representations of problems, have superior knowledge of task-relevant material, achieve expertise via extensive practise and are not necessarily superior in basic information processing capacities, when compared to the general population from which they are drawn (Green and Gilhooly, 1992).

2.2 Situation awareness

It has been argued that the active information on current events stored in memory is vital to understanding the reasons for erroneous decision-making and poor supervisory control performance. The user's current mental model of the situation has been termed the *situation awareness* of the user and it represents both current and previous acquired knowledge related to the on-going series of events. In her recent critical review of situation awareness Endsley (2000) identified the role of working memory in situation awareness but did not identify a specific model with which she associated the properties of situation awareness. This is clearly a significant issue because different models of short-term memory, working memory and information processing have different mechanisms that can be used to make different predictions about the factors shaping performance (see Campbell *et al.*, 1999). By not identifying a specific model of memory it is impossible to predict the ways in which performance is limited and the ways in which performance may fail. However, Endsley did identify two critical properties defining the effect of working memory on situation awareness and a similar analysis has been proposed by Cook *et al.* (2000):

- Limited working memory and attentional capacity.
- Limited ability for processing in novices compared to skilled individuals.

These limitations to human information processing are typical and identified to a greater or lesser degree in most human factors analyses of memory performance in respect of human performance (Wickens, 1992; Wickens *et al.*, 1998; Wickens and Hollands, 2000). Endsley's situation awareness model of working memory seems to incorporate elements of Baddeley's (1986) working memory model, in that she emphasises the role of capacity and attention, which suggests she is making reference to a combination of stores and a central processing executive. Baddeley's model of working memory can be criticised for its all-encompassing nature and some have suggested it describes the entire information processing system surrounding the short-term storage of information (Carlson, 1997). However, Endsley (2000) also made reference to changes in memory performance that occur as skill develops and she distinguishes strategy changes in processing from other changes that are determined by long-term experience. Even though there is no explicit reference it is possible that the changes are related to what is termed long-term working memory by Ericsson and Delaney (1999). Later in the same chapter, Endsley (2000) does reference the work of Ericsson and Kintch (1995) and notes that they have proposed the use of pointers in working memory which reference data held in long-term memory. A more detailed analysis of long-term working memory has recently appeared in Ericsson and Delaney (1999) and many of the properties of skilled performance discussed by Endsley (2000) are typical of long-term working memory characteristics. This later review provides a useful overview of the properties of long-term working memory; these reflect the more obvious properties of memory use in skilled behaviour typically described in reviews (cf. Proctor and Dutta, 1995; Carlson, 1997).

Endsley (2000) suggests that viewing situation awareness as either a function of working memory or long-term memory would be inappropriate. She continues by suggesting that situation awareness and working memory function is an activated subset of long-term memory, citing Cowan (1988) in support of this view. Cowan (1993; 1998) interprets working memory as currently active information and distinguishes two elements, the current focus and the background information. This view of working memory functions implicitly suggests that situation awareness will be better represented in the cognitive functions of skilled individuals because they will have created schematic representations that can be identified with current experience, to allow more effective encoding (Cook, 2000a). Endsley's global cognitive resources approach to situation awareness reflects a view that identifies it as an emergent property of coherence in function across many aspects of cognition and not as a property of an individual system. In this sense the first order effects related to the more effective use of memory are likely to be accompanied by second order effects in relation to the organisation of cognitive function that reflects meta-cognitive information processing management. Or, the effectiveness of the first order processing may be determined by the effectiveness of meta-cognitive process management skills. Endsley (2000) includes references to the use of mental models, schemata and scripts and various other forms of internal knowledge management, in her model of situation awareness. This is clearly not a new idea but Endsley uses the perceived requirement for knowledge organisation, and mental models in particular, to explain the central resource errors related to weak mental models, use of an inappropriate mental model and over-reliance on default values for the mental model, derived from frequent experience.

Endsley suggests that the original assessment of a situation generates a hypothesis or hypotheses that are used to guide the application of cognitive resources and bias the direction of processing. For example, the limited attentional capacity is more effectively applied when experience is used to focus the resource and similarly with regard to limited working memory capacity. This encourages a tendency for human decision-makers to stay with an original hypothesis that has been recognised for many years and it can be explained in a number of ways: either (a) the decision-maker develops a hypothesis and having invested cognitive effort in the hypothesis they are reluctant to change their view, as the change represents a further effortful process; or (b) accepting the need to change creates an aversive affective state because it suggests the initial assessment was wrong and uncertainty exists concerning the actual state of affairs. Thus, the decision-maker avoids the more aversive cognitive state of uncertainty but if they accept that, they would be less aware if they started task shedding to tackle the cognitive re-structuring task. In combination the demand for cognitive resources and the aversive affective state may influence the tendency to stay with a current hypothesis until it is clearly no longer tenable, a fact that has been well-recognised in the literature on thinking for many years (Gilhooly, 1983; Huey and Wickens, 1993). When the anxiety about the current interpretation is greater than that associated with the switch to searching for a new one, the decision-maker will consider alternatives. If the decision-maker delays a long time they may be in a heightened state of arousal that makes their processing even more inefficient and the burden of managing an unusual situation, that does not fit into available schema, is overwhelming.

Endsley (2000) accepts that pattern-matching appears to explain some aspects of situation awareness and this indicates another role for memory in retaining and returning the cues for identifying the pattern of events. There is a considerable body of work in a number of fields that support the view that expert decision-makers use pattern-matching from a number of potential stores in facilitating decision-making and improving the management of tasks. It is not evident if the pattern-matching that takes place is based on semantic memory, episodic memory, procedural memory, implicit memory, or a combination of these. Thus, it is not clear if the process is actually open to conscious evaluation even if the results of the process are and the *feeling* about the quality of the judgements is. It is certainly the case that some authors such as Gardiner and Java (1993) distinguish memory systems in terms of the contribution from explicit and implicit cognition. Meta-cognitive analysis of information processing management suggests that decision-makers can both express a decision and qualitatively judge how certain they are about the judgement. This distinction has been used in recent analyses of command team performance (McGuiness, 2000) and with regard to the analysis of systems (McGuiness *et al.*, 2000).

In addition to the initial situation assessment, Endsley (2000) recognises that human operators have preconceptions and expectations that influence the formation of situation awareness. This is clearly another area in which memory may influence the process of decision-making because those events in the task environment that are most frequent will bias the assessment, prior to commencement of the actual task. Clearly, from what is known of memory function, frequency will have a pronounced effect and highly improbable events will be discounted from consideration. The focus on frequent events has value in that it places attention and memory on events that are more likely, i.e. probabilistically will be correct a highly significant amount of the time. The problem is that some of the time the preconceptions will be wrong and a delayed response might be catastrophic.

While automaticity is more a property of the information processing process it is possible to identify it with changes in memory function because it implies access to memory with reduced conscious awareness. Endsley (2000) views this as potentially hazardous in that it may be difficult to extend the skilled behaviour when the link between perception and action is less effectively accessed from awareness and there may be a tendency towards reactive behaviour when automaticity occurs. It is significant that the shift towards skilled behaviour usually involves automaticity, use of long-term working memory and the application of expectations. This tendency towards cognitive resource short-cuts has been identified with the use of biases and heuristics that may be both advantageous and dangerous (Huey and Wickens, 1993).

2.3 Skill-based, rule-based and knowledge-based performance

In considering the relationship between memory and complex systems it is clear that the development of knowledge is an important issue because this informs best practice for training needs analysis. Some appreciation of the skilled memory development appropriate for system operation may help to define levels of skilled performance appropriate for managing certain levels of system responsibility. In addition, the appreciation of the memory demands relating to different components of the skill deployment may provide measures of skilled performance and priorities for the operator selection process. To successfully perform such an analysis there is a need for a framework to describe the general changes in performance and Rasmussen (1983; 1986) has provided such a theoretical framework. Rasmussen proposed a theory of skilled performance in which three modes of performance are defined and identified as knowledge-based, rule-based and skill-based. The knowledge-based behaviour typifies the early period of skill acquisition in particular and it is characterised by verbal mediation and conscious control. Knowledge-based behaviour can additionally occur late in skill development when an expert meets an unfamiliar or novel situation and is forced to examine the task environment carefully for cues. It is possible that as the task environment becomes more familiar that the operator can become aware of rules that match situational cues to actions which typify the rule-based mode of skilled behaviour. In time even this matching process is replaced by an intuitive, automatic processing of information typical of the skill-based mode of operation. The final stage of skill acquisition is marked by the capability to perform multiple tasks with minimum task interference and lower levels of subjectively experienced workload.

Thus, the process of skill acquisition seems to change from a consciously controlled process heavily dependent on working-memory resources to one dependent on a non-conscious long-term memory store of previous experience. This shift from conscious to non-conscious processing allows quick and automatic classification of the situation and execution of well-rehearsed sequences of actions that manage the task environment. This type of shift parallels the distinction of explicit memory functions and implicit memory functions described by Graf and Schacter (1985), Schacter (1987) and Dienes and Altmann (1997). Paradoxically, this may result in disadvantages that are typical of learning associated with skilled performance such as strong binding of knowledge to action and lower transfer of learning from one context to another (Healy and Bourne, 1995; Proctor and Dutta, 1995). This may in part explain observations that experts are significantly poorer at managing unfamiliar and unexpected situations because they re-apply logic from experience inappropriately (see Huey and Wickens, 1993).

A major concern with complex systems is the tendency for automation and agent-based protocols at the interface to force the operator to work at a rule- or knowledge-based level (Norman, 1990). Thus, even if the operator is an expert they find that working in co-ordination with automation they must work in a less effective manner. This may not be as designers intended, in that workload is higher than planned, but it may prevent the development of fixed behavioural routines, that may generate errors. Clearly, the emphasis in Rasmussen's model is a trend towards lower working memory demands as skills develop and emphasis is transferred away from the capture of all the likely information towards the selective analysis of prioritised information, on the basis of past experience. Rasmussen's model has many similarities with that of Endsley's (2000) latest iteration of the situation awareness model, because it explains the ability to operate more effectively while working with less effort.

A major element of Rasmussen's model is the adaptation of the processing mode to the dynamics of the environment and familiarity with the specific task situation, leaving an opportunity to recognise the context sensitivity that distinguishes human skilled performance from that of machine intelligence. A significant gap in the model is the management of the shift in processing from one mode to another and the presumably meta-cognitive processes that control the changes. In the amended model proposed by Cook and Elder (1998), affective and cognitive cues are used to indicate the need for a shift from one mode of processing to another and these changes may determine the management of memory. Cook (1999) has proposed that elements of the operator's trust related to factors such as confidence, uncertainty and predictability may prompt over- or under-reliance on automation. Some of the affective cues may actually reflect the results of memory searches because the failure to identify quickly representative schema matching the target situation may trigger the shift in processing mode downwards from skill-based to rule-based, and further to knowledge-based processing. The issue of shifting between modes of processing is discussed in Cook (2000a; 2000b) and the role of self-monitoring is discussed in Cook (2000c), and Cook *et al.* (1998).

The modified model of Rasmussen's skill operation shown in Figure 2.1 is very similar in content to a sophisticated model proposed by Bower and Forgas (2000) for social judgements, in recognising the dimension of familiarity as a modifier for meta-management of information processing. There are clearly situations in which the circumstances are very familiar and situations, at the other end of the continuum, where the context is novel and highly unfamiliar different information processing strategies are required. It is significant that Bower and Forgas (2000) highlight the importance of memory processes and the type of search strategy used in examining memory as a variable, according to the familiarity of the contextual cues.

2.4 Naturalistic decision-making

Klein (1993a; 1993b; 1997a; 1997b) has over a number of years developed and refined an analysis of decision-making which aims to explain real-time high stakes decision-making in real-world tasks and has identified it as a credible alternative to previous models of decision-making. Previous models, before naturalistic decision-making, were based on normative models and judgements of contrasting utility, with limited ability to explain real-world decision-making. Klein (2000a; 2000b), like Rasmussen (1983; 1986), describes a number of ways in which the information

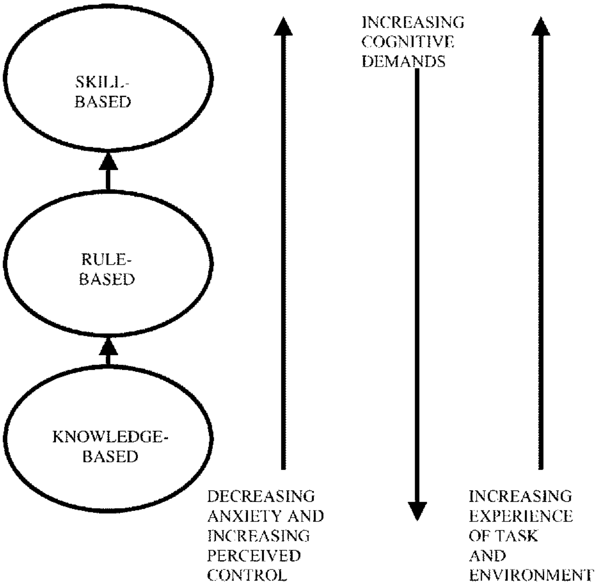


Figure 2.1. *A modified representation of Rasmussen's Skill Acquisition model emphasising the changes in affective state, demand on cognitive resources and experience of the task and the task environment.*

processing varies according to the circumstances but in Klein's case this relates to decision-making. Implicitly there are similar issues with variation in processing in both Klein's (2000a; 2000b) model and that of Rasmussen (1983; 1986) because the variations in processing are determined by the degree of match between expectancies and evaluation of the likely success of the potential action considered. The latter element, relating to the consideration of action efficacy is broadly similar to the projection phase of situation awareness (see Endsley, 2000). Most of the processing is largely defined by access to information stored in memory and this is defined as expectations, relevant cues, plausible goals and actions. Klein's model seems implicitly to endorse the tight binding of knowledge and action identified in the analysis of skilled activity in both perceptual-motor (Proctor and Dutta, 1995) and cognitive domains (Healy and Bourne, 1995; Carlson, 1997).

The information used in decision-making with regard to expectation concerns the result of a matching process between the information stored in memory and the cues from the environment. This seems to be an iterative process where the context creates expectations and a search for matches to the appropriate cues takes place, which in turn creates further expectations. The definition of the situation held in current memory helps to identify plausible goals in the current perceived situation and this allows the selection of appropriate courses of action. The development of skilled behaviour in this context is achieved by the acquisition of cues that both define and discriminate between different types of situation. Experience will further enhance this process by increasing the certainty with which relevant cues are identified, the speed of the identification process and the ability to prioritise the search for cues in the environment. These changes to the information processing should predict the effect of experience in skill development as increasing automaticity, lower workload and more effective performance.

2.5 Skilled long-term working memory

Skilled-memory, or long-term working memory as a function of knowledge and experience is probably the most significant mnemonic skill, or structure, that has direct significance to skilled performance (Groeger, 1997). However, a brief consideration of the properties affecting skilled memory may suggest that it is significantly more applicable to the operation of systems in highly familiar conditions and less applicable to the attempted management of unexpected catastrophic failures. Ericsson and Delaney (1999) give a long and detailed account of why skilled memory may be more significant in everyday skilled activities and Ericsson (2000) personally gave a similar review at the Naturalistic Decision-Making Conference in Stockholm in 2000. Many of the points that Ericsson makes regarding the development of expertise are supported by others (see Green and Gilhooly, 1992). It was suggested that expertise is built up by experience and repeated practice of skills, whether the skills are perceptual-motor skills or cognitive skills. In addition, it was suggested that this pattern of repeated activity created mental structures that allowed large patterns to be transferred rapidly from long-term memory to working memory and to relate new information to these knowledge structures. This clearly has significance for the kinds of decision-making model proposed by Klein (1993a; 1993b; 1997a; 1997b), and for the effortless automaticity of skill-mode processing in Rasmussen's (1983; 1986) model of skill acquisition. However, an argument can be made that this type of processing only applies to typical patterns of information at the interface and typical events in the real world. This in turn raises the possibility of other information processing strategies and the involvement of other memory processes when human operators of complex systems are faced with novel, unfamiliar, unexpected or surprising events. For example, it has been suggested that working-memory limitations may be more directly instrumental in limiting the ability to manage unexpected events because the information available cannot be encoded in the deep schema held in long-term memory. The task conditions can periodically determine a shift in processing strategy simply because the operator is unprepared for the pattern or structure of events that occur. An analysis of accident development suggests that this is a possibility and the use of weak mental models of the controlled system is an explanation that suggests an inferior form of information processing, where the operator is more likely to make errors. It is interesting to note that there are few critical remarks on memory in Reason's (1990) original analysis of errors but he does make wider references to the use of memory aids for supporting tasks. The development of his work (Reason, 1997) further into managing the risks of organisational accidents has largely omitted memory issues as a direct cause of accidents and is singularly focused on effective training as a method for accident reduction, related indirectly to memory. Thus, the implicit acknowledgement of memory in the memory human factors models and the unspoken assumption concerning the use of skilled memory in expert performance largely goes unchallenged. Most analyses of accident identify the workload of unexpected events but very few suggest ways in which this can be managed more effectively.

In this sense authors writing about skilled memory usage in expert performance have ignored the need to address more basic performance issues that appear in the absence of experience to guide action. This type of failure can be predicted by Endsley's (2000) model of situation awareness as a result of poor perception and/or poor comprehension of situation awareness. The role played by an inability to apply a structured schema or mental model in poor memory performance has been known for some time, because it had been demonstrated that chess players were less able to recall randomly placed pieces than legitimately positioned pieces (Chase and Simon, 1973). Extrapolating from this work it is possible to make a case for the effects of working memory on skilled performance under unexpected conditions and this may be exacerbated by the effects of fatigue that further reduce the memory capacity (Cook *et al.*, 2000).

A key issue in Ericsson and Delaney's (1999) analysis of skilled-memory performance was the unnatural nature of controlled laboratory experimentation in that the participant can use guesswork to estimate, with a high probability success, the task requirements on a trial by trial basis. This distinction is shown figuratively in Figure 2.2, where laboratory experiments are at the 'simply, reduce and measure capacity' end of the left hand side and more complex and real-world tasks are represented at the right hand side. In fact, normal operations of complex systems such as aircraft on approach, take-off, climbing, or en-route typically possesses similar highly predictable behaviour, broadly similar to that of laboratory experimentation. It is the unexpected, unusual and unfamiliar events that are less effectively described by Ericsson and Delaney's (1999) analysis of skilled memory and as a consequence it is difficult to apply their model to the management of exceptions that represent critical events in system operation.

Wickens *et al.* (1998) made reference to issues that may play a part in skilled memory performance such as schemas and mental models, and relate them to more effective designs that enable the development of such memory structures. It is significant that, in a later textbook on engineering psychology and human performance, Wickens and Hollands (2000) make similar references to memory in respect of performance but the links to specific aspects of supervisory control are relatively weak. Memory performance in the unexpected, novel and unfamiliar situations typical during accident development are not discussed in respect of changes in memory performance.

2.6 Prospective memory

In the strong multi-tasking environments associated with supervisory control it is common for operators to leave tasks unfinished and to assign a time to return to

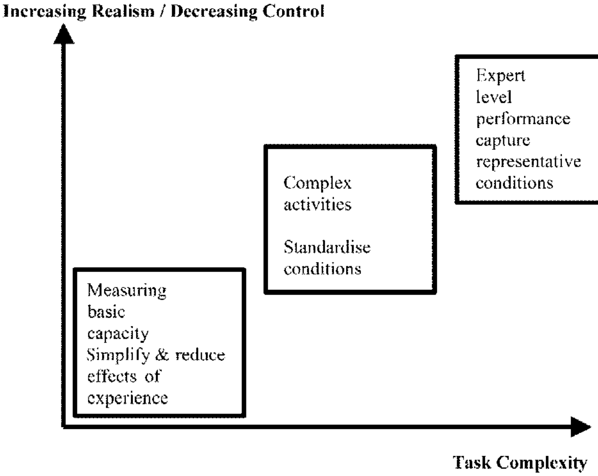


Figure 2.2. Figure representing the abstract and simplified nature of task performance in laboratory tasks where the number of factors potentially affecting performance are decreased, the role of experience decreased and the predictability of task events is increased. Adapted from Ericsson and Delaney (1999: p. 261).

them later. When tasks are set up for completion at some time in the future, memory acts as a placeholder for these tasks and is subject to disruption as tasks are placed between the point at which the memory is created and the future point in time when the task or instructions are to be recalled. This type of memory where future actions and operations are stored for later use is normally referred to as prospective memory and failures of this type of memory are accepted as a probable cause of error and accidents in complex systems (see Reason, 1990; Wickens *et al.*, 1998). Wickens *et al.* (1998) reported a case where an air traffic controller forgot to move an aircraft after positioning it at the end of a runway and caused a crash as a result, and this was interpreted as a prospective memory failure. Reason (1990) suggests that prospective memory probably accounts for a number of significant mistakes contributing to accidents and goes on to propose that it is a major source of human fallibility cognitive processing in complex tasks. Prospective memory is not normally studied directly but the effects of task interruption are. Carlson (1997) identifies management of task information processing as an important factor in successful completion of complex tasks and implicitly endorses the significance of prospective memory in multi-task environments.

2.7 Implicit memory and skilled performance

As experience is accrued with skilled activities it is generally accepted that a significant proportion of the skill-related knowledge is not accessible from conscious experience (see Groeger, 1997). Currently there are only a few studies that examine the effect on implicit cognition and implicit learning on the kinds of skilled activities that are important to designers of complex systems. There is a need to examine many more tasks because it is not clear if there are specific properties of tasks that encourage the use of implicit memory and the evidence for the involvement of implicit memory is relatively weak. However, implicit memory appears to have many of the properties associated with skilled memory usage in that there is a degree of modality specificity (Schacter and Graf, 1989) in the long-term durable knowledge and this is generally recognised as a feature of memory performance in skilled behaviour (Healy and Bourne, 1995; Proctor and Dutta, 1995). In addition, the difficulty with explicitly recalling detailed information in skilled behaviour (Healy and Bourne, 1995; Proctor and Dutta, 1995; Carlson, 1997) is typical of implicit memory function (Schacter, 1994).

There are reviews of implicit memory and memory performance that suggest that different types of cognitive systems are more or less conscious (Gardiner and Java, 1993). It is not clear how these analyses of different memory systems relate to complex supervisory control tasks, as a result of limited cognitive task analysis, and there is currently limited evidence to bridge that gap in knowledge.

2.8 Cognitive aspects of performance in the management of complex systems

It is suggested that a major mistake in many models of human error and human factors analyses of systems is the assumption that when skilled operators make mistakes they are likely to be operating at the highest level of performance in a skill-based mode of processing. It is assumed that the skilled operators will use their extensive experience to manage exceptions and any failures are as a result of faulty processing. If this were the case then skilled-memory would be the appropriate subject for analysis in expert operators and not other forms of memory, as many textbooks suggest.

It is argued here that memory processes relating to working, prospective, episodic and implicit memory need a much greater analysis because of the tendency for novel or unexpected events to present information that is not readily incorporated into already developed schema and scripts of actions. Many of the errors in these processes are likely to be significantly under reported in the literature because they will not be recorded by the current accident and incident analyses. As with fatigue, the effects of memory failure are difficult to identify at post-mortem unless cues can be identified that suggest a memory failure. It is only occasionally that the failures, like forgetting to push forward the engine throttles in the Potomac River crash, Washington DC, bear witness to a memory failure but the effects of distraction on memory performance are well accepted. It is clear that this area of research may benefit significantly from research in the laboratory because an unfamiliar event imposes many of the same demands in a simulation as in a real event, even if the affective response is much weaker.

2.9 Conclusions

Memory is often treated in quite a dismissive way in human factors analysis of systems with frequent references to the limited capacity of working memory and the fallibility in recalling the correct information at the appropriate time. A consideration of the index to textbooks on automation and human performance such as the collection of papers edited by Parasuraman and Mouloua (1996) fails to reveal memory in the index of key terms and many other textbooks exhibit a similar weakness. There are references to memory in widely used textbooks on human factors (Wickens, 1992; Wickens *et al.*, 1998; Wickens and Hollands, 2000). However, memory is not mentioned explicitly in textbooks proposing measures for human factors testing and evaluation (cf. O'Brien and Charlton, 1996). It is as if skilled performance with complex systems has no critical features related to memory performance. Rarely is the capacity for learning as a self-organising ability, which is clearly a part of memory, examined in a positive light and the seemingly infinite capacity for long-term memory is only implicitly recognised. In addition, the changes that take place, as novices become experts, are rarely examined in any great detail even though training programmes are usually developed for systems with any degree of complexity.

This analysis of memory and the way in which it is used within models commonly applied to human factors projects suggests that there are significant benefits to be gained from transferring knowledge developed in experimental cognitive psychology to applied settings. However, this analysis strongly suggests that there are significant limitations of laboratory based research with non-expert participants because the processes that occur in skilled operators are likely to be significantly different from those in the typical laboratory experiment. In addition, there are affective states that may not occur in laboratory experiments that would significantly alter memory performance in the field. For example, the failure to manage a significant event in a cockpit during the final approach, or in a nuclear power plant after a feed-water supply failure, would create affective states that would be an order of magnitude different from those in the laboratory. These affective changes would probably result in changes to information processing and memory performance. It is very likely that state dependent (cf. Groeger, 1997) blocking of memory will take place because the fault management procedures practised under benign conditions will be forgotten under the duress of the failure. Or, there may be an interaction between the less effective information processing typical in highly aroused states and the change in state-dependent cues for memory. This would explain the phenomena of latent learning or knowledge that has been used to explain failures to retrieve knowledge from simulation training during actual events.

Most models of skill development and deployment accept that there are significant changes in the nature of the processing and identify the changes somewhere between working memory and long-term memory. It is as if a new form of memory, which requires less attention and has larger capacity, can be accessed in highly practised tasks. It is difficult to determine if this is a structural, process, or coding change, or a combination of these. Skilled memory as it has been termed certainly explains a significant proportion of the time spent in skilled activity by experts but it may not explain the critical periods when the well-learnt activities are defeated by unexpected circumstance.

It is interesting to speculate that the inability of experienced operators to verbalise fully their skill suggests that at least part of the memory involved in highly practised behaviours and cognitive processes is implicit memory, as an unintentional, non-conscious form of retention (Graf and Schacter, 1985; Schacter, 1987). This link to implicit memory can be contrasted with explicit memory in which the conscious recollection of experiences occurs. It may be possible that explicit memory is more directly linked to episodic memory because of the relationship to explicit cues from past events, feelings and imagery. Implicit learning is more frequently assessed with tasks and differences in performance related to the presence and absence of material in implicit memory, the similarity between implicit memory and the functions in skilled behaviour have already been accepted by some (Dienes and Altmann, 1997). They concluded that implicit acquisition would be less likely to transfer flexibly and be bound tightly to the perceptual experience. The same view has been accepted in other accounts of skill development (Proctor and Dutta, 1995), accounts of experienced cognition

(Carlson, 1997) and other accounts of skilled memory performance (Shank, 1999). In the final analysis, much more work has to be conducted on memory performance during the use of complex systems.

Thus, there are three types of knowledge and skill development that can potentially occur in learning to use a complex system. The first is basic procedural knowledge, either part or whole task, which allows declarative knowledge to be integrated with practical experience to create basic rule-based knowledge. This first type of knowledge is largely associated with the refinement of bottom-up data-driven processes. The second type of knowledge is concerned with active selection of information from the operational domain and storage of information in schematic knowledge structures in long-term working memory. This second type of knowledge is largely a result of refining knowledge or processes that enhance top-down data-driven information processing. The third and final type of knowledge is associated with meta-cognitive processes that further refine the ability to manage the task-related information processing. All of these forms of knowledge must in some manner be stored in long-term memory for retrieval at some future time. For operators of complex systems the training requirement to achieve basic levels of competence and the different ways that memory performance, in its various guises, affects the performance outcomes in managing the system are significant issues.

References

1 BADDELEY, A. D.: 'Working memory' (Oxford University Press, Oxford, 1986)

2 BOWER, G. H., and FORGAS, J. P.: 'Affect, memory and social cognition', *in* EICH, E., KIHLSTROM, J. F., BOWER, G. H., FORGAS, J. P., and NIEDENTHAG, P. M. (Eds.): 'Cognition and emotion' (Oxford University Press, Oxford, 2000) pp. 87–168

3 CAMPBELL, C., COOK, M. J., ANGUS, C., and CRANMER, C.: 'Speech interfaces in real-time control systems'. Proceedings of *People in Control* Conference, Publication No. 463 (Institution of Electrical Engineers, London, 1999) pp. 428–33

4 CARLSON, R.: 'Experienced cognition' (Lawrence Erlbaum Associates, Hove, 1997)

5 CHASE, W. G., and SIMON, H. A.: 'Perception in chess', *Cognitive Psychology*, 1973, **4**, pp. 5–81

6 COOK, M. J.: 'Confidence, trust and uncertainty in air warfare'. Presentation to the staff of Naval Air Warfare Centre, Patuxent River, Maryland, USA, 1999

7 COOK, M. J.: 'Skilled memory, naturalistic decision-making and effective shared situation awareness in air warfare training'. Paper presented at the *5th Conference on Naturalistic Decision Making*, Stockholm, 2000a

8 COOK, M. J.: 'Skilled memory, naturalistic decision-making (NDM) and effective shared situation awareness in air warfare training', *in* KABER, D. B., and ENDSLEY, M. R. (Eds.): 'Human performance, situation awareness, and automation: user-centred design for the new millenium' (SA Technologies,

Marietta in conjunction with North Carolina State University, and Mississippi State University, 2000b) pp. 318–23

9 COOK, M. J.: 'Competence and confidence in air warfare automation'. Paper presented at *ATEDS/SA Symposium*, San Diego, CA, 2000c

10 COOK, M. J., ANGUS, C., BREARLEY, C., and DRUMMOND, K.: 'Decision making, product or process: the implications for training'. Paper presented at the 23rd European Association for Aviation Psychology Conference, Vienna, 1998

11 COOK, M. J., and ELDER, L.: 'Training group performance for biasing and de-biasing decision-making which avoids Groupthink', *RTO-HFM Collaborative Crew Performance in Complex Operational Systems*, **38**, (1–12), RTO-MP-4, AC/323(HFM)TP/2, (NATO (RTO HFM): (Nieully sur Seine, France, 1998)

12 COOK M. J., ELDER, L., and WARD, G.: 'Communication requirements in the cockpit', *RTO-HFM Collaborative Crew Performance in Complex Operational Systems*, **19**, (1–11), RTO-MP-4, AC/323(HFM)TP/2, (NATO (RTO HFM): (Nieully sur Seine, France, 1998)

13 COOK, M. J., REID, E., and WILSON, K.: 'The role of working memory in command and control tasks'. Paper presented at the 3rd Conference on *Engineering Psychology and Cognitive Ergonomics*, Edinburgh, 2000

14 COWAN, N.: 'Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system', *Psychological Bulletin*, 1988, **104**, (2), pp. 163–91

15 COWAN, N.: 'Activation, attention, and short-term memory', *Memory and Cognition*, 1993, **21**, pp. 162–7

16 DIENES, Z., and ALTMANN, G.: 'Transfer of implicit knowledge across domains: how implicit and how abstract?', *in* BERRY, D. C. (Ed.): 'How implicit is implicit learning' (Oxford University Press, Oxford, 1997) pp. 107–23

17 ENDSLEY, M.: 'Toward a theory of situation awareness', *Human Factors*, 1995, **37**, (1), pp. 32–64

18 ENDSLEY, M.: 'Situation awareness measurement in test and evaluation', *in* O'BRIEN, T. G., and CHARLTON, S. G. (Eds.): 'Handbook of human factors testing and evaluation' (Lawrence Erlbaum Associates, Mahwah, NJ, 1996)

19 ENDSLEY, M.: 'The role of situation awareness in naturalistic awareness in naturalistic decision-making', *in* ZSAMBOK, C. E., and KLEIN, G. (Eds.): 'Naturalistic decision-making' (Lawrence Erlbaum Associates, Mahwah, NJ, 1997) pp. 49–59

20 ENDSLEY, M.: 'Theoretical underpinnnings of situation awareness: a critical review', *in* ENDSLEY, M., and GARLAND, D. J. (Eds.): 'Situation awareness analysis and measurement' (Lawrence Erlbaum Associates, London, 2000)

21 ERICSSON, K. A.: 'Skilled memory'. Presentation to the 5th Naturalistic Decision Making Conference, Stockholm, Sweden, 2000

22 ERICSSON, K. A., and DELANEY, P. F.: 'Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance', *in* MIYAKE, A., and SHAH, P. (Eds.): 'Models of working memory: mechanisms of active maintenance and executive control' (Cambridge University Press, Cambridge, 1999) pp. 257–97

23 ERICSSON, K. A., and KINTSCH, W.: 'Long term working memory', *Psychological Review*, 1995, **102**, pp. 211–45

24 GARDINER, J. M., and JAVA, R. I.: 'Recognising and remembering', *in* COLLINS, A. F., GATHERCOLE, S. E., CONWAY, M. A., and MORRIS, P. E. (Eds.): 'Theories of memory' (Lawrence Erlbaum Associates, Hove, 1993) pp. 163–88

25 GILHOOLY, K. J.: 'Thinking: directed, undirected and creative' (Academic Press, London, 1983)

26 GRAF, P., and SCHACTER, D. L.: 'Implicit and explicit memory for new associations in normal and amnesiac patients', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1985, **11**, pp. 501–18

27 GREEN, A. J. K., and GILHOOLY, K. J.: 'Empirical advances in expertise research', *in* KEANE, M. T., and GILHOOLY, K. J. (Eds.): 'Advances in the psychology of thinking: volume one' (Harvester Wheatsheaf, London, 1992) pp. 45–70

28 GROEGER, J. A.: 'Memory and remembering' (Longman, Harlow, 1997)

29 HEALY, A. F., and BOURNE, L. E. Jr.: 'Durability and specifity of knowledge skills', *in* HEALY, A. F., and BOURNE, L. E. Jr. (Eds.): 'Learning and memory of knowledge and skills: durability and specificity' (Sage Publications, London, 1995) pp. ix–xiv

30 HUEY, M. B., and WICKENS, C. D.: 'Workload in transition' (National Academy Press, New York, 1993)

31 KLEIN, G.: 'Naturalistic decision-making'. Wright Patterson Air Force Base, Ohio, Crew Systems Ergonomics Research Information Analysis Centre (CSERIAC), 1993a

32 KLEIN, G.: 'A recognition-primed decision (RPD) model of rapid decision-making', *in* KLEIN, G. A., ORASANU, J., CALDERWOOD, R., and ZSAMBOK, C. E. (Eds.): 'Decision making in action: models and methods' (Ablex, Norwood, NJ, 1993b) pp. 138–47

33 KLEIN, G.: 'An overview of naturalistic decision-making applications', *in* ZSAMBOK, C. E., and KLEIN, G. (Eds.): 'Naturalistic decision-making' (Lawrence Erlbaum Associates, Mahwah, NJ, 1997a) pp. 49–59

34 KLEIN, G.: 'The current status of naturalistic decision-making framework', *in* FLINN, R., SALAS, E., STRUB, M., and MARTIN, L. (Eds.): 'Decision making under stress: emerging themes and applications' (Ashgate Publications, Aldershot, 1997b) pp. 11–28

35 KLEIN, G.: 'Cognitive task analysis of teams', *in* SCHRAAGEN, J. M., CHIPMAN, S. F., and SHALIN, V. L. (Eds.): 'Cognitive task analysis' (LEA, London, 2000a) pp. 417–30

36 KLEIN, G.: 'Analysis of situation awareness from critical incident reports', *in* ENDSLEY, M. R., and GARLAND, D. J. (Eds.): 'Situation awareness analysis and measurement' (Lawrence Erlbaum Associates, Mahwah, NJ, 2000b) pp. 51–71

37 McGUINNESS, B.: 'Workload and situation awareness assessments in a simulation of digitized command and control'. Paper presented at *People in digitized command and control* Conference, Shrivenham, UK, December 12–14, 2000

38 McGUINNES, B., FOY, L., and FORSEY, T.: 'Battlespace digitization: SA issues for commander', *in* KABER, D. B., and ENDSLEY, M. R. (Eds.):

'Human performance, situation awareness, and automation: user-centred design for the new millenium' (SA Technologies in conjunction with North Carolina State University, and Mississippi State University, Marietta, 2000) pp. 125–30

39 NORMAN, D. A.: 'The "problem" with automation: inappropriate feedback and interaction, not "over-automation"', *Philosophical Transactions of the Royal Society of London B*, 1990, **327**, pp. 585–93

40 O' BRIEN, T. G., and CHARLTON, S. G. (Eds.): 'Handbook of human factors testing and evaluation' (Lawrence Erlbaum Associates, Mahwah, NJ, 1996)

41 PARASURAMAN, R., and MOULOUA, M. (Eds.): 'Automation and human performance: theory and applications' (Lawrence Erlbaum Associates, Hove, 1996)

42 PROCTOR, R. W., and DUTTA, A.: 'Skill acquisition and human performance: advanced psychology texts' (Sage, London, 1995)

43 RASMUSSEN, J.: 'Skills, rules, and knowledge: signals, signs and symbols, and other distinctions in human performance models', *IEEE Transactions on Systems, Man and Cybernetics*, 1983, SMC-13, pp. 257–66

44 RASMUSSEN, J.: 'Information processing and human-machine interaction: an approach to cognitive engineering' (North-Holland, Amsterdam, 1986)

45 REASON, J.: 'Human error' (Cambridge University Press, Cambridge, 1990)

46 REASON, J.: 'Managing the risks of organizational accidents' (Ashgate Publishing, Aldershot, 1997)

47 SCHACTER, D. L.: 'Implicit memory: history and current status', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1987, **15**, pp. 3–12

48 SCHACTER, D. L.: 'Understanding implicit memory: a cognitive neuroscience approach', *in* COLLINS, A. F., GATHERCOLE, S. E., CONWAY, M. A., and MORRIS, P. E. (Eds.): 'Theories of memory' (Lawrence Erlbaum Associates, Hove, 1994) pp. 387–412

49 SCHACTER, D. L., and GRAF, P.: 'Modality specificity of implicit memory for new associations', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1989, **15**, pp. 3–12

50 SHANK, R. C.: 'Dynamic memory revisited' (Cambridge University Press, Cambridge, 1999)

51 WICKENS, C. D.: 'Engineering psychology and human performance' (Harper Collins, London, 1992)

52 WICKENS, C. D., GORDON, S. E., and LIU, Y.: 'An introduction to human factors engineering' (Longman, New York, 1998)

53 WICKENS, C. D., and HOLLANDS, J. G.: 'Engineering psychology and human performance' (Prentice Hall, Upper Saddle River, NJ, 2000, 3rd edn.)

Chapter 3
Vigilance

Craig Donald

3.1 Introduction

Human vigilance underlies the successful implementation and operation of many technical systems. This chapter examines some of the issues and implications of vigilance for performance in technology intensive operations at work-stations and in control rooms and command centres. Despite the increasing sophistication and automation of many of the functions previously performed by people, the role of operators within complex operating environments remains essential. Indeed, as routine operations become more automated, the shift to a more defined and focused role for the operator in detecting critical incidents or conditions often occurs. The ultimate identification and interpretation of conditions and signals by people and subsequent decision-making remains the cornerstone of system performance and incident management. Among others, this applies to process controllers in plants such as chemical, manufacturing and power plants, air traffic controllers and pilots in aircraft, X-ray screeners, closed circuit television (CCTV) surveillance operators, emergency control operators, medical monitoring functions, military personnel in confrontational settings and quality inspection functions.

The concept of vigilance receives mixed levels of attention in the design, implementation and operation of technology driven systems. A shortfall in focus on the area may be due to perceptions that the work of human operators is simple and non-demanding. The involvement of operators is not seen to be significant because it is felt that the technology can perform or be developed to perform any significant task required of an operator. A second and sometimes related perspective is that operator jobs are routine and mundane, and that there are no special skill requirements for such a person. The perspective of this chapter, however, is that operators are an integral part of the control system and are critical in realising the performance potential of the technology. Further, although there may be extensive routine in these positions, sustained vigilance has been demonstrated to impose appreciable demands on information processing and working memory (Parasuraman, 1979). In line with this, Wickens (1992) noted that conclusions from the analysis of operator performance in vigilance research indicate that operators are far from optimal in their performance and demonstrate a higher miss rate than desirable. The complex visual, and at times audio, analysis associated with work such as air traffic control, airport X-ray screening and military encounters requires extensive processing of information. Incident conditions, unusual events and system malfunctions in particular can lead to highly demanding cognitive, communication and emotional responses and require rapid but critical decision-making. Because of these kinds of demands and associated responsibilities, stress and fatigue are common factors in vigilance intensive positions (Sawin and Scerbo, 1995).

3.2 The concept of vigilance

Drawing strongly on signal detection, resource, expectancy and arousal theory among others (Davies and Parasuraman, 1982; Wickens, 1992; Huey and Wickens, 1993), definitions of vigilance generally incorporate a number of key themes. These include sustained attention, signals, detection, staying alert, being able to identify targets and maintaining performance over time. Rather than covering these theories, the chapter will draw out general conclusions and highlight some of the main research findings. Broadly, vigilance in this context refers to a capacity for sustained effective attention when monitoring a situation or display for critical signals, conditions or events to which the observer must respond. Incorporated into this perspective on vigilance is the ability to identify, recognise and interpret the information that is being monitored.

Operator tasks can require active observation and the continual analysis of images and/or information. Perceptual processing and recognition are also important to outcomes. This goes beyond simple sensory recognition to being aware of the context of the situation and the identification of patterns within the displayed information. Information being viewed must be put in context and the status of the situation must be continually evaluated. The operator may also need to deal with separate sources of information and allocate attention dynamically according to the pattern or relevance of signals received. Huey and Wickens (1993) described the vigilance requirements of a battle tank crew acquiring a target under combat conditions. It included searching the ground and sky for target signatures (critical signals) involving both acoustic and visual patterns (such as engine noise, broken vegetation, weapon smoke and glare from objects) using both their own senses as well as viewing aids under extreme physical conditions. Although the job of air traffic controllers or X-ray screeners may not be as extreme, the responsibilities of the position and the complex visual analysis that is carried out on an ongoing basis is nevertheless highly demanding, with a frequent flow of bags to be screened or planes to be monitored (Bennett and Fobes, 1994; Wickens, *et al.*, 1997).

3.2.1 Measuring vigilance

Vigilance intensive operations have as their main challenge the requirements of consistently accurate attention and detection, and the maintenance of performance levels under these conditions. However, measurement of such performance is more difficult than may be apparent. Vigilance research typically uses a standardised environment and relatively simplistic signals or target behaviour with fairly frequent occurrences of such signals within the defined time period. This contrasts with real world situations where more complex stimuli occur in often ambiguous circumstances in an unpredictable, event-driven environment (Wiener, 1987). Further, signal rates differ from the frequency found in research studies, often occurring at intervals of weeks or even months apart. Measurement of vigilance levels in some operational concerns is also complicated by the fact that, unlike research settings, one is often unaware of the incidents that were missed. Although the logging of anomalies outside parameters within a process control room provides a context for missed events and near-incident reports reflect air traffic controller performance, a CCTV operator may never be aware that something has been missed. The missed incident gets lost in the general flow of things and the incident is not brought to anyone's attention unless there is a serious consequence from this lapse. Reviews of CCTV video tapes, for example, can in many cases detect a higher percentage of incidents than obtained during live surveillance (Donald, 1997).

The most obvious of the measures of vigilance is detection rate. However, detection rates need to be benchmarked against something if they are to be an effective measurement. In a 'free response' environment (Wickens, 1992) the incident or target event may occur infrequently and non-events are not defined. The level of vigilance can be established by reports on the impact of clearly missed violations (e.g. reports from other departments or police reports) or relative to audit criteria such as logs of parameter violations or the tape reviews mentioned above. A possible, but at times misleading, method is to evaluate relative to the level of detection for other operators in equivalent situations. However, it is still not apparent what the actual detection rate standard should be. In the 'inspection' environment, events occur at fairly regular intervals and are differentiated into target and normal events (e.g. quality control). This allows the detection rate per signal rate to be used as a measure of vigilance, although it is still not obvious what the actual detection rate should be. Other measures of vigilance can relate to the identification of priority versus non-priority events, response times and the level of false alarms. Because of the difficulty in getting a benchmark measure, statistics on consequences are often used as an indicator of vigilance. This could include complaints from customers for goods supposed to have been subject to quality inspection and control, police reports on crime incidents not detected on the street, avoidance actions by pilots, or incident and irregularity reports filed downstream by other departments in a process control department.

3.3 Vigilance and performance

Vigilance within the control room is a consequence of a number of factors and the degree to which these are in alignment and complement each other will determine the ultimate level of vigilance of the responsible person. The initial job definition and subsequent design of the work-station/control room play the most fundamental role in determining vigilance requirements. However, vigilance levels cannot simply be explained in terms of the nature of job demands. The physical environment plays a part in determining the level of vigilance and whether required levels can be maintained, as does the way the situations are managed over the extended period of operation. Further, the staff used to operate the technology need to be reconciled with the job and work environment demands. There is often an assumption in vigilance research that all people will perform the same given identical conditions. The strengths and weaknesses of different people will, as in any other occupation, make them more or less suited to different vigilance tasks and environments. Training in what to look for and how to look for it will also influence the effective level of vigilance. This chapter will focus on the areas of job demands and the nature of human vigilance as the design of the work environment and its impact on people is addressed in detail in other chapters.

It is important to recognise that the nature of vigilance is not uniform across different positions. Performance efficiency in vigilance tasks is closely tied to the nature of the stimuli that need to be dealt with (Huey and Wickens, 1993). Different kinds of operations, jobs and situations make different demands on vigilance. The job profile should reflect the vigilance demands and signals that define the core elements of vigilance required on the job. These demands can then be accommodated and facilitated in the job design, work methods, work-station and control room design and management functions. In addition, the actual sustained level of vigilance required in monitoring the situation needs to be established, as well as whether activities can lead to a decrement in vigilance that could comprise efficiency or effectiveness. The concept of 'vigilance decrement', namely the fall off in performance over time, is central to many concerns over operators' job performance. The following section examines factors in the work context that influence vigilance and impact on the sustainment of the level of vigilance.

3.4 Work and task factors and vigilance

3.4.1 The influence of workload

The workload of the operator has a number of implications for vigilance levels and decrement. The term 'workload' can refer to mental and associated cognitive demands, or workload in terms of physical and temporal demands. Both types of work overload can lead to reductions in vigilance as the person struggles to maintain accuracy and judgement under information and time pressure. However, work underload can also produce a lack of stimulation and introduce boredom into the job role, resulting in mental disengagement from activities being worked on and lowered vigilance. While operators engage in adaptive practices to deal with overload situations by reprioritising, working 'smarter' or shifting responsibilities, overload may still reduce the level of attention and level of vigilance given to areas that need to be monitored. Underload, on the other hand, potentially leads to displacement of attention and mental disengagement and can substantially reduce effective monitoring as well as response time. Mackie *et al.* (1985) conducted a comparative study of stressors and the impact of these on vigilance for sonar operators on submarines, surface ships and helicopters. First on the list for submarine and surface ship operators was boredom and monotony, while this was second for helicopter operators. Fatigue and tiredness, difficulty with displays/controls and operator overload featured prominently among the top six stressors for all groups. More importantly, operators judged that these stressors had a substantial impact on their vigilance levels and overall effectiveness.

Vigilance intensive positions are generally associated with prolonged periods of attention within a relatively routine task where critical incidents or signals need to be identified. Diversifying the information or tasks that are monitored or conducted is one method of enriching the job and introducing some variety. This is seen to keep people more involved in their work and it also has the potential to reduce vigilance decrement from occurring. However, diversifying inputs also reduces the amount of vigilance to any one of the specific functions being performed. For example, CCTV operators often provide a communications and alarm monitoring service as well as conducting surveillance. Although the overall level of vigilance is high, the demands of communication on the one hand and real time surveillance on the other is likely to lead to a reduced level of vigilance and detection in surveillance. However, the dilution of vigilance may also occur with more requirements within the same function. Wallace *et al.* (1995), for instance, found a significant deterioration of detection rates with increased numbers of CCTV monitors to be viewed. The effect on task performance of viewing one, four, six and nine monitors led to accuracy scores of 85 per cent, 74 per cent, 58 per cent and 53 per cent respectively. Although CCTV surveillance operators can rate the risk factors associated with situations on the different monitors and focus on priorities, the need to cover multiple sources of information reduces the absolute level of vigilance for each source. In control rooms with 15 to 30 monitors per operator, effective vigilance is seriously impaired by the need to distribute attention across so many information sources. The minimum vigilance levels for specific functions must therefore be clearly determined on the basis of operational requirements whenever multiple functions are part of the job.

3.4.2 Work complexity

The complexity of situations or targets being monitored leads to a substantial increase in cognitive demands and information processing. The use of working memory and cognitive interpretation to evaluate available information and to decide on deviations from normal conditions also becomes more pronounced under these conditions. Davies and Parasuraman (1982) made the point that successive tasks that involve appraisal and absolute judgement against standards retained in working memory are more resource demanding than simultaneous tasks where comparative judgements based on distinguishing information in the signals themselves are made. Consequently, successive tasks are more demanding and can lead to greater vigilance decrements. This may be exaggerated in situations where there is a time parameter that has to be adhered to and the operator is forced into a pressurised vigilance state where signals, images or other information have to be evaluated and decided on quickly but accurately. Background signal noise is also associated with greater complexity, creating greater ambiguity as targets become less distinct or more difficult to pick out. This background noise may take the form of additional information, activities around the operator, or sounds. As it occurs, the ability to discriminate an incident condition becomes more difficult. The higher the level of background noise or the more frequently changes occur in the background, the greater the distraction will be. Given these factors, one can appreciate that high complexity typically leads to increased mental workload and potential for stress.

The level of workload is, therefore, critical to operators under high complexity conditions. The balance needs to be struck between processing time and the job demands of greater throughput or greater simultaneous coverage. Technology that assists operators by highlighting particular cues or potential risk conditions to facilitate speed of processing are important elements in handling complexity. Similarly, making the signal more conspicuous by increasing the amplitude of the signal to noise ratio of critical signals can improve attention and consistency over time. For X-ray screeners at airports, for example, the detection of various explosive or weapon components within luggage can be complicated by the overlay of various levels of goods within the baggage that mask some of the characteristics of the target items. X-ray technology has consequently evolved such features as different penetration levels, organic/inorganic stripping and assisted detection features. However, while these provide potential distinguishing information on potential target items and cue screeners to certain features, screeners still have to engage in substantial visual analysis and cognitive interpretation. This also occurs with appreciable workflow rates leaving only a few seconds to view and analyse each piece of luggage.

3.4.3 Optimal work periods

Deciding on the work periods that will provide optimal vigilance within the shift will depend on the reconciliation of four main factors. These are optimal concentration time, required viewing consistency, handover and work variety. The optimal concentration span that emerges from research and actual operations for a person is generally seen to be about 25 to 30 minutes, although some parties see it being as short as 20 minutes. However, concentration needs to be reconciled against the need for consistency of observation. Some work situations need to be tracked and developments noted across time. Short viewing periods tend to break up the viewing pattern and the operator's overall situation awareness (e.g. in CCTV surveillance). The requirements of handovers from person to person also influence what work period is appropriate. Where the work covered in the various work periods is discrete and few handover briefing requirements and complications exist, shorter work periods can be implemented without much difficulty. Handovers which require more extensive briefing and familiarisation in order to generate appropriate situation awareness, such as air traffic controllers, are less conducive to short work periods. Operators who perform a variety of tasks as part of their work vary their concentration levels and focus, and have the opportunity to refresh their minds and viewing perspectives. Where work involves viewing situations where little or nothing may be going on, boredom is increased and it becomes difficult to maintain vigilance levels. The intense visual analysis and the fact that each bag of luggage is treated as a separate component in the work of X-ray screeners make work periods of 20 minutes realistic to avoid a decrement in vigilance and too much fatigue. After that the screener rotates with another member of the section and engages in other security work such as searching before returning to screening. Work schedules for air traffic controllers in the US, on the other hand, call for an overall shift of eight hours to be broken into two-hour maximum work periods and a full hour of breaks within the shift period (Wickens *et al.*, 1997). While intensive in nature, the handover process and the need to form a picture of traffic as part of their situation awareness makes longer work periods more appropriate than with the X-ray screeners.

3.4.4 Signal characteristics

Signal characteristics and their impact on vigilance and vigilance decrement represent part of the core of much of the vigilance theory research. Despite criticism of the simplistic nature of some of this research and questions about the relevance to real world and more complex situations (Mackie, 1987; Wiener, 1987), signal characteristics need to be considered for any kind of implementation and ongoing operation requiring high levels of vigilance. This section will attempt to summarise some of the key findings that should be considered with a focus on the practical implications of these. Extensive theoretical and research discussions are available in the literature for readers who want to pursue these areas in more depth, particularly discussions of signal detection theory.

3.4.4.1 Consistency of the signal

The consistency of the signal, target or situation profile will have a marked effect on detection levels. The more often the target signal occurs and the more regular it is, the better the vigilance level and the less likely the decrement. Frequent signals maintain the operator's attention more effectively than when signals occur at long intermittent periods. Further, operators are generally quick to pick up the patterns of signals and the existence of regular intervals allows them to predict their occurrence and adjust their approach accordingly. The time between signal intervals may be used for psychological rest breaks by operators. A similar situation can occur with the risk factors and viewing by CCTV operators. Where no risk factors are displayed on screen, operators may reduce concentration levels and their focus on the screen and switch to a routine monitoring mode, becoming more vigilant when risk factors increase or targets come on screen. The main problem is that if signals fall out of the expected pattern or the operator is not sufficiently knowledgeable about risk conditions, the chances of missing them become appreciably higher as the operator is likely to become less focused and more relaxed during these ‘quieter’ or ‘safe’ periods. This expectancy of signal occurrence plays a strong part in vigilance and as Huey and Wickens (1993) indicated, signal detection in vigilance is determined as much by what transpires in the interim between signals as by the characteristics of the signals themselves.

3.4.4.2 Signal predictability

Most vigilance intensive positions fall outside the frequent target and regular interval scenario. Unpredictable, and with targets or threat conditions occurring weeks or even months apart, these types of positions pose particular challenges to vigilance. If the operator is uncertain when the signal is likely to occur or its likely location, both response times and effectiveness of detection are likely to be lower. Response time to incidents also tends to become longer as the likelihood of an incident becomes more uncertain. In one example, to combat such attention issues, aviation authorities have introduced an artificial increase in incident rates for X-ray screeners through the use of threat image projection (TIP) (Bennett and Fobes, 1994). TIP substitutes images containing target material for the images of standard passenger bags being reviewed at X-ray work-stations. The increased incidence of target images (which can be set through the computer) creates operator expectations, and stimulates and motivates them. In addition, it provides a performance measure of operator vigilance and ultimately that of the whole operation.

3.4.4.3 Signal distinctiveness

Signals need to be distinctive in order to be detected. A signal may be too subtle or too sensitive for the operator, or alternatively the operator may be undecided about the precise status of the signal. The duration of the display of the signal will also influence how easily it is recognised although prolonged displays tend to fade from attention. The use of dual mode signals such as the simultaneous combination of audio and visual stimuli to draw attention to a condition is also likely to improve performance. Positioning of the signal is critical in terms of how it relates to the operator’s typical line of sight during operations. However, the extent of other signals around it and ‘background’ noise can reduce attention to the signal appreciably. Increasing the signal to noise ratio of critical signals can help improve attention and consistency. One needs to be careful of the ‘cry wolf’ syndrome with signals, however. If an alarm or other signal becomes too frequent or too much a feature of the operator's work without enough justification, it tends to get pushed towards the status of being background noise and becomes a lower monitoring priority, even when the signal is showing. Similarly, having a high number of signals with high visibility that are shown excessively not only creates additional processing demands on the operator, but can potentially lead to a reduction in sensitivity to the signals. Grouping and positioning of such signals is an important element in effective allocation of attention and this needs to be reconciled with the need for distinctiveness.

3.4.4.4 Signal interpretation

The degree to which the signal puts a load on working memory and requires interpretation and memory recall will lead to increased stress and fatigue. Recognition rather than recall generates faster addressing of signals and the more the signal is self explanatory, the less demand it will place on cognitive processes. Having a large number of factors that require absolute judgement rather then comparative judgement places additional demands on the operator's information processing resources and are likely to impact on the efficiency of vigilance (Davies and Parasuraman, 1982). On the other hand, signals similar in nature to those that have been memorised are also likely to be seen more quickly as people recognise the signs. Unique and different signals or conditions that require interpretation increase the response time and likelihood of detection.

3.4.5 Shifts, circadian rhythms and sleep disruption

Shift work is a common feature in many of the occupational positions requiring high vigilance. It is commonly recognised that alertness and vigilance can be adversely affected during night shifts by circadian rhythms that regulate one's body cycle over a period of approximately 24 hours. The circadian rhythms are set by exposure to sunlight and darkness which influence a central mechanism in the hypothalamus so the internal state of the person varies predictably over 24 hours. These rhythms govern psychomotor and cognitive performance allowing people to be alert and perform effectively during the day but causing them to be less alert at night. During the latter part of the night, people are at their mental and physiological trough and maintaining high levels of alertness and performance at this time is exceedingly difficult (Huey and Wickens, 1993). For shift workers, the circadian clock maintains a strong regularity over time even when the normal cues are withheld from the person and is capable of continuing for several months. Huey and Wickens provide a concise summary of the impact on night shift workers of misaligned circadian rhythms, stating

> they are attempting to stay awake and perform complex tasks at a time of day that the human circadian clock has scheduled the central nervous system to be asleep, and then attempting to sleep at the scheduled time of peak arousal. In and of itself, this misalignment of circadian phase results in deterioration of alertness and performance. In addition, because of the effect

> of circadian phase alignment on sleep, night shift work inevitably results in acute and chronic sleep deprivation, which can independently impair performance (Huey and Wickens, 1993:125).

Problems of sleep deprivation, sleeplessness and fatigue can combine with circadian rhythms to have major implications for worker adjustment and levels of vigilance during shifts (Sawin and Scerbo, 1995). Sleep-related disruptions appear to be most prevalent in low-load vigilance based monitoring tasks (Huey and Wickens, 1993). These problems can lead to mental and visual fatigue, variable attention levels, reduced sensitivity in detection and increases in incidences of false alarms. While 'phase delayed' shift designs allow for the easier adaptation of circadian rhythms by spreading the work week over a longer period of time with greater and more regular sleep opportunities, Wickens *et al.* (1997) indicated that workers prefer a phase advanced schedule. This is because it allows compression of duty into a shorter working week and allows longer non-work weekends. Decreasing the frequency of shift changes is seen as another option in order to maintain some consistency of rhythms once the person starts a new schedule. More recently, the use of circadian lighting has been introduced where lighting levels are systematically varied to shift circadian rhythms so operators remain alert at night. Periods of restorative sleep have also been explored along with identification of personnel most suited to work at night times.

3.5 Work environment influences on vigilance

Lighting, noise, ventilation, vibration and temperature are all factors in the working environment that can impact on levels of vigilance. In addition, motion in ships and aircraft can also affect vigilance in these environments. Issues relating to control room environments will be addressed in more depth in other chapters but are briefly covered here. As indicated above, workplace lighting levels have been systematically varied to improve worker alertness and performance during working hours. The degree of operator comfort with the lighting is often a personal issue and while recommendations vary from 200 to 500 lux, the option of operators being able to control lighting around their own work-stations needs to be seriously considered. In the case of noise

> performance is degraded by loud noise (above 90dB SPL) when the information processing or resource demands of the task are high while it remains unaffected by loud noise when the processing demands are low. In addition, performance under low demand conditions can be facilitated by low-level (approximately 64dB SPL) continuous noise that is variegated in character (Huey and Wickens, 1993: 150).

Higher levels of noise tend to make operators focus on more central priority events and this leads to a fall off in vigilance tasks that are seen as more peripheral to the job. Vibrations, also a contributor to noise, in extremes can cause headaches, vision problems and a lack of concentration, all of which can impact on vigilance levels. Guidelines on temperature conditions generally agree on a range between

20–26 degrees centigrade providing a comfort range with greater discomfort and performance drops as the temperature moves to extremes.

3.6 The vigilance of people

In much of the vigilance research, the impact of personal characteristics on the performance of the participants is not seen as an intervening variable and it is assumed that personnel will perform at a standard level within the experimental conditions. A similar trend occurs in the design and operation of vigilance intensive positions. Given appropriate ergonomic design in areas such as layout, information display and control interfaces, together with a system of policies and procedures to ensure standardised operation, it is expected that performance requirements will be met. Yet if the people responsible for the design, implementation and management of the systems are asked if they would like to perform that kind of work, they often do not relate to it. The same distinguishing levels in performance that differentiate good from weaker performers in any area apply to vigilance type tasks. In discussing air traffic controllers, for example, Wickens *et al.* (1997) noted that not all controllers experience extremes of workload and most use various adaptive strategies to manage their performance and subjective perceptions of task involvement. Donald (1999) has found that CCTV operators identified as high performers on a vigilance assessment instrument can identify up to twice as many incidents as personnel identified as lower performers. The performance of personnel within the control room environment can differ markedly, and simple errors of attention or oversights may occur despite intensive attention to the ergonomic interface. The identification of the kind of people suited for a vigilance intensive environment and the selection of these personnel are therefore fundamental to any effective system implementation.

3.6.1 Vigilance characteristics and qualities in people

A person's vigilance is not a single quality or skill but results from a combination of abilities and temperament factors that can be matched against the position. Somebody who shows high vigilance in a process control centre may not demonstrate the same level of suitability in a quality inspection role, for example. However, there is a general set of characteristics applicable to most vigilance intensive positions. Starting from basic physical requirements to more advanced perceptual processing of information and finally the ability to apply this on a sustained basis, the model progressively builds on the qualities outlined in more depth in Table 3.1.

The gradient illustrated in Table 3.1 is important in that each level is a key component in defining the final vigilance level of a person. A shortfall at any level will have ramifications all the way through in terms of limiting subsequent vigilance levels. Selection practices such as biographical backgrounds, interview-

Table 3.1. The gradient of competencies for human vigilance.

Area	*Sample competencies*	*Explanation*
Physical capacity	• Visual acuity • Colour differentiation • Audio acuity.	The physical capacity of the person provides the foundation for vigilance related skills.
Sensory recognition	• Presence and loss of signal • Colour changes • Audio tones/signal differentiation • Sensitivity to motion/speed • Perceptual clustering.	The sensory recognition abilities define the sensitivity of the operator to signals and conditions in the environment. These abilities relate to monitoring and detection functions and define how likely somebody is to be able to detect changes in information or data.
Perceptual processing	• Pattern recognition • Sensitivity to critical cues • Recognition of anomalies • Use of memory (short/medium/long term) • Seeing information in context • Relating disparate sets of information together • Screening of background noise • Speed of becoming orientated to new information sources.	Perceptual processing refers to perceptual and cognitive processing abilities. These abilities are essential in allowing the operator to make sense of the information presented. These also define how well the operator can 'tune in' to the patterns and trends within the information and carry this through to an overall situation awareness. The capacity to pick up deviations, anomalies and inconsistencies in signal, data, or behavioural trends and to interpret these in the context of variation from expected conditions is critical to effective vigilance.
Observation skills	• Scanning efficiency of information sources • Depth of coverage • Division of attention • Attention to detail • Use of peripheral vision • Search techniques.	Observation skills refer to the actual skills and methods that the operator uses to search for information within an information source as well as covering multiple data sources within the work environment. The operator may have the perceptual processing skills to analyse and interpret such data, but unless it is effectively accumulated, vigilance levels will be impaired.
Sustained attention	• High concentration levels • Ability to deliver sustained attention • Consistency over time • Tolerance for routine • Patience • Emotional stability • Self discipline • Goal orientated.	Effective human vigilance requires the sustained and consistent application of the above skills in the working scenario. This requires certain abilities and personality or temperament characteristics from the operator. The absence of many of these characteristics may mean that although the person has the vigilance skills required, these cannot be applied in a sustained manner in order to meet performance requirements.

ing, psychometric assessment and the use of vigilance assessment exercises are methods that can be used to assess such qualities.

3.7 Enhancing vigilance

Conditions that detract from the effectiveness of vigilance discussed previously have been accompanied by some methods of addressing these circumstances. In addition, a number of practical steps to address vigilance shortfalls can be adopted. These are described under the categories of management, training and technology aided interventions.

3.7.1 Management-driven enhancement methods

The development of a strong and supportive work group environment can maintain cohesion and commitment to responsibilities. Besides the generation of morale and involvement, increasing the ratio of operators within the same area has the potential benefits of concentrating attention by using multiple operators to monitor situations with one operator picking up conditions that another may have missed. In situations where operators are pursuing intermittent targets (e.g. CCTV) the success of other team members also enhances overall team motivation as all can share in the success, occasional though it may be. Multi-skilling and task rotation can create variation in the nature of work and the flow of the shift and change the nature of concentration requirements. Providing appropriate work and rest schedules also form an important part in maintaining the vigilance levels of personnel (Mackie, 1987).

The provision of a feedback loop on operator performance enhances the capacity to differentiate priority from non-priority effects and increases motivation when the individual is given recognition for positive accomplishments. Feedback on actual situations is ideal, but the TIP-based system that provides X-ray screeners with simulated targets is also a method of providing real time testing and feedback. The use of simulated targets during the normal work process needs to be implemented carefully as it can generate employee resistance, but it has potential for a range of other vigilance intensive settings. Biofeedback-associated 'alertness indicators' on the actual operator themselves may also assist them and management to maintain concentration levels more effectively.

The use of personnel selection procedures to identify and place personnel with the capacity for high vigilance has the potential to enhance the system effectiveness. Assessing a candidate's suitability for the position would include a review of qualifications, training, experience and competencies. A relatively small expenditure on assessment of the person's vigilance capacity during the placement of suitable personnel can lead to significant benefits in overall system performance (Donald, 1999).

3.7.2 Training to enhance vigilance

Training and increased awareness in what to monitor have been associated with higher detection rates (Parasuraman, 1986; Mackie, 1987). This training could involve sensitisation of operators to target conditions, or target cueing, so that incident conditions can be recognised more quickly and interpretation reduced, e.g. signs of a developing incident for CCTV operators, configurations of different conditions or reference standards for comparative purposes for a process controller. This type of training also sensitises operators to distracters, low or ambiguous signal conditions and hazards in certain courses of action. Novel situations and those where a direct response is not laid out can cause delays in interpretation and response time. Training in procedural steps to take in response to anticipated incidents and rehearsal of optimal responses in different scenarios provides a mental model for operators to follow in the event of the incident occurring. Huey and Wickens (1993) have pointed out that the participant's detection goals, expectancies about the nature of the stimuli and anticipated consequences of correct and incorrect responses contribute to the likelihood of a detection response. Training also has the potential to enhance the interest factor by making the operator more aware of the implications of detection and the broader consequences of detection or failure.

3.7.3 Technology-driven enhancement methods

Issues relating to the enhancement in the detection of signals were discussed earlier. However, technology increasingly has the capacity to assist rather than replace operator performance and by doing so, to enhance detection. Computer-aided signal identification can operate on a proactive basis, notifying operators of potential risk conditions, or on a reactive basis allowing operators to examine conditions with computer-aided investigation. Computer-aided risk analysis uses patterns, behaviour, or face recognition to highlight greater risk conditions where operators have to improve their vigilance levels. For example, the use of face recognition to identify and alert the operator to a number of people who have been classified as high risk and who have congregated together will be as relevant for surveillance of a football game as a diamond operation. Similarly, the use of different configurations of signals or measurement conditions can provide automated notification of various warning levels about conditions to operators in process control plants.

The amplification of critical signals through computer-aided image enhancements allows operators to view threat conditions in a number of ways. This can promote both vigilance and greater attention to objects that otherwise may have been seen as too ambiguous to be classified as an incident condition. For example, the image enhancements and different viewing methods (such as X-ray organic/inorganic differentiation) for X-ray screeners provides the option for operators to confirm judgements with greater confidence and contributes appreciably in reducing false alarms and misses. Another aspect to this is the use of computer-aided filtering to show only relevant information to the incident being examined and to reduce background noise and clutter.

3.8 Conclusion

Automation will have an increasing influence on technological systems where people are responsible for vigilance functions. However, both currently and in the future, personnel will retain their critical role in monitoring, interpretation and decision-making. One needs to ensure that personnel will be able to operate in a manner and environment that will allow them to be capable of high levels of vigilance and that this performance can be sustained during working periods. Vigilance enabling conditions need to be addressed from the initial operational requirements, the specification of job and work arrangements, and the design of the console and the physical operating environment. This needs to be followed up with what kind of people and skills are appropriate, and how they will be selected, trained and managed on an ongoing basis. An important part of the technology system implementation is how to ensure that technology facilitates vigilance rather than merely assuming some of the functions previously done by people and reducing it. It may well be that the role of people within a particular system is redefined and their responsibilities changed. However, this needs to be done in a manner that results in the creation of job roles and work processes that facilitate vigilance and operator decision-making. The cost of lowered vigilance levels and the consequences of problems in overall system performance can otherwise be extremely high.

References

1. BENNETT, T., and FOBES, J. L.: 'Screener proficiency and evaluation reporting system (SPEARS): human factors functional requirements and procedures'. FAA Technical Center, report: DOT/FAA/CT-94/65, 1994
2. DAVIES, D. R., and PARASURAMAN, R.: 'The psychology of vigilance' (Academic Press, London, 1982)
3. DONALD, C. H. M.: 'Implications of different surveillance roles', *Electronic Security Systems*, 1997, **3**, (5), pp. 46–7
4. DONALD, C. H. M.: 'Assessing the human vigilance capacity of control room operators'. Institution of Electrical Engineers, *People in Control* – An International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres, University of Bath, Bath, UK, June 21–23, 1999
5. HUEY, B. M., and WICKENS, C. D. (Eds.): 'Workload transition: implications for individual and team performance' (National Academy Press, National Research Council, Washington, DC, 1993)
6. MACKIE, R. R.: 'Vigilance research – are we ready for countermeasures?', *Human Factors*, 1987, **29**, (6), pp. 707–23

7 MACKIE, R. R., WYLIE, C. D., and SMITH, M.: 'Comparative effect of 19 stressors on task performance: critical literature review (what we appear to know, don't know, and should know)'. Proceedings of the Human Factors Society 29th Annual Meeting, Santa Monica, CA, (Human Factors Society, 1985) pp. 462–9

8 PARASURAMAN, R.: 'Memory load and event rate control sensitivity decrements in sustained attention', *Science*, 1979, **205**, pp. 924–7

9 PARASURAMAN, R.: 'Vigilance, monitoring and search', *in* BOFF, K. R., KAUFMAN, L., and THOMAS, J. P. (Eds.): 'Handbook of perception and human performance: volume II, cognitive processes and performance' (Wiley, New York, 1986), pp. 43.1–43.39

10 SAWIN, D. A., and SCERBO, M. W.: 'Effects of instruction type and boredom proneness in vigilance: implications for boredom and workload', *Human Factors*, 1995, **37**, (4), pp. 752–65

11 WALLACE, E., DIFFLEY, C., BAINES, E., and ALDRIDGE, J.: 'Ergonomic design considerations for public area CCTV safety and security applications' (Police Scientific Development Branch Publication, UK, 1995)

12 WICKENS, C. D.: 'Engineering psychology and human performance' (Harper Collins, New York, 1992)

13 WICKENS, C. D., MAVOR, A. S., and McGEE, J. P. (Eds.): 'Flight to the future – human factors in air traffic control' (National Academy Press, Washington, DC, 1997)

14 WIENER, E. L.: 'Application of vigilance research: rare, medium or well done?', *Human Factors*, 1987, **29**, (6), pp. 725–36

Chapter 4
Situation awareness

Carl Sandom

4.1 Introduction

This chapter will introduce a safety perspective into the development of systems that include people in control. Human systems are increasingly integrated into social contexts where their correct design and operation is essential in order to preserve the safety of the general public, the operators and the environment. However, many studies of human systems have in the past considered safety predominantly, or even exclusively, from a technical perspective. Unfortunately, these studies are typically limited to addressing the hazards arising through hardware failures alone, despite the fact that human failures are more common in safety-related systems.

A consideration of the human factors during systems development often reveals a complex set of problems that are extremely challenging. The hazards associated with human failures are very different from the hazards which have historically been the concern of system designers since they arise directly from the use of the system and therefore require some understanding of the cognition and actions of users in context. The identification of interaction hazards arising during system use may help designers to improve the system interface such that the associated risks are mitigated or even eliminated. However, in order to study these interaction hazards, appropriate models and techniques are required to help systems developers.

Situation awareness (SA) is one phenomenon that can be profoundly affected by the design of human–computer interactions; particularly when a system is situated in a dynamic environment (Hopkin, 1995). People in control of such systems must often pay attention to a large volume of information from a variety of sources including sensors and other operators in order to acquire an awareness of the situation in question. SA has therefore been the subject of much research in recent years, particularly within the field of aviation and other similarly complex domains (see, for example, Garland and Endsley, 1995; Harris, 1997; Sandom, 2000). Studies such as these have shown that SA should be a major safety consideration when developing interactive systems.

This chapter will examine the dominant perspectives of SA and the major themes will be drawn from this examination to form a Situated Cognition perspective. From this perspective, a generic Systems Situation Awareness Model is introduced to represent both the human and technical factors which affect the SA of people in control. A Situation Awareness Process Analysis Technique (SAPAT) for applying the Systems Situation Awareness Model to the analysis of interactive control systems will be introduced. Finally, the chapter will conclude with an explanation of how SAPAT can be used to identify those areas of an interactive system where safety should take precedence over usability.

4.2 Situation awareness – a perspective

Sarter and Woods (1991) identified SA as a critical, but ill-defined, phenomenon in complex, dynamic systems. SA has become a common phrase for both system designers and operators who often base its use on an intuitive understanding of its definition. Endsley (1995a) argued that a need for a commonly accepted definition is a particular requirement for practitioners attempting to design and evaluate systems that rely upon operator awareness.

In the context of human–machine interaction, current definitions of SA are generally based on conflicting views of SA as either a cognitive phenomenon or as an observer construct; these can respectively be referred to as the Cognitive or Interactionist perspectives. The Cognitive perspective is the most prevalent view of SA as a cognitive phenomenon that occurs 'in the head' of an individual. In contrast, the relatively new Interactionist perspective regards SA as an abstract concept located 'in the interaction' between actor and environment. Despite the philosophical differences, a number of themes can be drawn from these perspectives to form the basis of a Situated Cognition perspective of SA.

4.2.1 The cognitive perspective

Proponents of a cognitive perspective of SA view it as a phenomenon that occurs 'in the head' of an actor in a similar fashion to the dominant cognitive framework of the human as an information processor (Card *et al.*, 1983). Indeed, some theorists even suggest that SA is yet another 'black box' component or sub-process within the human information-processing model (see Endsley, 1995b). However, Cognitive perspective theorists often confusingly refer to SA as a cognitive process, a state of knowledge or both. With this distinction, *product* refers to the state of awareness with reference to knowledge and information, whereas *process* refers to the various cognitive activities involved in acquiring and maintaining SA. A typical process-oriented definition of SA has been proposed by Sarter and Woods (1991: 52):

> Situation awareness is the accessibility of a comprehensive and coherent situation representation which is continuously being updated in accordance with the results of recurrent situation assessments.

Cognitive definitions of SA also generally provide a rich description of key elements of decision-making activities in complex systems such as perception, comprehension and projection, as suggested by the definition of SA proposed by Endsley (1995b: 36):

> 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.

Having implied the process-oriented nature of SA, however, Endsley (1995c: 18) also confusingly states that 'SA can be described as a person's state of knowledge or mental model of the situation around them'. To add to this confusion, SA has also been defined as both a product and process as in the following definition by Issac (1997: 185):

> SA refers to a cognitive state or process associated with the assessment of multiple cues in a dynamic situation. It may refer to a person's knowledge and reference to their status within a space and time continuum (pilot) or an operator prediction within a known space and time continuum (air traffic controller).

These different definitions of SA suggest an apparent lack of coherence within the cognitive perspective of SA. Nonetheless, Endsley's (1995b) theoretical model of SA that is based on the role of SA in human decision-making in dynamic systems, has been widely cited and highly influential in cognitive science research. This model represents a typical cognitive perspective of SA and it proposes three different levels of SA which are relevant to this chapter:

- Level 1 SA. *Perception* of the status, attributes and dynamics of relevant elements in the environment.
- Level 2 SA. *Comprehension* of the situation based on a synthesis of disjointed Level 1 elements to form a holistic 'picture' of the environment.
- Level 3 SA. *Projection* of the near-term future of the elements in the environment.

Endsley's (1995b) model suggests that SA is based on more than simply perceiving information about the environment, which is often the intuitive definition of the phenomenon.

Many cognitive accounts of SA suggest that after information concerning relevant elements is perceived, a representation of the situation must be formed before a decision can be made based upon current SA. Kirwan *et al.* (1998) contended that air traffic controllers have a mental representation of the air traffic situation which includes what has happened, what could happen and what they would like to happen based on their goals and objectives. Kirwan *et al.* also suggested that this mental representation (or model) can be visual, verbal or both. Mental models such as these may be regarded as dynamic mental representations of a situation that allow operators to make predictions about future states and to make inferences regarding situations not experienced before. Clearly, there are striking similarities between this general definition of a mental model and Endsley's (1995b) process-oriented definition of SA given previously.

4.2.2 The interactionist perspective

In contrast to the cognitive school, there is a competing and developing view of SA which can be termed the Interactionist perspective. Interactionists share a common view of SA as an observed construct associated with the user's interaction with the system. From this perspective, SA is regarded to be an abstraction that exists only in the mind of an observer. SA is thus considered as a useful description of a phenomenon that can be observed in humans performing work through interacting with complex and dynamic environments (Billings, 1995; Flach, 1995a). The description is developed by considering observable behaviour in the environment – what the user does, how the system performs – but is not concerned with directly relating these with cognitive states of the user.

In one sense this might be associated with traditional behavioural psychology. A behavioural stance may simplify the discussion of SA by removing (or at least marginalising) interest in the user's mental state in favour of a reliance on observable action. A behaviourist stance is, however, much less rich as a research perspective, since no attempt will be made to relate action to intention on the user's part. In moving the SA debate forward, and looking for rich models to explain SA, identify hazards and ultimately inform the design of safety-related systems, it is suggested here that cognitive views of SA are necessary.

Yet, there are competing views of SA which do not fit neatly into the information-processing position predominantly taken by the cognitive school, but which might be useful in developing an informed stance on SA. Smith and Hancock (1995), for example, proposed a view of SA as adaptive and externally directed consciousness, arguing that there is currently an artificial and contentious division evident within the literature relating to general perspectives of SA as exclusively either knowledge (i.e. cognitive state, or product) or process.

From the interactionist view, SA specifies *what must be known* to solve a class of problems posed when interacting with a dynamic environment. Smith and Hancock (1995) also criticised the lack of dynamism exhibited in the cognitive perspective, contending that SA is a dynamic concept that exists at the interface between a user and their environment. Moreover, they argued that SA is a generative process of knowledge creation and informed action taking as opposed to merely a snapshot of a user's mental model.

There are merits in many of the competing perspectives of SA, and the range of views that exists highlights the complexity and the general immaturity of research in this area. The mental state of the user is important in trying to understand the awareness that the user builds up of a situation. Yet often only observable interaction data is available, tempting researchers to marginalise the mental state as a concern and focus on explaining SA without reference to the user's cognitive processes.

4.2.3 A situated cognition perspective

A synthetic, and perhaps pragmatic, perspective sees SA as a measure of the degree of dynamic coupling between a user and a particular situation (Flach, 1995b). This view attaches importance both to the user's cognitive state and to the context or situation in which they are interacting. This reflects a move away from traditional information processing models of cognition characterised by the ubiquitous model human processor proposed by Card *et al.* (1983) towards a situated cognition (and situated action) perspective.

Proponents of the situated perspective generally agree that the information processing approach to human–computer interaction (HCI) has neglected the importance of how people work when using computer systems situated in the context of the real world (see, for example, Winograd and Flores, 1986; Suchman, 1987; Hutchins, 1995; Nardi, 1996). Landauer (1987: 5) summed up this link between cognition and context aptly: 'There is no sense in which we can study cognition meaningfully divorced from the tasks and contexts in which it finds itself in the world'.

A Situated Cognition perspective of SA addresses how the current awareness of a situation effects the process of acquiring and interpreting new awareness in an ongoing cycle. This view is similar to Neisser's Perception-Action Cycle (Neisser, 1976) which has been used to model SA (see Adams *et al.*, 1995; Smith and Hancock, 1995) in an attempt to capture the dynamic nature of the phenomenon. Central to this view of SA is the contribution of active perception on the part of the user in making sense of the situation in which they are acting. Such active perception suggests informed, directed behaviour on the part of the user.

Neisser (1967) proposed a cognitive framework, which has been highly influential in cognitive psychology research into human behaviour in complex systems. His original framework partitioned the human information-processing system and subsequent research was directed at quantifying constraints, such as memory capacity, within each stage. Neisser (1976) subsequently expanded his model of cognition and he proposed the Perception-Action Cycle to reflect his assertion that *active perception* will unavoidably encounter unexpected situational elements or even fail to find them.

A tangible benefit of this perspective of SA is the focus on the inseparability of situations and awareness (Flach, 1995b). Discussions of SA focus attention on both what is inside the head (awareness from a cognitive perspective) and also what the head is inside (the situation which provides observable data) (Mace, 1977). Generally, this stance suggests that the user's current awareness of a situation affects the process of acquiring and interpreting new awareness from the environment in an ongoing cycle.

4.3 A situation awareness process model

As the preceding discussions have highlighted, there are competing and sometimes confusing views on SA and its relation to people and the situation in which they are acting. Four important themes will now be drawn from the theoretical perspectives discussed and these themes will be used as a framework to develop a model for the evaluation of the SA process within the human component of a system.

4.3.1 Awareness

As the discussion of the competing perspectives highlighted, the term SA is often used to describe the experience of comprehending what is happening in a complex, dynamic environment in relation to an overall objective or goal. Regardless of theoretical perspective, it is generally accepted that this experience involves both acquiring and maintaining a state of awareness (Endsley, 1995b; Smith and Hancock, 1995). This view is shared by Dominguez (1994) who, in an attempt to define SA as both a process and a product, compared 15 definitions and concluded that the perception of expected information in the environment occurs in a continual cycle which is described as *continuous extraction*. To be useful therefore, a perspective of SA should reflect the equal importance of both the continuous process of acquiring and maintaining SA and the state of SA itself.

4.3.2 Situated action

An area that is seen as important, but on which there is much academic disagreement, is consciousness. Compare, for example, the description of Endsley's (1995b) model of SA with that prescribed by Smith and Hancock (1995). If research in SA is to take a broader perspective than that offered by the information-processing model, it will have to concern itself with issues which reflect deliberate action on the part of those being studied in the specific context in which they are acting. A perspective informed by this stance would have to acknowledge the existence of situated action (Suchman, 1987), and reflect that an operator's awareness of a situation consciously affects the process of acquiring and interpreting new information in a continuous, proactive cycle.

4.3.3 Context

The positions taken on awareness and situated action reflect the importance of the operator making sense of situations in a particular context, and frame SA in this light. Any perspective of SA should explicitly reflect this, showing that accurate interpretations of a situation cannot be made without an understanding of the significance of the situation within a particular context. In other words, the context in which an operator is acting has to be understood in order to appreciate the importance of particular situations and their likely relationship to SA. This coupling of situation to context is suggested as a key issue and, as discussed earlier, is one that has emerged as a theme of increasing importance in cognitive science and HCI.

4.3.4 Dynamism

When an operator is making sense of the situation in which they are acting, their understanding is informed by them extracting relevant information from their environment. This information is temporal; the same information at different times (and therefore in different situations) may mean different things to an operator. The continuous information extraction process in which the operator is engaged implies that SA requires operators to diagnose past problems and provide prognosis and prevention of future problems based on an understanding of current information. This suggests that a perspective of SA must be inherently dynamic, reflecting the development of SA over time, and that it must be responsive to environmental changes, e.g. in the information available to the operator.

As can be seen, SA is an ill-defined, but critical phenomenon for operators in complex, interactive systems. However, as discussed here, one of the problems in making use of SA is the conflicting theoretical perspectives from which SA has been described and researched. Whilst it is recognised that theoretical debate is both healthy and necessary, it is suggested here that a situated cognition perspective may be a more immediate way of contributing to system design. The four themes outlined above form the basis of what can be described as a situated cognition approach to SA based upon a synthesis of important concepts from a review of the different theoretical perspectives. These themes are therefore used here to frame a model of the SA process shown in Figure 4.1.

The SA Process Model in Figure 4.1 is adapted from Neisser's Perception-Action Cycle (1976) which focuses on the adaptive, interactive relationship between an actor and their environment. Pictorially, the SA Process Model owes much to Boehm's Spiral Model of the software development life-cycle (Boehm, 1988) which is also centrally concerned with issues of iteration and dynamism. It also shows that awareness information is continuously extracted from a real-world situation and that this is integrated into an operator's awareness to form a mental representation upon which decisions are based and exploratory actions are taken.

The SA Process Model shows the inseparability of the SA acquisition process and the resulting (product) state of awareness that recursively direct the selection of

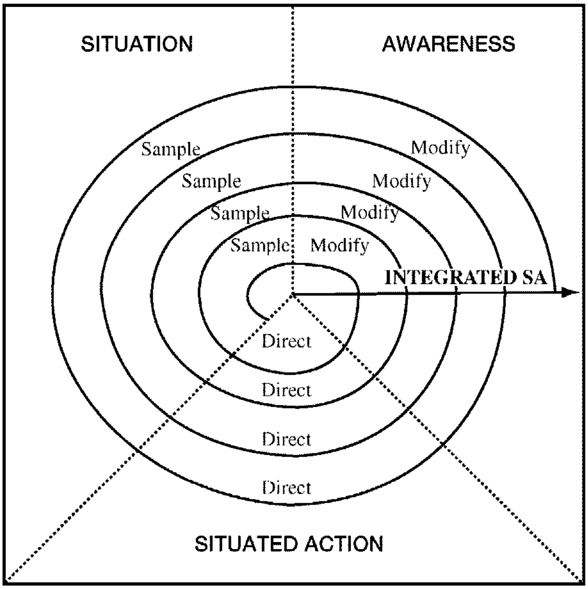


Figure 4.1. An SA process model (adapted from Neisser, 1976).

relevant situational information in a continuous cycle. It is worth noting that Norman's well-cited action model (Norman, 1988) appears very similar to Neisser's Perception-Action Model. An important difference, however, is that Neisser maintains that knowledge (or awareness) leads to anticipation of certain information that directs the sampling strategy and increases an operator's receptivity to some elements of the available information.

In Figure 4.1, the three terms 'sample', 'modify' and 'direct' are used. In Neisser's model, these terms are related to the 'environment', 'knowledge' and 'action' respectively. In the adapted SA Process Model the terms relate directly to the areas of situation, awareness and situated action. For the purpose of using Neisser's model in the context of SA, the terms 'situation' and 'awareness' are substituted for 'environment' and 'knowledge' to imply that only a subset of elements of the environment and knowledge relevant to a specific task are considered. This is consistent with the view of SA espoused by Endsley (1995b).

As the operators begin to interact in their environment, they can be considered as moving along the spiral in the model from the central point. Operators may start anywhere in the cycle as, for example, a routine may take over to provoke initial action. Starting arbitrarily, the operators will *sample* the situation, building a perception of it by extracting and interpreting information content. This may lead the operators to *modify* their awareness, developing their subjective mental representation of the situation in which they are interacting. Changes in the operators' interpretation of the situation cause them consciously to *direct* their action (including what/where to sample next), anticipating future states in which they might find themselves and acting accordingly. The 'sample-modify-direct' cycle, through which the operators can be thought of as having passed, will have developed their awareness in a particular way. As time progresses the operators will cycle through these phases, building an integrated awareness that grows with each iteration.

4.4 A system situation awareness model

An SA Process Model was developed in the previous section to represent the SA acquisition process of the human operator *in situ*. However, to carry out a true systems analysis, what is required is a model of the complete human and technical system in terms of SA. If we need to undertake an analysis of the hazards in an interactive control system we need to consider SA-related interactions from a controller's perspective. In particular, we need to identify those that are hazardous to enable the risks associated with these interactions to be mitigated.

It is therefore necessary to develop a model of a typical control system (based on its functionality) and the SA-related interactions that can affect the operator. Such an SA Interaction Model could then be integrated with the SA Process Model in Figure 4.1 (representing the human factor from an operator's perspective) to constitute a generic System SA Model. A model such as this is presented in Figure 4.2.

The System SA Model in Figure 4.2 is a generic systems model intended to represent a typical interactive system where people are in control of a process such as an air traffic control system, for example. These systems typically present situational data to an operator through a communications sub-system, a display and (possibly) a remote method of optimising the situational data collection sensors involving other system operators. For example, a railway control room operator will sample situational data from radio and/or telephone communications, a large screen display and a computer console, and the signals can be manually adjusted by signallers co-located with the remote signal sensors to optimise the railway system. Clearly, the model can be adapted to represent the interaction of a specific control system.

Together with the Systems SA Model in Figure 4.2, we now have a model to represent both the human and technical factors within a system. This model can now be used to undertake an analysis of the hazardous interactions that can affect the SA of an operator in the context of a control system's environment. Moreover, the model represents *both* the human and technical factors that can affect SA and the people in control of safety.

4.5 SAPAT – SA process analysis technique

Having described a model for the evaluation of SA, it is necessary to develop a method of applying the model to the analysis of interactive control systems and ultimately an evaluation of system safety. For theoretical coherence, a method that

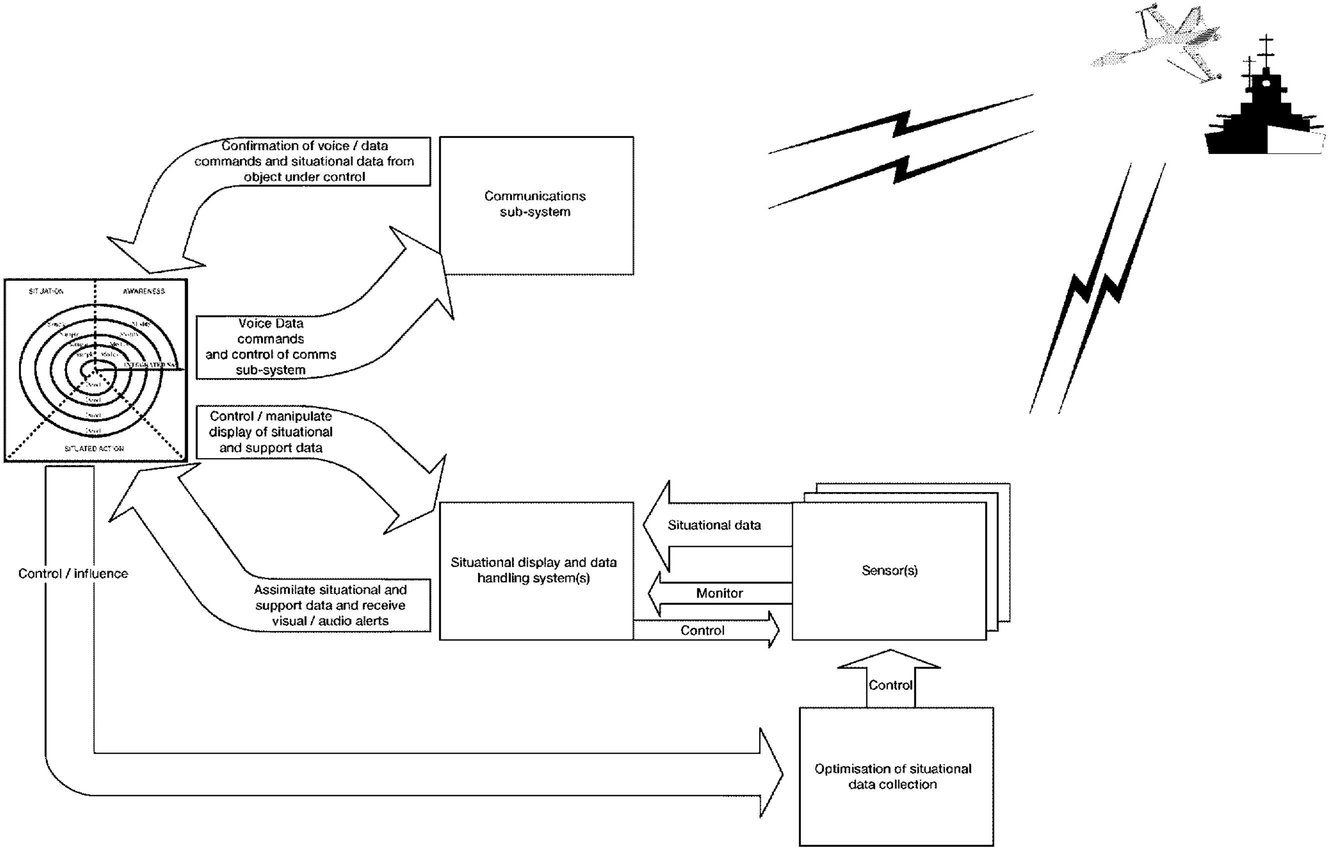


Figure 4.2. A system SA model.

uses the Systems SA Model (and specifically the integrated SA Process Model) as a tool must also be consistent with the Situated Cognition perspective of SA discussed in Section 4.2.3.

The SA Process Analysis Technique (SAPAT), shown in Figure 4.3, is a technique that can be used for the analysis of the SA process.

The SAPAT diagram in Figure 4.3 shows five distinct analysis stages and a brief description of these follows.

4.5.1 Stage I: identify system SA sources

The aim of this stage is to validate or amend the generic SA Interaction Model (shown in Figure 4.2) for the specific system under analysis. A tailored System SA Model can then be produced to represent both the human and technical factors in the context of use and it is then used as a framework for developing an initial list of interaction hazards. It is clearly not practical to analyse all system interactions and this system model enables a Preliminary Hazard Identification (PHI) to be undertaken to consider SA-related interactions from a controller's perspective and more specifically, to identify those which are hazardous.

4.5.2 Stage II: preliminary hazard identification

A PHI is undertaken in Stage II to focus the following analysis stages on the hazardous system interactions which are considered to be safety-related. The PHI can use the System SA Model validated in Stage I as the basis for the application of a HAZard and OPerability (HAZOP)-based technique (MoD, 1995) to identify SA-related hazards.

Briefly, the aim of HAZOP is to identify, in a comprehensive and structured manner, the hazard and operability problems that may be associated with an operation, process or system. HAZOP is a widely used and well-established hazard identification technique which is used in a range of industries (MoD, 1995). The technique is particularly useful for the identification of operator or system errors which may lead to hazard or operability problems. A summary of the HAZOP process is given in Figure 4.4.

The HAZOP technique involves a structured, systematic and comprehensive examination of designs or operations to identify potential hazard or operability problems. It can be seen from Figure 4.4 that a HAZOP begins with a system model (such as the System SA Model) identifying the interconnections between nodes or components within the system and determining the corresponding interactions. These interactions may consist of the physical flow of material from one node to another or, for information systems, may represent the flow of data between components. Each system component possesses certain attributes denoting correct system operation, for example the *value* or *latency* of situational data may be important in a specific context. For each node, the effects of deviations from these

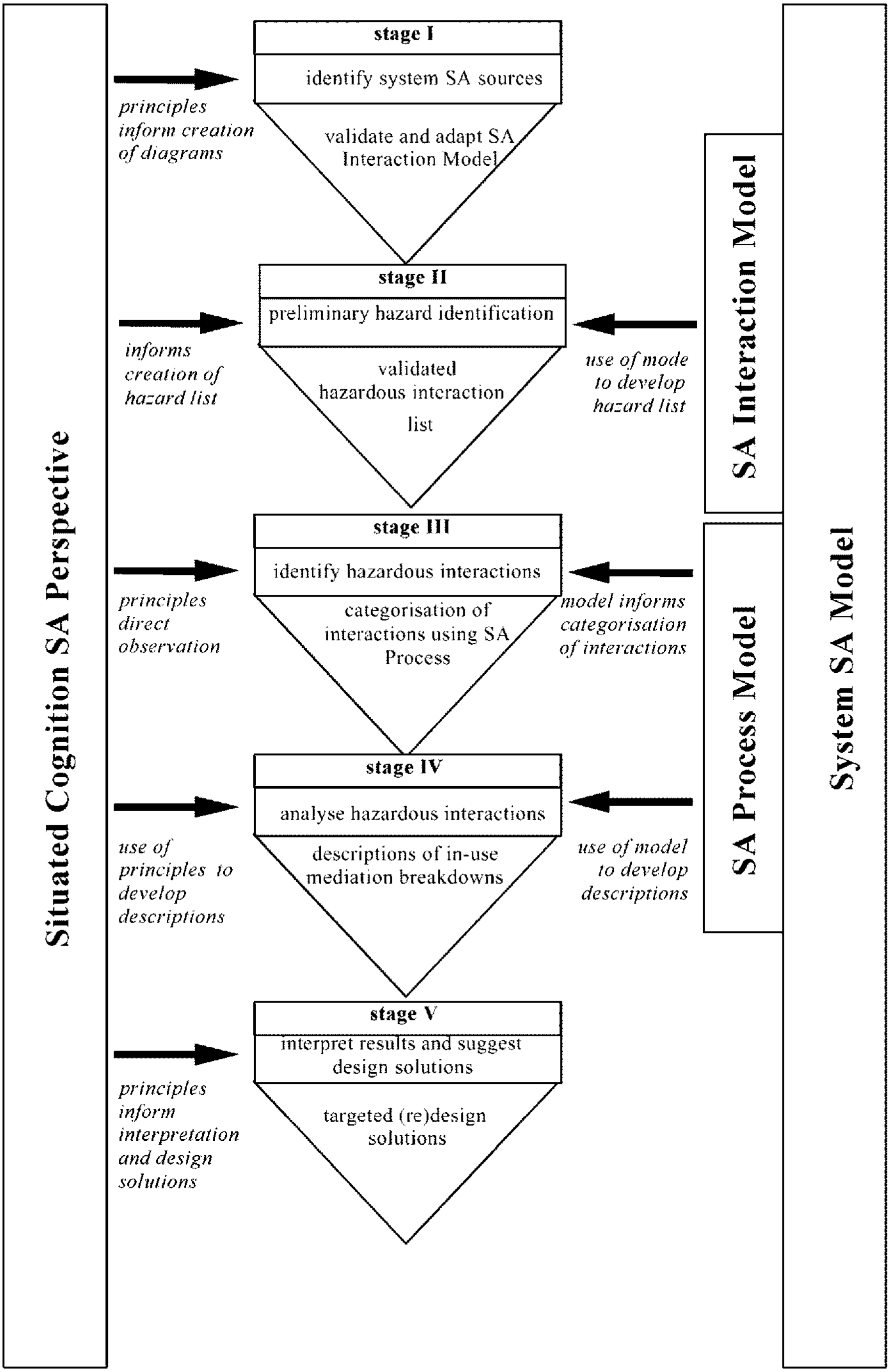


Figure 4.3. SA process analysis technique (SAPAT).

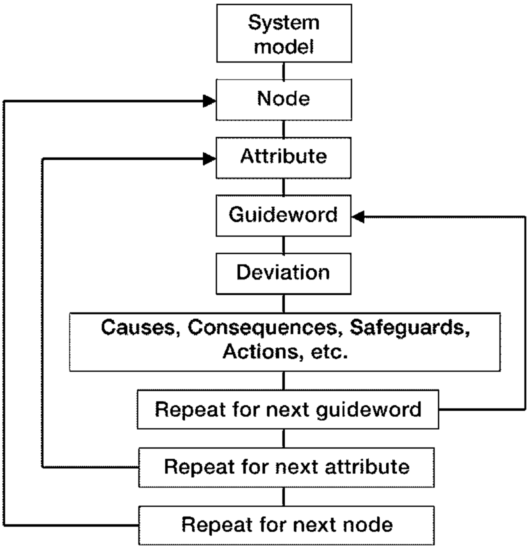


Figure 4.4. HAZOP summary.

attributes are considered using appropriate guidewords such as *inaccurate* or *none*. A HAZOP analysis will consider each system component or node in turn, as shown in Figure 4.4. A detailed explanation of the HAZOP technique can be found in MoD (1995).

4.5.3 Stage III: identify hazardous interactions

In this stage, initial problem actions and operations resulting from interaction breakdowns are identified using the SA Process Model and applying the principles of Situated Interaction to direct the observation. The division of the SA Process Model into different areas of operator activity (sample-modify-direct) provides a structure to analyse and categorise SA-related problems.

4.5.4 Stage IV: analyse hazardous interactions

The aim of this stage is to analyse and describe the observed in-use interaction breakdowns using the SA Process Model and applying the Situated Interaction perspective as a guiding framework. The structure of the SA Process Model partitions different areas of interest and it enables the interaction boundaries between these partitions to be considered separately.

4.5.5 Stage V: interpret results and suggest safe design solutions

In this final stage, the findings from the preceding stages are interpreted. An understanding of the possible situated interaction breakdowns and their associated hazards will lead to informed re-design solutions which can be justified from a system safety perspective. Generally, SAPAT analyses, concerning interaction breakdowns and automated interactions, deal with the design trade-off between usability and safety. SAPAT can therefore be used to identify those areas of an interactive system where safety should take precedence.

4.6 SAPAT and safe interactions

It has been suggested here that SAPAT can be used as a framework for the identification and analysis of situated hazards relating to operator awareness in the context of system use. Specifically, there are two ways in which SAPAT can contribute to the design of safer systems: identifying interaction breakdowns and identifying automatic interactions, both of which are key to SA. The hazards associated with these interactions can be related to the concepts of conscious and automatic cognition.

Differentiating between these two modes of cognition using SAPAT enables us to highlight and compare different aspects of human action which will be of use to the improved design of safety-related systems and these are discussed in the following sections.

4.6.1 Automatic interactions

It is possible to use SAPAT to identify hazardous interactions which are carried out automatically without the operator modifying their awareness. If the specific interaction has been identified in the SAPAT PHI stage as hazardous, it is possible to design the system to prevent an automatic interaction. A simple example shows how this can be achieved. An exit menu used in a control system may ask the operator the final question:

Are you sure you want to exit? Y/N.

Typically, such a system would use a Windows, Icons Menus and Pull Down (WIMP) style of interface. This particular interaction is often designed so that the operator can select with a mouse from two buttons marked either 'Yes' or 'No' which are always positioned in the same location relative to each other on screen. An analysis of this type of interaction would typically show that the action of selecting a button from the exit menu on a normal system shutdown will develop into an automatic operation without the operator being consciously aware of the interaction – until an erroneous menu selection is made.

This is an example of an *unplanned* automatic operation that is carried out without the conscious formulation of a plan. Reason (1990) asserted that the term *human* error can only be meaningfully applied to planned actions that fail to achieve their desired consequences without some unforeseeable intervention. It has been argued here that *system* errors can be caused through automated, unplanned operator interactions and these can be identified through the application of SAPAT. This example was not particularly hazardous in this context; however, a simple design solution to this may be to require a text string (not simply one keypress) in response to the question, 'Are you sure you want to exit? Y/N.' Forcing the operator to input a text string would increase the probability that the question is consciously considered and a plan of action is formulated before the action is carried out.

There are also more subtle automated interactions that can lead to what Reason (1990) has called the knowledge-based mistake. These automated interactions are ones where an interaction has developed from a conscious action to an automatic operation and SA-related information is not assimilated as a result. This can lead to incomplete operator awareness which, in a specific context, can result in the operator formulating the wrong plan and making what Reason calls the knowledge-based mistake. To reduce the risk associated with these *planned* automatic interactions, it is again necessary to design the interaction to force it to become conscious to increase the probability of the operator's awareness being modified. Note that it is only suggested here that this will increase the probability of the operator's awareness being modified as a breakdown may also occur during the 'Modifying Awareness' stage of the SA Process Model due to a distraction, for example.

4.6.2 Interaction breakdowns

The SA Process Model used within SAPAT can also provide design guidelines relating to analysing difficulties that affect the user-system coupling, such as interaction breakdowns. The division of the model into areas of activity on the individual's part (sample-modify-direct) provides a structure to analyse and categorise SA-related problems. For example, the SA Process Model could be used to question where the problems in particular situations might have arisen: what information did the operator sample from their environment? How did this lead them to modify their awareness? (What information was available through the interface?). And how, subsequently, did this direct the operator's situated actions?

The structure of the SA Process Model partitions different areas of interest to allow system developers to concentrate on each as a distinct dimension contributing to awareness which in turn can bring its own set of potential problems. It also permits a consideration of the interaction boundaries between these partitions, which is where many SA difficulties can be identified. As operators integrate sampled information, for example, the erroneous modification of their awareness may lead to an overall reduction in SA.

Clearly, substituting a text string input for a WIMP's button selection, as in the example above, will adversely affect usability metrics relating to the speed of interaction for example. However, for hazardous interactions identified in SAPAT Stage II, safety is more important and this leads to the following general design guideline for safety-related systems:

> If an interaction is potentially hazardous, and the design will allow it to develop from a conscious action to an automatic operation, the design of the interaction should force the interaction to remain a conscious action. (Sandom, 2000: 134)

SAPAT analyses which identify interaction breakdowns and automated interactions can help systems designers to deal with the trade-off between usability and safety. SAPAT can be used to identify those areas of an interactive system where safety should take precedence. For a detailed explanation of how SAPAT has been used for an exploratory analysis of interaction safety in a complex system where people in control can affect safety, see Sandom (2000).

4.7 Conclusions

Developing an understanding of how the SA of an operator can be affected by the design of systems interactions within the context of use is an important safety issue. People in control of complex systems interact and operate using a remarkable cognitive process which requires the creation and maintenance of SA. However, there are many different definitions of SA and this chapter provided a review of the literature from which a Situated Cognition perspective was presented.

A system comprises both human and technical components and a comprehensive system model must therefore address both of these issues. An SA Process Model was introduced to model the process of a human operator acquiring and maintaining SA *in situ*. A generic SA Interaction Model of a typical control system was also described which is based on a system's functionality and the SA-related interactions that can affect the operator. Together the SA Process Model and the SA Interaction Model constitute a generic System SA Model and this represents both the human and technical factors that can affect the SA of people in control.

In conclusion, the design of a control system will normally entail many trade-offs including that between safety and usability. During system development, human factors experts will often consult with the intended end users of the system who will invariably support the view that system interactions should be made 'simple' and 'intuitive'. However, in the context of a system's use, it may be hazardous for people in control to interact without consciously considering their actions and the SAPAT was developed specifically to help systems designers to identify when safety should take precedence over usability.

References

1 ADAMS, M. J., TENNEY, Y. J., and PEW, R. W.: 'Situation awareness and the cognitive management of complex systems', *Human Factors*, 1995, **37**, (1), pp. 85–104

2 BILLINGS, C. E.: 'Situation awareness measurement and analysis: a commentary', *in* GARLAND, D. J., and ENDSLEY, M. R. (Eds.): 'Experimental analysis and measurement of situation awareness' (Proceedings of International Conference, FL, USA, November 1995)

3 BOEHM, B. W.: 'A spiral model of software development and enhancement', *IEEE Computer*, 1988, pp. 61–72

4 CARD, S. K., MORAN, T. P., and NEWELL, A.: 'The psychology of human computer interaction' (Lawrence Erlbaum Associates, Hillsdale, NJ, 1983)

5 DOMINGUEZ, C.: 'Can SA be defined?', *in* VIDULICH, M., DOMINGUEZ, C., VOGEL, E., and McMILLAN, G. (Eds.): 'Situation awareness: papers and annotated bibliography', 5–15, Report AL/CF-TR-1994-0085, (Wright-Patterson AFB, Ohio, 1994)

6 ENDSLEY, M. R.: 'Measurement of situation awareness in dynamic systems', *Human Factors*, 1995a, **37**, (1), pp. 65–84

7 ENDSLEY, M. R.: 'Towards a theory of situation awareness in dynamic systems', *Human Factors*, 1995b, **37**, (1), pp. 32–64

8 ENDSLEY, M. R.: 'Theoretical underpinnings of situation awareness: a critical review', *in* GARLAND, D. J., and ENDSLEY, M. R. (Eds.): 'Experimental analysis and measurement of situation awareness' (Proceedings of International Conference, FL, USA, November 1995c)

9 FLACH, J. M.: 'Situation awareness: proceed with caution', *Human Factors*, 1995a, **37**, (1), pp. 149–57

10 FLACH, J. M.: 'Maintaining situation awareness when stalking cognition in the wild', *in* GARLAND, D. J., and ENDSLEY, M. R. (Eds.): 'Experimental analysis and measurement of situation awareness' (Proceedings of International Conference, FL, USA, November 1995b)

11 GARLAND, D. J., and ENDSLEY, M. R. (Eds.): 'Experimental analysis and measurement of situation awareness' (Proceedings of International Conference, FL, USA, November 1995)

12 HARRIS, D. (Ed.): 'Engineering psychology and cognitive ergonomics volume 1: transportation systems'. Proceedings of the 1st International Conference, EP&CE, Stratford-upon-Avon, October 23–25, 1996 (Ashgate Publishing, Aldershot, 1997)

13 HOPKIN, V. D.: 'Human factors in air traffic control' (Taylor and Francis, London, 1995)

14 HUTCHINS, E.: 'Cognition in the wild' (MIT Press, Bradford, 1995)

15 ISSAC, A. R.: 'Situation awareness in air traffic control: human cognition and advanced technology', *in* HARRIS, D. (Ed.): 'Engineering psychology and cognitive ergonomics volume 1: transportation systems' (Ashgate Publishing, Aldershot, 1997)

16 KIRWAN, B., DONOHOE, L., ATKINSON, T., MACKENDRICK, H., LAMOUREUX, T., and PHILLIPS, A.: 'Getting the picture: investigating the mental picture of the air traffic controller' Proceedings of the Annual

Conference of the Ergonomics Society, (Taylor & Francis, London, 1998) pp. 405–8

17 LANDAUER, T. K.: 'Relations between cognitive psychology and computer systems design', *in* CARROLL, J. M. (Ed.): 'Interfacing thought: cognitive aspects of human-computer interaction' (MIT Press, Cambridge, MA, 1987)

18 MACE, W. M.: 'Ask not what's inside your head but what your head's inside of', *in* SHAW, R. E. and BRANDSFORD, J. (Eds.): 'Perceiving, acting and knowing' (Erlbaum, Hillsdale, NJ, 1977)

19 MoD: 'A guideline for HAZOP studies on systems which include a programmable electronic system'. Draft Interim Defence Standard 00–58, July 1995

20 NARDI, B. A. (Ed.), 'Context and consciousness: activity theory and human-computer interaction' (MIT Press, London, 1996)

21 NEISSER, U.: 'Cognitive psychology' (Appleton, Century-Crofts, New York, 1967)

22 NEISSER, U.: 'Cognition and reality: principles and implications of cognitive psychology' (W. H. Freeman, San Francisco, 1976)

23 NORMAN, D. A.: 'The psychology of everyday things' (Basic Books, New York, 1988)

24 REASON, J.: 'Human error' (Cambridge University Press, Cambridge, 1990)

25 SANDOM, C.: 'Situation awareness through the interface: evaluating safety in safety-critical control systems'. Proceedings of IEE People in Control: International Conference on *Human Interfaces in Control Rooms, Cockpits and Command Centres*, Conference Publication No. 463, University of Bath, June 21–23, 1999

26 SANDOM, C. W.: 'Situation awareness and interactive system safety analysis' (Ph.D. Dissertation, Department of Information Systems and Computing, Brunel University, UK, 2000, pp. 119–124)

27 SARTER, N. B., and WOODS, D. D.: 'Situation awareness: a critical but ill-defined phenomenon', *International Journal of Aviation Psychology*, 1991, **37**, (1), pp. 45–57

28 SMITH, K., and HANCOCK, P. A.: 'Situation awareness is adaptive, externally directed consciousness', *Human Factors*, 1995, **37**, (1), pp. 137–48

29 SUCHMAN, L.: 'Plans and situated actions' (Cambridge University Press, Cambridge, 1987)

30 WINOGRAD, T., and FLORES, F.: 'Understanding computers and cognition: a new foundation for design' (Ablex, Norwood NJ, 1986)

Chapter 5
Teamworking

Raphael Pascual, Matthew Mills and Simon Henderson

5.1 Introduction

> Greater than fifty percent of the nuclear industry's events which occur are attributable to human performance problems. A significant portion of these events is due to some breakdown in co-ordination among members of the nuclear control room teams (Frye, 1988: 574).

This chapter discusses the importance of effective teamwork in the successful operation of control room environments. It provides an initial review of core teamwork competencies and, together with Chapter 9, outlines a number of approaches and principles for team training and technology support in these settings.

Teams, both military and civil, operate in complex and dynamic control room environments. Such teams include those working in military headquarters, nuclear power plant control rooms, on the bridges of large ships and in train operation control centres.

Although the functions of the teams and the roles of the individuals within them may vary, the quality of their teamwork behaviour and performance is often influenced by a set of common factors. For many of these teams, the day-to-day operation of the control room involves overseeing systems monitoring the environment, the conduct of safety checks and the testing of equipment. In these circumstances, decision-making and team co-ordination whilst important, is relatively proceduralised (Gaddy and Wachtel, 1992). The real challenges for effective teamwork arise in response to emergency situations, such as the failure of a cooling system in a nuclear power plant, or the arrival of unidentified aircraft near a warship. At these times, control room teams are frequently faced with the interpretation of large amounts of information from multiple systems, ambiguous goals, high time and risk stakes and the unstructured and rapidly changing nature of many of the problems faced. The requirement for accurate and timely decision-making, shared situation awareness and co-ordinated team performance is particularly critical if mistakes are not to be made.

Another operating facet of many modern organisations is the increasing use of distributed control team structures. In the military domain, for example, dispersing control personnel and assets help reduce their vulnerability to enemy attack. Advances in technology have also led civil organisations to distribute control room functions. Many of these teams communicate and maintain co-ordination using networked electronic media (e.g. through the use of GroupWare, Internet and video-conferencing facilities). Even when these personnel may share the same physical space, the requirement to monitor continuously complex automated control systems often results in severely limited opportunities for face-to-face communication and for performance feedback. For both military and civil control room structures, the distribution of team members makes it more difficult to maintain a shared understanding of both the operating environment and of the actions of colleagues. Similarly, team leaders have more difficulty in imparting their presence at remote locations and in communicating and subsequently monitoring expectancies for future team member behaviour (Mills *et al.*, 1999).

The challenges to team performance are also compounded by the fact that the composition of control room teams is often subject to change. This may be due to shift rotations in order to keep the control room continuously operational, individuals moving post, or illness among team members. Other, more transparent examples of *ad hoc* control room teams include those rapidly formed to co-ordinate the response to a major train crash, or the multinational military team brought together to co-ordinate disaster relief efforts. Team members may be relatively unfamiliar with each other, have had limited opportunities to train together, but must still sustain high levels of team performance. Any instability in the team composition makes it more difficult for team members to build up a shared understanding of team norms and to respond in a co-ordinated manner to critical situations. Team identity and cohesion may also be lower, creating the conditions where implementing essential team behaviours, such as providing situation feedback and backing-up colleagues, becomes harder to achieve effectively (Pascual *et al.*, 1999).

5.2 Control room team failures

Given the complex operating characteristics of control room settings, it is perhaps not surprising to learn that history is littered with examples of teamwork failures contributing to the poor resolution of critical incidents. A few of these are briefly discussed here.

In 1988, the Aegis cruiser *USS Vincennes* was patrolling the waters of the Persian Gulf as part of the enforcement of an embargo on Iran. In the midst of a minor skirmish with Iranian gunboats, the command information centre (CIC) also tracked, engaged and destroyed a commercial Iranian airliner, killing all aboard

(Klein, 1998). Although a number of factors underpinned the disaster, the teamwork dynamics in the CIC during the incident compounded the situation. Relevant teamwork issues included the development of 'groupthink' among the control room personnel, as they reinforced each other's mistaken perceptions concerning the commercial aircraft's identity and intentions. Information emerging from the complex monitoring systems that appeared to contradict their fears was also either ignored or misinterpreted, leading to breakdowns in shared situation awareness. Communications protocols became fragmented and role conflict occurred within the control room team, as officers became distracted managing multiple tasks.

In 1990, two fires broke out on a passenger liner, the *Scandinavian Star*, which was cruising between Norway and Denmark. As a result, 158 people were killed. The ship's command was considered to have failed in their ability to maintain a shared understanding of the unfolding situation, with no protocols having been established for the transmission of status updates. Information was distributed in a haphazard and inconsistent fashion among crew and passengers, and co-ordination with other vessels in the area during the incident was also judged to have been poor. In addition, many of the team leaders aboard the liner did not hold a shared understanding of the emergency procedures and plans, leading to considerable confusion in the aftermath of the fires (Flin, 1996).

Deficiencies in teamwork communications and co-ordination have also been cited in relation to the aetiology of various accidents in the nuclear industry. In 1979, an event occurred at the Three Mile Island Unit 2 reactor in the United States that resulted in melted fuel, prior to the situation being brought back under control. As the situation unfolded, operating and monitoring personnel persistently failed to recognise critical information cues and to co-ordinate their responses accordingly (Rogovin *et al.*, 1980).

Several years later in 1986, the Unit 4 reactor at the Russian Chernobyl installation exploded during the test of the plant's turbine-generator system, resulting in 30 fatalities and widespread radioactive contamination. Among the human factors issues contributing to the disaster were failures in pre-planning for the inter-team co-ordination requirements for such an incident, poor demarcation of roles and responsibilities and breakdowns in communications among team members (NEA, 1995).

As a result of the lessons learned from these and other similar incidents, organisations have initiated programmes of research in order to understand better the nature of control room teamwork and teamwork failures (Cannon-Bowers and Salas, 1998). There has also been a drive to improve the form and content of training courses, to include teamwork skills training and to enhance shared team learning opportunities (Gaddy and Wachtel, 1992). In addition, increasing attention has been paid to control room layout and the design of technology support systems (NEA, 1998). Ultimately, the aim is to help prevent accidents and, when necessary, to improve incident management. The remainder of this chapter seeks to provide a better understanding of what is meant by the notion of a team and the type of teamwork competencies that underpin the support concepts and imperatives that are discussed in Chapter 9.

5.3 Defining the team and teamwork

Although many definitions of a team have been evolved over the years, the following by Salas *et al.* (1992) provides a clear statement of a team's key characteristics. A team is defined as 'a set of two or more people who interact dynamically, interdependently and adaptively towards a common and valued goal'. Salas *et al.* also argued that team members have specified functions or roles and the team itself has a limited life span of membership.

In a similar fashion, researchers have evolved various models to explain the way in which teams operate, the factors that influence their performance and to provide valuable insights into team training requirements. A useful starting point is to examine the Team Evolution And Maturation (TEAM) model (see Figure 5.1) developed by Morgan, Salas and Glickman (1994). The TEAM model describes how a task-oriented team matures over a number of developmental stages. This is based on Tuckman's (1965) well-known team development stages of forming, storming, norming and performing. Morgan extended this model to include 'pre-forming' and 'deforming' stages, arguing that teams do not materialise out of thin air and usually do not remain intact forever. A 're-forming' phase was also added, to highlight how effective teams will try establishing new structures and methods of working to seek additional team performance benefits.

The other important concept in the TEAM model is the inclusion of two separate tracks of skill and knowledge development, *taskwork* and *teamwork*. Although they start widely separated, they need to become increasingly integrated as the team matures in order to achieve successful team performance. The taskwork track

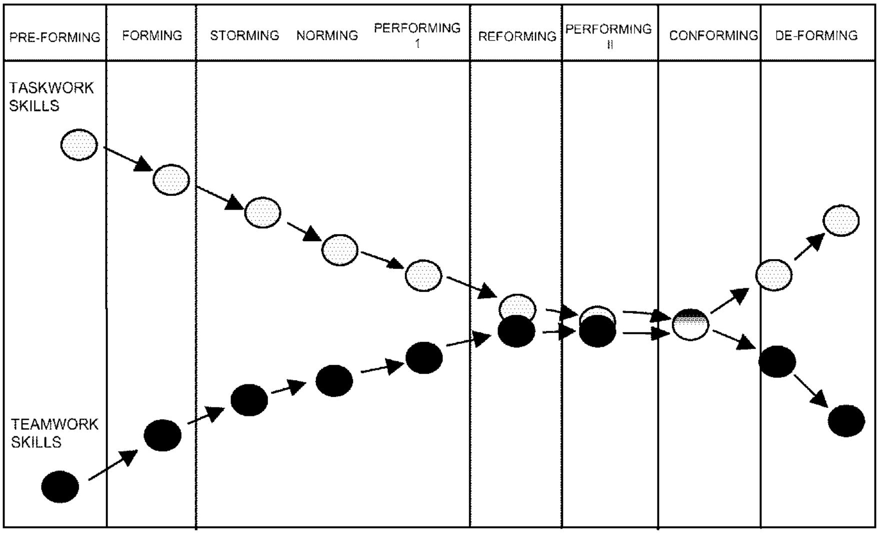


Figure 5.1. Simplified version of the team evolution and maturation (TEAM) model.

involves the operations-related activities of the team and focuses on the technical accomplishment of individual and team tasks. For example, knowledge of the procedures involved in correctly responding to a control room alarm signal. The teamwork track includes those activities that enhance the quality of the functional interactions, cohesion, communication and co-ordination of team members. For example, team members providing regular feedback to each other to maintain a shared awareness of overall control room workload.

Both of these development components are equally important for effective team performance. However, in the past, organisations have failed to pay sufficient attention to the training of teamwork skills and the development of the underpinning knowledge to apply those skills effectively. In part, this has been due to the difficulties of defining exactly what needs to be trained, and how to relate teamwork training to task requirements within organisational constraints.

A useful way of visualising the complexity of teams and their inherent make-up is through a process model. The model illustrated in Figure 5.2 is a simplified version that has been tailored to a generic command and control team domain, and is loosely based on an original team model concept by McGrath and Hollingshead (1994). The model is essentially made up of a series of input, teamwork process and output variables.

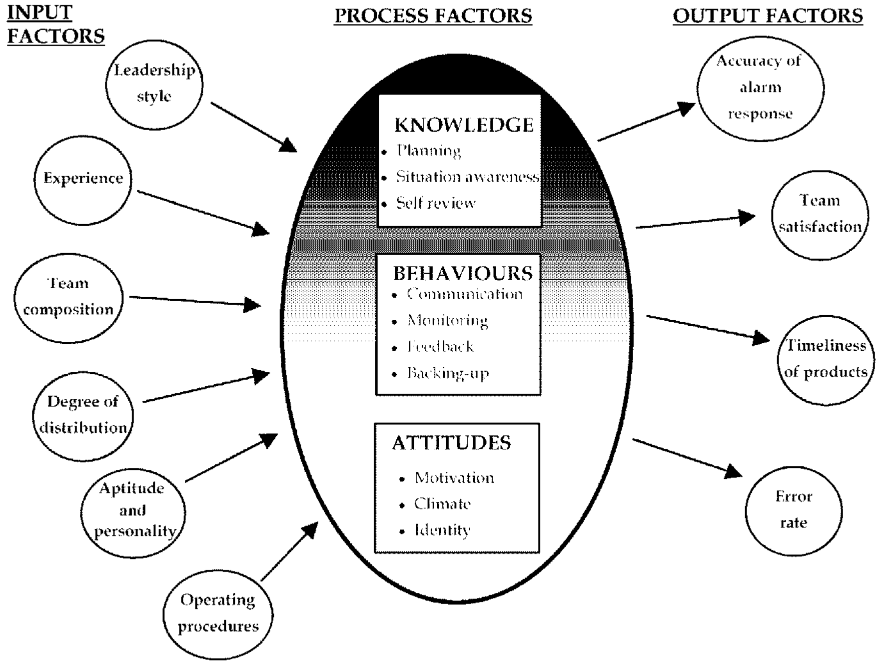


Figure 5.2. Team process model (after McGrath and Hollingshead, 1994).

The example input factors highlighted in Figure 5.2 include external environmental conditions (e.g. physical, sociocultural and temporal factors) imposed on the team and individual and group properties (e.g. leadership styles, degree of shared team experience). These factors shape the ways in which the team operates and the nature of the teamwork required. For example, tasks characterised by low complexity and organisation within a hierarchical work setting are likely to require limited communication between team members (Swezey and Salas, 1992).

Outcome factors are largely domain specific and are usually both readily recognisable and measurable, e.g. the amount of targets missed by a military control room team. Measures of team effectiveness should be taken at both the outcome and the process levels, in order to diagnose and evaluate fully the quality of team performance.

The effective conduct of the teamwork processes described, increases the team's potential for producing desirable outcomes, whilst reducing its potential for producing undesirable products. In this representation, effective teamwork is founded on 10 core teamwork principles, which can be broadly divided into the areas of teamwork knowledge, behaviours and attitudes. These principles, drawn from the outputs of a range of theoretical models and empirical studies (Cannon-Bowers *et al.*, 1990; McIntyre and Salas, 1995; Orasanu and Fischer, 1997) are summarised briefly here.

- **Planning and goal setting**. These are key processes by which a team establishes who will do what, with whom, when and why. Clarifying interdependencies enhances team co-ordination, particularly in distributed structures. Failure to adequately plan future responses is likely to lead to degraded co-ordination in emergency situations, with team members suffering from role and responsibility conflicts.
- **Situation awareness**. At the team level, maintaining shared situation awareness enables team members to hold a common understanding of what is currently happening in the working environment and to anticipate future situation changes and, thus, colleagues' requirements. The realisation of effective shared situation awareness in teams is strongly related to the way in which other teamwork competencies are maintained during task performance.
- **Team self-review**. Teams should review their teamwork performance, either after an exercise or incident, generating examples of team strengths, as well as identifying where and why teamwork broke down, was hindered or challenged. Such reviews develop team learning and can also be conducted during exercises to improve team approaches to dynamic task conduct.
- **Monitoring**. Team members should aim to keep track of the activities of their colleagues to help catch errors, or problems, before they escalate. Monitoring requirements and processes should be explicitly identified during planning, thus strengthening awareness of team interdependencies.
- **Feedback**. This can be viewed as the follow-up activity to monitoring. Team members should offer goal-focused taskwork and teamwork opinions and advice to colleagues. This not only helps maintain shared situation awareness,

but also helps promote team identity and cohesion in distributed team structures.

- **Backing-up**. Team members should aim to intervene pro-actively to assist colleagues when a need for help is perceived, rather than waiting to be asked. This is particularly important in crisis situations, when workload in the control room is likely to rise.
- **Communication**. Effective communication underpins all aspects of teamwork knowledge, behaviour and attitudes, and should be characterised by brevity, clarity and lack of ambiguity. Effective teams are explicit in making predictions and warnings, providing explanations and stating goals and strategies. They are also efficient in their communications in high workload situations. Team members should also engage in 'closed-loop communication', ensuring that information sent is interpreted by the recipient as intended.
- **Climate**. This relates to the degree to which team members are satisfied and enjoy working within the team. Team members facilitate effective operational climates created by proactively intervening to alleviate potential friction within the team, and resolving problems and anxieties as quickly as possible.
- **Identity**. Effective teams understand the performance benefits to be gained from working in teams and see their team's success as taking precedence over individual successes. Effective teams increase team identity by involving all team members in planning processes and by reinforcing the relationships between individual tasks and team goals.
- **Motivation**. In a team context, the motivation to engage in effective teamwork and taskwork will influence team performance. Team motivation is influenced by the explicit support of task identity and significance, by reinforcing task interdependencies, and through the provision of timely feedback to *all* team members on their outputs.

5.4 Conclusions

Team science has received much research attention in the last 25 years, as organisations have become more project and team centric. The importance of understanding and supporting team processes is now widely considered a key requirement for promoting organisational effectiveness. This chapter has provided a brief overview of current theory in the team science domain. Laboratory and field-based team research has generated a number of valuable theories, models, conclusions and recommendations concerning effective teamwork. It is hypothesised that teamwork incorporates cognitive, behavioural and attitudinal competencies and is as critical as taskwork for effective control room operations. Analyses of a number of large-scale accidents have shown that over half originate from breakdowns in control room teamwork. Many control room failures can be explained in terms of a lack of specific teamwork competencies. Consequently, team research is now beginning to pinpoint and explain teamwork competencies critical for promotion and support in control room environments. In Chapter 9, a systems approach to the development and implementation of team training will be outlined, providing a framework within which these teamwork principles can be taught and subsequently practised by control room teams. In a similar fashion, Chapter 9 will also discuss the relationship between these competencies and the design of technology to support control room teamwork.

References

1 CANNON-BOWERS, J. A., and SALAS, E.: 'Making decisions under stress' (American Psychological Association, Washington, DC, 1988)

2 CANNON-BOWERS, J. A., SALAS, E., and CONVERSE, S. A.: 'Cognitive psychology and team training: training shared mental models and complex systems', *Human Factors Society Bulletin*, 1990, **33**, pp. 1–4

3 FLIN, R.: 'Sitting in the hot seat' (Wiley, Chichester, 1996)

4 FRYE, S. R.: 'An approach to enhanced control room crew performance'. Paper presented at 4th Conference on *Human factors and power plants*, Monterey, CA, IEEE, 1988

5 GADDY, C. D., and WACHTEL, J. A.: 'Team skills training in nuclear power plant operations', Chapter 14 *in* SWEZEY, R. W., and SALAS, E. (Eds.): 'Teams, their training and performance' (Ablex, Norwood, NJ, 1992)

6 KLEIN, G.: 'Sources of power' (The MIT Press, Cambridge, MA, 1998)

7 MCGRATH, J. E., and HOLLINGSHEAD, A.: 'Groups interacting with technology' (Sage, Thousand Oaks, CA, 1994)

8 MCINTYRE, R. M., and SALAS, E.: 'Measuring and managing for team performance: lessons from complex environments', *in* GUZZO, R. A., and SALAS, E. (Eds.): 'Team effectiveness and decision-making in organisations' (Jossey-Bass, San Francisco, 1995)

9 MILLS, M. C., PASCUAL, R. G. *et al.*: 'Supporting distributed team interaction'. Unpublished DERA Report, 1999

10 MORGAN, B. B., SALAS, E., and GLICKMAN, A. S.: 'An analysis of team evolution and maturation', *Journal of General Psychology*, 1994, **120**, (3), pp. 277–91

11 NEA: 'Chernobyl ten years on – radiological and health impact'. An assessment by the NEA Committee on Radiation Protection and Public Health, OECD Nuclear Energy Agency, November 1995

12 NEA: 'Issue brief – an analysis of principal nuclear energy issues', *The human factor in nuclear power plant operation*, 1998, (2)

13 ORASANU, J., and FISCHER, U.: 'Finding decisions in natural environments: the view from the cockpit', Chapter 32 *in* ZSAMBOK, C. E., and KLEIN, G. (Eds.): 'Naturalistic decision-making' (Lawrence Erlbaum Associates, New Jersey, 1997)

14 PASCUAL, R. G., MILLS, M. C., and BLENDELL, C.: 'Supporting distributed and ad-hoc team interaction'. Proceedings of IEE People in Control: International Conference on *Human Interfaces in Control Rooms, Cockpits and Command Centres*, Conference Publication No. 463, University of Bath, June 21–23, 1999, pp. 64–71

15 ROGOVIN, M., FRAMPTON, G. T., and the Nuclear Regulatory Commission (NRC) Special Inquiry Group: 'Three Mile Island: a report to the commission and the public'. NUREG/CR-1250, Vol. 1, US Nuclear Regulatory Commission, Washington, DC, 1980

16 SALAS, E., DICKINSON, T., CONVERSE, S.A., and TANNENBAUM, S.I.: 'Toward an understanding of team performance and training', *in* SWEZEY, R. W., and SALAS, E. (Eds.): 'Teams, their training and performance' (Ablex, Norwood, NJ, 1992)

17 SWEZEY, R. W., and SALAS, E.: 'Guidelines for use in team-training development', *in* SWEZEY, R. W., and SALAS, E. (Eds.): 'Teams, their training and performance' (Ablex, Norwood, NJ, 1992)

18 TUCKMAN, B. W.: 'Developmental sequence in small groups', *Psychological Bulletin*, 1965, **63**, pp. 384–99

Chapter 6
Training for control room tasks

Andrew Shepherd

6.1 Introduction

Control rooms provide places for staff to undertake supervision and control of complex systems. Often, staff are removed from the actual environment and must monitor the system through symbolic displays, rather than direct observation. To acquire the appropriate skills an operator or system supervisor must learn how to co-ordinate actions that can be taken in the control room to deal with different circumstances that arise. Control room design decisions prescribe how the world is represented to operating staff and, therefore, substantially influence what has to be trained.

Control rooms are built where materials are hazardous, such as in the nuclear and petrochemical industries. They are also common where systems are dispersed over a large area and where crucial control parameters, such as temperatures or pressures, can only be viewed accurately through instrumentation. This remote control is also encountered in transportation systems, such as in air traffic control. In some environments the supervisor needs to have access to a view of the real world in order to relate instrumentation readings to direct observation of events. For example, in intensive health care, nurses and doctors need direct access to patients as well as continually monitoring parameters such as heart rate and blood oxygen. Hence, the control room must be adjacent and accessible to the patient's room. Equally, control staff in railway depots need to view the rail-tracks to ensure they are clear as well as monitoring displays showing the status of signals and points before agreeing to the movement of trains. In these cases, the control room designer must provide a suitable view, either through strategically placed windows or closed-circuit television cameras. The characteristics of the system being controlled directly affect the decision to build a control room. They also substantially influence the nature of the tasks that control staff must carry out.

Control room designers may choose to represent information to supervisory staff in a variety of ways, through numerical or pictorial displays, for example. Hence, the designer may provide parameters that supervisors in a particular domain have not encountered. For example, chemical plant operators may have access to on-line analysis rather than having to take routine samples. In light railway systems, staff may be able to monitor the weight of trains in order to estimate how many people are on board. Thus, when making decisions about whether to introduce extra trains, they must learn to use this new information in conjunction with the video screen observations about crowding on platforms that they already make. The designer may also choose to integrate certain parameters to represent the system in novel ways, or even automate parts of the control of the system. The resultant design presents operating staff with a set of operating goals, process information, control facilities and an operating context which, taken together, constitute the task that must be trained.

The characteristics of the task and its environment also constrain how training can be developed. For example, in controlling hazardous systems such as nuclear power plants, the consequences of error can be severe, thereby justifying costly or extensive training provision. But in some environments operating problems may be dealt with easily and safely, provided that a competent on-job instructor is on hand – this is how most people learn to drive after all. In some operational contexts events may be predictable, therefore using job aids may be a suitable method of operating. In other situations it may be impossible to anticipate all of the events that control staff are expected to deal with and so staff must be trained. If the consequences of making incorrect decisions are unacceptable and cannot be recovered easily, such as in air-traffic control, then extensive training provision, including simulation, is justified. Like other areas of provision, training investment must be justified in terms of risks and benefits.

Since training decisions are always affected by design decisions, design decisions should be cognisant of training implications. If thought is given to tasks and how personnel are trained, then some aspects of control room design can be modified to make training more effective and system integration far more successful. Developing training is, generally, the province of training departments. However, training staff must take decisions in full consultation with staff engaged in specifying performance requirements and providing resources. These include operations management, safety specialists, designers and commissioning management. Design decisions can also influence the manner in which training can be developed. By acknowledging the link to training, the designer can make it easier and less expensive to provide effective training. By ignoring the link, the designer can severely constrain provision of training.

6.2 Relating training to the task

6.2.1 Task analysis

The characteristics of different tasks mean that different approaches to training are justified in different contexts. Task analysis methods are used in order to understand tasks for the purpose of designing training. Task analysis helps to set out operating goals and indicates which actions and decisions are necessary for accomplishing these goals. It indicates those parts of the task that must be carried out in a rigourous way in order to maintain standards of safety and indicates other parts where the operator is required to demonstrate expertise and discretion. It also helps clarify the context in which work is undertaken. This can help the training designer decide which different training methods are appropriate.

An important method of task analysis is Hierarchical Task Analysis (HTA) (Duncan, 1974; Shepherd, 1989; 2001). In HTA, a task is first expressed in terms of a goal that the operator is required to attain, then redescribed into sub-goals and a plan that governs when each sub-goal should be carried out. Each sub-goal is then considered in turn to decide whether suggestions can be made to ensure that performance will be satisfactory or whether further detail is warranted. The result of this process is illustrated in Figure 6.1. This shows part of a typical process control operation. The plan is crucial since it specifies parameters to be observed to guide the next action – time and temperature in this case. Even in such a simple example, it becomes clear that the location of the temperature display, the manner in which temperature is displayed and the extent to which control is automatic, all influence the skill that the operator must master. As well as supporting training design, HTA can help designers express how tasks must be carried out in order that they can make other design decisions to support human performance.

Figure 6.2 shows a more extended example of HTA – part of the analysis of a continuous process operation. The top level of this analysis is typical of the sorts of ways in which continuous processes, indeed many system supervision processes, can be described. There is a phase of starting up the system, of running the system

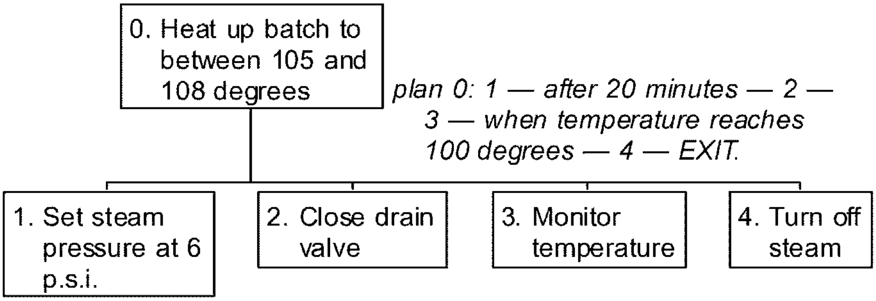


Figure 6.1. HTA of a simple heating operation.

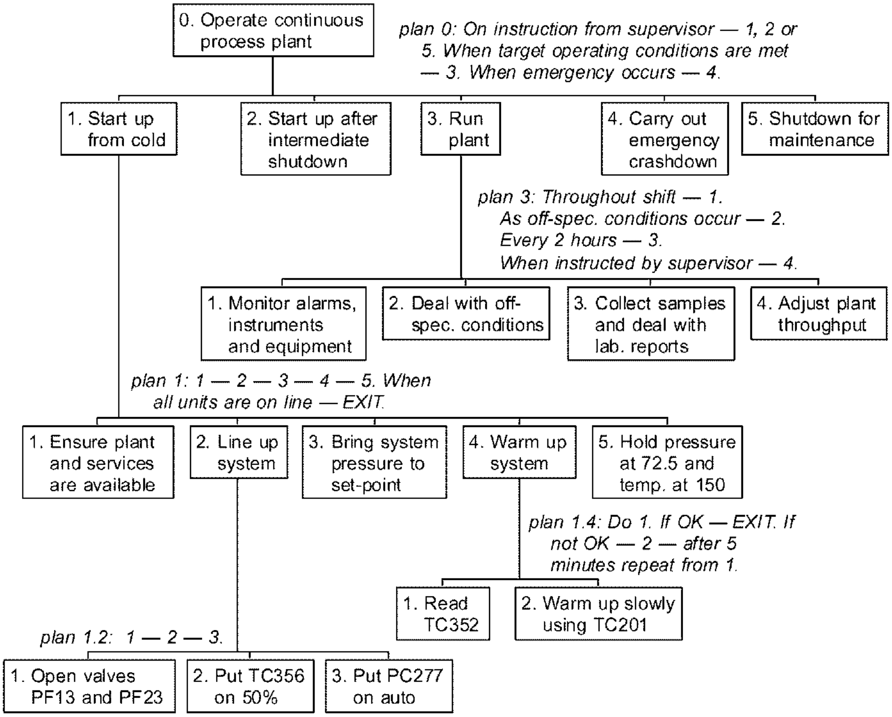


Figure 6.2. HTA of a continuous process control task.

and of shutting it down again. Starting a plant up from cold and empty following routine maintenance, for example, is a lengthier but generally more predictable process than re-starting the system after it has been temporarily shut down following, for example, some localised maintenance. The procedures for starting-up from cold and empty can be anticipated, written into operating instructions and rehearsed. Starting up the plant from an intermediate state will need more understanding on the part of the operator, because certain steps may need to be varied according to prevailing conditions. Therefore, these two aspects of the task will have different training implications.

Running the plant entails ensuring that the process plant is working according to design. Monitoring system performance may entail paying attention to alarms when they occur. It is also generally regarded that the good operator is alert to trends and aware of where to pay particular attention so that s/he can respond more quickly and more appropriately to deviations when they arise. Responding to alarms entails listening and seeing, whereas attention to trends requires greater interpretation and understanding of a range of factors.

Different systems require this sort of task to be accomplished in various ways. In some cases people are employed as machine watchers, who register an alarm as quickly as possible, then respond with a fixed procedure or by referring the problem to a colleague who then assumes responsibility. In other cases, the person responsible for system monitoring will also be responsible for dealing with the problems that have arisen. Each of these variants implies a different training requirement.

Dealing with off-specification conditions is done by a combination of tasks, including diagnosis, compensating for the problem, rectification of the problem, then recovering the situation to return to the desired operating conditions – Figure 6.3 shows how operation 3.2 from Figure 6.2, dealing with 'off-spec.' conditions, can be represented.

This general pattern of start-up, run and shutdown is characteristic of many system supervisory tasks. Other examples can be found in the task of supervising an automated railway system. Thus, the railway supervisory task reflects phases of starting the system and running it. Equally, the general pattern of dealing with off-specification conditions, shown in Figure 6.3, is quite common. Supervising the running of the railway involves monitoring performance to ensure that trains run to time and dealing with any disturbances. Minor disturbances can be dealt with by minor adjustments to the speed of trains; more serious disturbances need to be investigated and dealt with appropriately, possibly by bringing in or removing trains from service or rerouting trains and dealing with hazards on the line. Shepherd (2001) illustrates and discusses HTA for a wide range of tasks from different domains.

6.2.2 Task analysis in human factors design

As well as being crucial in designing and developing training, task analysis makes an important contribution to each of the major stages of human factors design, including: the allocation of function between humans and equipment, information requirements specification, team and job design, and interface design (Wilson and Rajan, 1995). It is also a prerequisite to making judgements of system reliability or interface usability. Without using task analysis methods to aid these design decisions, the control room designer cannot properly understand what functions the control room environment must serve (see Kirwan and Ainsworth, 1992).

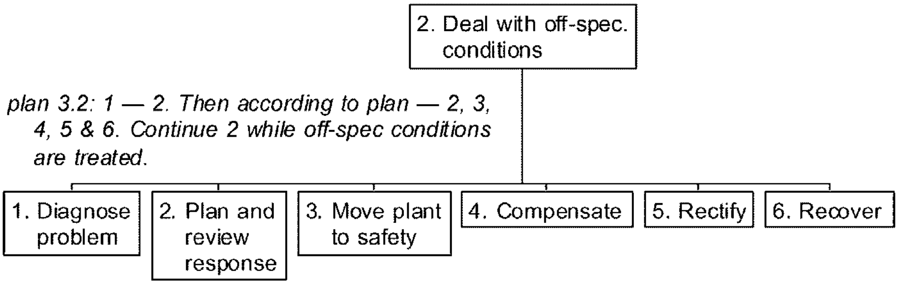


Figure 6.3. HTA of dealing with off-specification conditions.

Both training design and control room design require a proper understanding of the task. When task analysis is carried out only later in the design process in order to develop training it often highlights fundamental design weaknesses that cannot be overcome by operating skill. For example, it may become apparent that information necessary to make decisions is not readily available to the operating staff. This can be due to inadequate instrumentation, failure to provide crucial views of outside events or poor sight-lines in the control room. Where information is inadequate it can hamper or prevent development of the necessary operating skill. Task analysis can also highlight sub-tasks that would have proven unnecessary had they been anticipated earlier in the design process. Recognising the need to clean out parts of the system or the frequency with which aspects of the system need inspecting can prompt fundamental changes in basic design, either by simplifying tasks or eliminating the need for them all together. It must be accepted that training cannot always overcome poor task design. By recognising the relevance of task analysis to control room design, task analysis can be commenced earlier in the design process with both objectives in mind. This is economical, leads to greater consistency between different aspects of system design and encourages trainers and designers to work together to the benefit of overall system performance.

6.3 Setting the criteria for performance

The extent to which achievement during training leads to success at the job is called ‘transfer of training’ (see Patrick, 1992 for a fuller discussion of transfer and other training concepts). In some situations it is necessary to ensure that the operator is fully competent to deal with particular circumstances on the first occasion that they occur. For example, the airline pilot must land successfully first time. This contrasts with situations where a lower initial standard is permitted because the operator is able to learn from experience. Thus, the benefit of a computer-based training programme could be demonstrated by showing that operators who followed it became reliable sooner than those who did not. These two types of transfer measure can be referred to, respectively, as ‘first shot’ and ‘savings’ measures.

In principle, transfer of training is measured formally in order to demonstrate the benefits of training. In practice, this is often impossible because formally measuring how well people perform in the real task may be too difficult to achieve. This is especially the case in tasks where critical events are unscheduled and infrequent, making it impossible to observe different people in similar circumstances. Sometimes, we may be able to judge training effectiveness by seeing how people perform in simulated tasks, because a representative set of situations can more easily be provided. But this is often limited because some people perform well in simulators but are stressed in the real environment and vice versa. So, in order to judge whether people will operate effectively in a real task, performance in a simulator should be complemented by judgements about the trainee’s knowledge of systems and operations and their confidence in dealing with problems under pressure.

While we are limited in formally assessing transfer of training, the concept is useful in helping to make training decisions. In training hazardous tasks, for example, we are likely to be interested in first-shot performance. In training evacuation drills, we are concerned that people evacuate safely on the first occasion that it is required of them, so it may be necessary to invest in simulation and provide frequent emergency exercises. In some operations, however, it is possible to permit mistakes, provided they are quickly identified and safe conditions quickly recovered. This means that some aspect of learning through properly supported on-the-job experience can be used with an experienced colleague anticipating and dealing with problems. This implies savings in the transfer of training because the aim of training is to reduce the length of time that the experienced colleague must remain on hand. The choice must be made on the merits of the situation.

Evidence that training is effective must be sought from an assessment of how well the operator carries out his or her task. In a highly routine task, an assembly task for example, we would have some evidence of whether or not training and support is working by judging the quality of finished work – although this would not indicate whether work was carried out safely. In most process control tasks events are far less straightforward. Often we must rely on supervisory judgements to decide whether or not a person is performing satisfactorily. Such judgements include:

- the extent to which the operator attempts to do the right thing;
- the extent to which the activity is completed or targets are met in an acceptable time;
- whether the pace and rhythm of carrying out the activity is satisfactory;
- how well the activity integrates with other activities;
- whether the activity yields an accurate outcome;
- whether an acceptable method is used;
- whether staff are conscious of the risks they are managing;
- whether performance is achieved with confidence;
- whether performance is maintained under conditions of stress.

The judgements used will need to vary for different tasks and work domains; these need to be determined through task analysis. In some situations it may be necessary to comply with a specified procedure to minimise risks, while in other cases there may be less concern about how things are done as long as they are carried out successfully or quickly. Railway supervision, the management of aircraft movements in air traffic control, the recovery from a pump failure in a process plant are all tasks that need to be done quickly to return the system to profitability, but not at the risk of safety. In intensive health care, speed of diagnosis may be paramount to avoid death, but the speed of diagnosis may result in error anyway. In many process plants, speed in diagnosis avoids down-time and hence can avoid financial penalties, whereas the consequences of error in diagnosis may be catastrophic.

A characteristic of fault diagnosis tasks is that they are required infrequently and irregularly. An operator who is observed doing a task competently may simply be demonstrating greater familiarity with the faults that arise frequently. Unfortunately, similar symptom patterns may arise with less frequent faults and this can lead to error. In these circumstances, it is important to establish that the operator is following a systematic strategy that more carefully explores the differences between situations. It may be necessary to use some form of task simulation to make these judgements. It is usually necessary to obtain different sorts of information about performance, including supervisory ratings, product measures of performance showing the effectiveness of decisions taken and process measures showing that satisfactory adaptive strategies are being followed (e.g. Duncan, 1987). Wilson and Corlett (1995) contains a wide selection of papers detailing different sorts of measurement that can be undertaken in assessing performance.

6.3.1 Performance measurement in system design and development

While measuring operator performance is important for trainers, it is also important for system designers because measures of human performance must be used to assess the complete design from the perspective of usability and effectiveness. Such evaluation can lead to a design review where some tasks become better supported, simplified, automated, eliminated, or distributed between team members in ways that had not previously been envisaged.

Knowing how performance will be assessed can also influence facilities provided in the control room. For example, it might be helpful to provide training desks, where trainees can sit under the supervision of their instructor. The instructor may need remote access to monitor the trainee's choice of action and may need these to be recorded, either to make judgements about progress, or talk through different problems or strategies with the trainee on a later occasion. It may be important to ensure that instructors observing the trainee have proper access to the views necessary to make a judgement without getting in the way of other control room activities. It may be useful to provide access to scenarios that can be used during practice periods when other activities allow training or assessment to take place.

6.4 Providing training

6.4.1 Aspects of skill and skill acquisition

Since experimental tasks rarely represent the complexities of real tasks, much of the research that underpins our knowledge of skill acquisition must be treated with caution. However, theoretical accounts of skill are useful in order to provide an insight into the processes involved in skill acquisition and to help us to relate the processes of learning to training methods.

6.4.1.1 Information and skill

A basic representation of the relationship between an operator and his or her working environment is shown in Figure 6.4. The operator must monitor input information to discern system states in need of change, then make planning and control decisions to determine a course of action to move the system towards the required goal state. In order to control action, the operator needs to monitor action feedback. In a process control example, the operator's goal might be to maintain a neutral level of acidity. On a particular occasion, the operator may decide that this requires caustic flow to be increased by five per cent. To increase caustic flow by five per cent the operator must be able to see the controlling instrument and ensure that it is adjusted in the appropriate way to obtain the required effect. When the action has been carried out, the operator must monitor control feedback to be satisfied that this has resulted in the required change to the system. Then, the operator needs to monitor system feedback to establish that the desired outcome – a neutral level of acidity – has been achieved. This usually means waiting for a period for the effect to start to take place. If the change proves unsatisfactory, then further planning and control decisions must be made to enable correction. If all of these elements are integrated successfully, then the operator can be assumed to have demonstrated 'skill'. Planning and control skill involves a number of cognitive elements. Using the information from the senses entails processes of perception.

While most people may see control room displays as banks of instruments and display screens, experienced controllers organise information much more

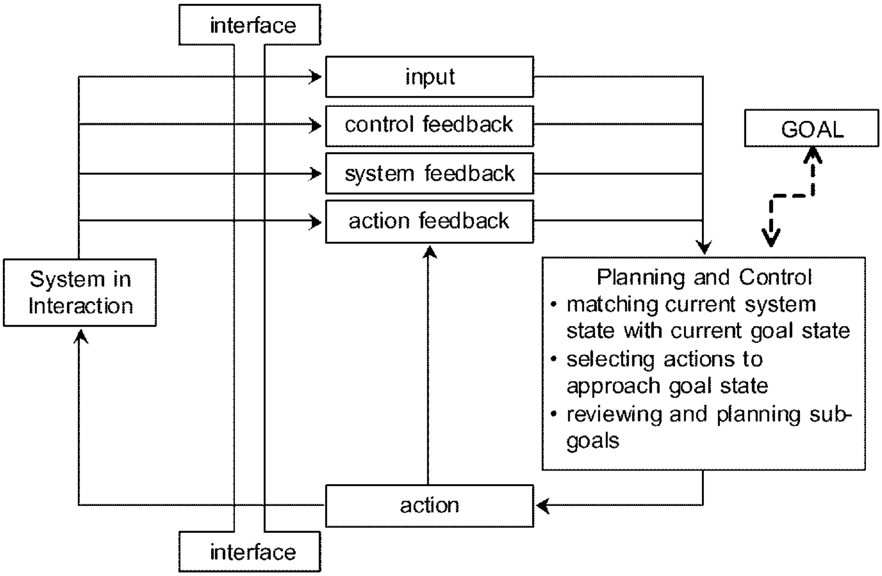


Figure 6.4. A basic informational model of skill.

efficiently in terms of system functions. The operator uses this perception of the system to evaluate whether intervention is required, then takes decisions concerning how to deal with any perturbation or other need for change. Typical task elements involved in dealing with system perturbations include the cognitive processes – perception, decision-making, action-planning and action-control. These are fundamental cognitive skills that people need to acquire to become skilled at system supervisory tasks (see Wickens, 1992 for a more extensive review). For a fuller account of the research underpinning skill acquisition, see Annett (1991).

6.4.1.2 Knowledge, skill practice and feedback

Operating skill develops when the trainee is given the opportunity to practise parts of the task; s/he develops new areas of competence by applying knowledge and existing skills to deal with new problems (see Stammers, 1996). Such practice could occur in the real situation as an operator gains experience, but this would be limited to the events that arise during operations and cannot benefit from the same instructional support. The same processes occur during training as the operator is given new tasks to master. Knowledge helps the trainee fashion appropriate responses to situations encountered. The trainee evolves strategies to distinguish between situations warranting different treatment. During practice, knowledge may serve to guide action, make choices (or informed guesses), suggest analogies and indicate the relevance of appropriate task feedback. When outcomes are shown to be successful through appropriate feedback, learning may arise. The practical implication of this is that knowledge and practice must both be provided in training. The training context is explicitly designed to help the trainee learn, whereas the operating context is not.

6.4.1.3 Stages in learning

An established notion within skill acquisition is that practical skills develop through a series of stages (Fitts and Posner, 1967). In motor skill development the initial stage has been called the ‘cognitive’ phase as the learner attempts to generate an appropriate set of actions, based on an application of knowledge or through mimicking the actions of an instructor. This is followed by the ‘associative’ phase where the correct response patterns are laid down by practice. Finally, the ‘automation’ phase enables response patterns to become automated. Control room skills are essentially cognitive and they develop in a similar way as the trainee moves from conscious reasoning about processes to more automatic judgements (Anderson, 1990). With practice at cognitive skills, operators use information far more proficiently and economically. People become expert because they have encountered more situations and so can deal more quickly with different circumstances as they arise, either by recognising previous instances, by developing a flexible strategy, or bits of both approaches (Shepherd and Kontogiannis, 1998). Practice is essential to enable the trainee to see how best to adapt this knowledge so decisions can be made more efficiently.

6.4.1.4 Individual differences

People differ in the knowledge they possess at the start of training, the manner in which they are able to use their knowledge, the ways in which they are best able to learn new materials, the speed with which they can assimilate new materials and their confidence in dealing with new materials (see Patrick, 1992). These individual differences mean that training methods should be flexible and adaptive to individuals where trainees progress at a rate that suits their individual needs. Some people may be slower to learn than others, but still manage to become fully skilled. The main aim of industrial training is to ensure that all personnel attain a satisfactory level of competence. If one trainee takes more time to achieve a required and reliable level of competence than a colleague, then this may be of no consequence beyond the marginal differences in the cost of training each of them. The extra resources needed to provide adaptive training will be off-set by avoiding the wastage caused by people failing because the pace of training was not appropriate to their needs.

6.4.1.5 The role of extrinsic information in instruction

The various forms of information identified in Figure 6.4 are intrinsic to the task. This means that the operator can make use of this information in the real task. In training, it is often helpful to provide additional input and feedback to the trainee, i.e. extrinsic information. This could include information to prompt the trainee what to do next, help the trainee deal with difficulties and provide feedback to tell the trainee how s/he is progressing and how to rectify errors. Providing extrinsic information is generally necessary to help the trainee start to perform difficult tasks effectively. As training progresses, sources of extrinsic information must be removed to ensure that the trainee learns to cope with only those features that will be present during real operations. Careful management of extrinsic information is essential to ensure that the trainee does not become reliant on features that are not present in the real task.

6.4.2 The instructional cycle in mastering operational skills

Providing training and instruction entails a number of complementary activities in which skills can develop in the manner just discussed. Figure 6.5 shows a typical cycle of instruction. The instructor identifies an appropriate opportunity to practise and prepares materials pertinent to the session. When the trainee is ready, s/he is set to practise the training task under scrutiny of the instructor who provides support and maintains the well-being of the trainee. If this training is 'on-the-job', the instructor must also maintain the safety and security of the plant or system. The instructor may choose to intervene or to leave the trainee to deal with problems encountered without assistance. When the practice session is complete, the instructor can provide feedback, either verbally or by reference to records or a recording of what has transpired. In this way, weaknesses can be ironed out and encouragement given. After the session, the instructor must update records of

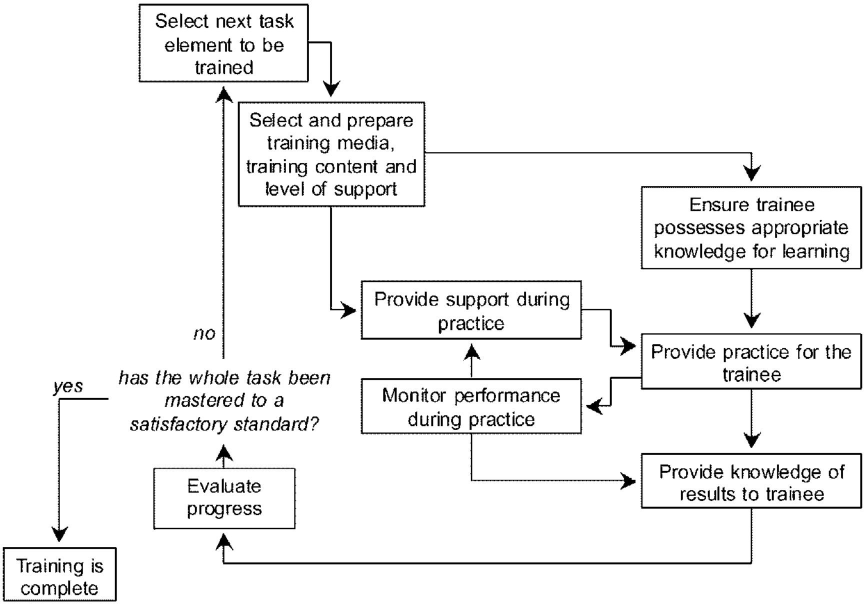


Figure 6.5. The instructional cycle.

achievement in order to provide a basis for progression to the next training session. At the next session, the instructor may choose to consolidate skills already mastered or move to fresh material. In moving through the training programme, therefore, the instructor must manage the support of the different phases of learning according to how the trainee is progressing. A fuller account of these processes is given in Shepherd (2001).

6.4.3 Common forms of training

In practice, training is delivered by one of a handful of practical methods including classroom instruction, on-the-job instruction, simulator training, computer-based learning and supervised experience. Often, training choices are dictated by what circumstances and resources permit. The various training methods can all be useful, but they may often be compromised because of costs or lack of time to make adequate preparation.

Classroom instruction is often used to teach knowledge concerning the system being controlled. It may include descriptions of the structure and function of systems, elementary physics or chemistry, and business objectives. It can be used to motivate people to learn by showing them how their role fits into a wider picture. It can teach the names of equipment and parts of the system, justifying certain procedures and safety measures, and the rules and principles of operation.

However, this form of training is often over-used because it is relatively easy to administer and develop. A common weakness of classroom training is the tendency to lump all such teaching together and to teach all trainees together. This is administratively convenient but some trainees will be unable to keep up and many trainees will be overloaded with 'knowledge' and not understand which elements are relevant when they come to practise. Knowledge should be presented selectively in order to support the acquisition of different skills. A more satisfactory approach is to use task analysis to identify knowledge requirements to present alongside practice opportunities. It may be less convenient to administer, but with techniques such as computer-based training, where knowledge can be presented on an individual basis, administrative difficulties can often be overcome.

Computer-based learning is becoming more commonplace. It offers some obvious advantages in enabling instruction to be delivered on an individual basis as and when needed. Computer-based learning can be used to provide knowledge training and can also provide opportunities to practise different parts of the task using part-task simulations. Thus, knowledge teaching can be linked directly to relevant parts of the task. A particular advantage is that such training can be adaptive to the needs of the individual learner.

On-the-job instruction is where a trainee is introduced to tasks in the real system under the scrutiny and control of an experienced colleague. This provides an opportunity for the trainee to see the work properly paced, to see the actual signals and to gain a proper feel of the controls. It helps new members of staff integrate with their colleagues. It also avoids the necessity to build specialist training environments. However, there are negative aspects which are particularly pertinent to control room work. One problem is that, since many control rooms are concerned with hazardous processes, the consequences of action by unskilled people is a risk that cannot be taken. A second implication of such situations is that crucial events occur infrequently and intermittently, offering little or no opportunity for practice on the job. For on-the-job instruction to be most effective, the instructor must be trained, should have an instruction programme to guide what needs to be trained and know the standards to be attained. In essence, the instructor should comply with the processes represented in Figure 6.5 and relate knowledge to practice in an effective way. But these training procedures will still not overcome the inherent problem of dealing with the infrequent and irregular events that are so crucial in many control room tasks.

Simulation training entails using some kind of representation of the task to be learned in order to provide a trainee with the opportunity to practise. It is the task rather than the equipment that should be simulated. Simulators need not be close physical representations of control consoles and control panels. High fidelity may be required for some purposes, but it is only through analysing the task that precise simulation requirements can be established. Training simulators must be designed not merely to simulate, but also to train. The issue of simulator training will be discussed more fully in Section 6.5.

Often, the ideal learning environment to support skill is a mixture of classroom learning, on-the-job instruction, computer-based learning, simulator training and learning through experience. If a control room designer were to engage with a training designer, then far more satisfactory learning arrangements could be provided. Such provision could include training consoles where trainees and their instructors could work more effectively, video and software facilities to monitor control actions and decisions, computer-based learning facilities made available to enable revision of various principles or practice at various task elements relevant to new skills being acquired. Systems would include training modes where trainees could experience the task through dummy data, including video footage of relevant views of the outside world or an audio library to convey the sorts of message with which trainees must learn to cope. This sort of collaboration between training and other design staff will result in greater system effectiveness.

6.5 The role of simulation in training for control room tasks

Simulation for training has a central role with regard to control room tasks. Control room tasks are often associated with hazardous processes and systems. These tasks also entail supervision of systems where some critical events occur infrequently and irregularly, requiring monitoring and decision-making skills to be practised. Operating risk and infrequency of events are also the conditions that tend to justify adopting task simulation. Therefore, simulation is often justified in training control room tasks.

Creating satisfactory simulations of real environments is rarely straightforward. A principle reason is cost. Replicating the physical attributes of a real environment can be expensive in itself. Representing the events with which the trainee must cope can be particularly problematic. Because of costs, simulation design often entails compromises. These compromises often result in reduced fidelity, in which the skills practised may no longer transfer satisfactorily to the real situation. The training simulation designer must make careful judgements about which compromises can be managed effectively to minimise the risk of negative transfer of skill as the trainee moves to work in the real system.

6.5.1 Representing the task

Simulation seeks to represent enough of a task to enable the trainee to practise the skills that have to be acquired. We can gain a clearer picture of the problems facing simulation design from Figure 6.6 (a variant of Figure 6.4), which indicates how different aspects of operating can influence each other. In the real situation, the operating team receives information about the system under control through displays and effects changes through controls. Information can emanate from control panels and visual display terminals. It can also come from communications such as telephones, tannoys, e-mail or direct verbal interactions from colleagues. It can also come from looking at, listening to, smelling or sensing the environment through vibration or other physical sensations. For example, the aircraft pilot

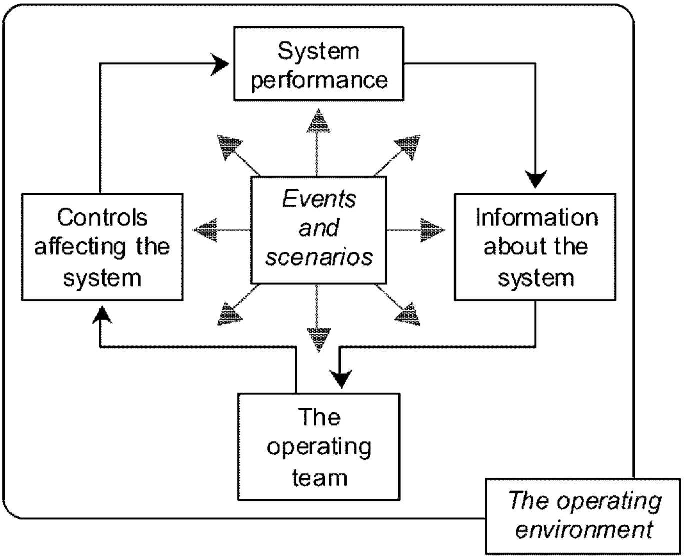


Figure 6.6. *Elements in an operating environment.*

relates information from instruments, from the cockpit view and from physical sensations of movement. Controls that effect changes to the system – knobs, switches, keyboards, touch-screens, telephones – must be mastered. Events affect the sorts of information conveyed and the frequency with which different patterns of information are conveyed. Infrequent events will, by definition, mean that fewer opportunities for practice are represented. Events and scenarios can affect the reliability of control systems. They also affect the expectations of operating teams concerning what is likely to have happened in different circumstance. Also, the operating environment itself can affect system performance – heat and vibration can affect the reliability of equipment and it can also affect the performance of the human operator (Hartley *et al.*, 1989).

A problem in any simulation design is knowing which aspects to include and which aspects can be left out. This is especially of concern when the cost of developing the simulator is addressed. Effective skilled performance does not make use of all aspects of a working environment, but what are the risks of leaving out expensive features? Task analysis should be used to help clarify which features are essential. Particular care should be taken to ensure that the trainee does not come to rely on features of the simulator that are not represented in the same way as in the real task.

6.5.2 Representing the system being controlled

Simulation design does not simply entail replicating the physical appearance of the real task. The performance of the system also substantially affects how the trainee experiences events. If the simulated system does not respond to control changes in the same way as the real system, then skills mastered may be inappropriate. This is easy to see when changing from one motor car to another. It is easy to adapt in a short space of time, but initially, the different characteristics of the two vehicles will affect the smoothness of the ride – the driver might brake too hard or accelerate too quickly and this could lead to an accident. The same influences arise in simulation. If the trainee has mastered a set of characteristics in a simulator that did not properly represent the real system, then negative transfer can result.

Another aspect of representing system performance is providing events according to their frequency in the real world. This is not the same sort of technical problem as creating appropriate system dynamics; it simply amounts to deciding when to schedule events with which the trainee is required to gain experience. This presents a dilemma for the person administering the training simulator, since by increasing the frequency of occurrence of rare events in order to allow more practice the trainee now starts to expect problems at a higher frequency than in the real world. For example, a civil airline pilot undertaking a simulator exercise could encounter an engine failure, a terrorist attack and a colleague suffering a heart attack all within the space of two hours, yet never experience any of these in reality. There is no simple solution to this training problem. It is often advisable to ensure that trainees master sensible strategies for dealing with problems before they undertake extensive simulation training, otherwise, there is a risk that they will come to rely on selecting responses simply according to expectation. To reduce this risk, simulation must be used carefully, within a well-thought-out training programme. It is important that trainees master some of the difficult events with which they may eventually be confronted. But since, by definition, they may also encounter situations they have not seen before, it is important that emphasis is placed on helping them to develop more flexible strategies.

6.5.3 Providing simulation in practice

The purpose of simulation in training is to provide opportunities for practice. This does not necessarily depend solely on building a dedicated simulator as actual equipment or approximations of real equipment can also sometimes be utilised for this purpose.

Using real equipment may entail deliberately inducing fault conditions or may involve engaging various other personnel to convey information about problems to the trainee. Such approaches are often used for safety drills. Using this approach, the trainee will experience events in a similar manner to how they would occur in reality. This general approach is only feasible where safety consequences can be controlled. Real equipment can also be used where simulated data is available to drive displays in a training or practice mode. In a control room, with several similar work-stations, it may be possible, to assign one work-station for training purposes. This is an effective way to train new personnel and provide practice drills to maintain the skills of all staff, because it allows them to deal with critical situations they are unlikely to encounter during normal operations. Providing such facilities should be a principle concern for control room design.

In some circumstances, it is also possible to simulate a working environment using low fidelity elements. Some team aspects of control room tasks may require people to become familiar with control room layout and to communicate with colleagues, other people in the control room environment and emergency services. In such situations it can be appropriate to simulate the control room by providing equivalent space, a warehouse, for example, with control desks represented by tables and equipment represented by cardboard boxes. This approach is entirely legitimate provided the activities with which the trainee is engaged correspond to the activities they aim to represent. Such approaches are also extremely useful in assessing control room layout early in the design and development cycle.

Another sort of simulation entails focusing on specific parts of the task only. These can be called 'part-task simulators'. Careful task analysis indicates where task elements and their associated skills may be practised separately to good effect. This can be a very powerful solution. In some decision-making tasks, where the onus is on the operator to collect information rather than carry out controlling actions, it is only necessary to represent the signals associated with events and not to replicate the response features or the full dynamics of the system. These skills may then be further practised in a full simulation, or applied and refined in the real situation. On its own, the part-task trainer may be insufficient to ensure an adequate level of skill. However, used in conjunction with other methods, part-task training simulation can be very effective because it can be used to help trainees build up more versatile strategies. Part-task training simulation in fault diagnosis, for example, provides the trainee with opportunities to focus on critical aspects of performance in circumstances where a range of events can be presented. These skills can then be transferred to a less supportive operational or simulated context where the trainee can learn how to apply the strategies learned in the full-task context.

Simulator training is rarely an efficient method of taking a trainee from a low level to a high level of skill. The simulator may be too complex an environment to enable the novice to learn a task from the beginning. Moreover, full-mission simulators are often expensive and therefore few are available. A more sensible solution is to use the full-mission simulator only towards the end of a training programme, utilising other training methods to build up skills. Part-task trainers enable the trainee to accomplish the earlier stages of learning more efficiently so that expensive simulator training may then be used more effectively to consolidate skills. Personal computers are relatively inexpensive and are cost effective because they are useful for other purposes, so they can be provided in greater numbers. Individual time on the full simulator can be scheduled while other trainees continue with computer-based part-task training exercises.

Tasks for which controls rooms are built often exhibit the characteristics that justify using training simulation. They can entail risk, which means that errors by people undergoing training must be minimised; they involve system supervision, which means that the events with which they must learn to cope occur infrequently and irregularly and so offer insufficient opportunity for practice. Given that simulation training is often required for control room tasks, it is perhaps fortunate that several simulation problems are easy to solve. Control rooms are artificial environments, so simulating the physical aspects of the control room should be straightforward. The difficulty rests with representing the dynamics of the system being controlled.

To provide simulation for training, it would be necessary to address the issues of freezing, rerunning and controlling the presentation of events. It would also be important to provide suitable tools by which a trainer can monitor and record trainee actions. This could involve software to record control actions, trends and screen shots, and video facilities to monitor the trainee's movements in the simulated control room. But all of these facilities would be of use to engineers evaluating system performance anyway. It would be helpful if designers of these simulations could collaborate with training staff in order to adapt facilities to serve both purposes. Indeed, collaboration could be helpful to both parties, because the costs of simulation can be spread across these two aspects of system development.

6.6 Conclusions

Many managers and engineers regard the development and administration of training as a company function separate from the development of methods to support operators, but control room design and training are two interrelated facets of system design and development and so should be linked together. Control room design entails the fashioning of a built environment in which people are required to undertake tasks. This entails providing facilities to display information and to enable system control, and an environment in which people can work together. Acquiring skills involves learning how to relate system information to control actions in order to meet system goals. Therefore, skills are partially determined by the environment that the control room design provides. It is essential that the human factors requirements of control room tasks are understood in order to provide tasks that people can carry out. If this is not done, then the benefits of training will be severely limited. It is beneficial if design personnel can collaborate with training staff in the proper analysis of tasks so that tasks can be well designed, training is properly informed and that both aspects of design complement each other in meeting systems' requirements.

It is also beneficial if control room designers can appreciate the facilities that would serve the best interests of training, because this can ensure the most effective mastery of skills. Indeed, many of the facilities that trainers seek to enhance training could also be put to good effect in helping control room designers evaluate their designs and ensure that control rooms will be used most effectively.

References

1 ANDERSON, J. R.: 'Cognitive psychology and its implications' (Freeman, New York, 1990, 3rd edn.)

2 ANNETT, J.: 'Skill acquisition', *in* MORRISON, J. E. (Ed.): 'Training for performance' (John Wiley & Sons, Chichester, 1991)

3 DUNCAN, D. K.: 'Analytical techniques in training design', *in* EDWARDS, E., and LEEDS, F. P. (Eds.): 'The human operator and process control' (Taylor & Francis, London, 1974)

4 DUNCAN, K. D.: 'Reflections on fault diagnosis expertise', *in* RASMUSSEN, J., DUNCAN, K. D., and LEPLAT, J. (Eds.): 'New technology and human error' (John Wiley & Sons, Chichester, 1987)

5 FITTS, P., and POSNER, M. I.: 'Human performance' (Brooks Cole, Belmont, CA, 1967)

6 HARTLEY, L. R., MORRISON, D., and ARNOLD, P.: 'Stress and skill', *in* COLLEY, A. M., and BEECH, J. R. (Eds.): 'Acquisition and performance of cognitive skills' (John Wiley & Sons, Chichester, 1989)

7 KIRWAN, B., and AINSWORTH, L. K.: 'A guide to task analysis' (Taylor & Francis, London, 1992)

8 PATRICK, J.: 'Training: research and practice' (Academic Press, London, 1992)

9 SHEPHERD, A.: 'Analysis and training in information technology tasks', *in* DIAPER, D. (Ed.): 'Task analysis for human-computer interaction' (Ellis Horwood, Chichester, 1989) pp. 15–55

10 SHEPHERD, A.: 'Hierarchical task analysis' (Taylor & Francis, London, 2001)

11 SHEPHERD, A., and KONTOGIANNIS, T.: 'Strategic task performance: a model to facilitate the design of instruction', *International Journal of Cognitive Ergonomics*, 1998, **2**, pp. 349–72

12 STAMMERS, R. B.: 'Training and the acquisition of knowledge and skill', *in* WARR, P. B. (Ed.): 'Psychology at work' (Penguin, Harmondsworth, 1996, 4th edn.)

13 WICKENS, C. D.: 'Engineering psychology and human performance' (HarperCollins, New York, 1992)

14 WILSON, J. R., and CORLETT, E. N. (Eds.): 'Evaluation of human work' (Taylor and Francis, London, 1995)

15 WILSON, J. R., and RAJAN, J. A.: 'Human machine interfaces for systems control', *in* WILSON, J. R., and CORLETT, E. N. (Eds.): 'Evaluation of human work' (Taylor & Francis, London, 1995)

Part Two

Methods

Chapter 7

Humans and machines: allocation of function

Neville Moray

7.1 Introduction

The major classical reviews of the human operator in process control are now over 25 years old (Edwards and Lees, 1973; 1974), and summarise much of what was then already known about the manual control of large complex industrial systems. They contain many hundreds of references to the role of the human operator, but the advance of automation means that the relation of humans and machines in industrial tasks needs to be re-examined.

Automation is present when a function that could be performed by a human operator is performed by a machine that may or may not be a computer. Since automation is used to make systems more efficient than those using manual control, why is the role of human operators of concern? All too frequently accidents and system failures are ascribed to 'human error' rather than hardware faults, and therefore many engineers try to 'design out' human operators, reduce the possibility of human error and increase productivity and safety. Why then retain humans at all? Is their presence merely a transitional stage before complete automation?

Close examination of modern automation suggests that such is not the case, and recently research and industrial practice have begun to emphasise the virtues rather than the weaknesses of operators. It is now agreed that in automated systems humans are needed at least for two purposes: as the last line of defence in hazardous operations and to improve productivity. While the first is generally accepted, the latter may be more surprising, since during automation it is often possible to design a system to perform in a mathematically optimal way. However, it can be shown that humans and machines when combined can sometimes exceed the performance of either, even if machines are fully optimised in a strong mathematical sense. For example, discrete manufacturing scheduling algorithms are optimised over long time horizons, and may not handle efficiently real time modifications for just-in-time operations. Humans can use local windows of opportunity to exceed the performance of schedules optimised over long time horizons (Sanderson, 1989; Singalewitch, 1992; Wiers, 1997; Higgins, 1999). The combination of human and machine is then more powerful than either entity alone.

Even if automation were to replace operators entirely, it would not remove the opportunity for human error: rather, the locus of human error would change. An increased burden would be placed on the designer, who would have to foresee all possible failure modes of the system and design the automation to cope with all of them. However, not all stresses to a plant can be foreseen, and modifications to plant in the years following commissioning may so alter its characteristics that algorithms or interlocks designed into safety systems may become ineffective, and if the conditions under which they are required are very rare, this may go unnoticed. The plant may then contain what Reason (1990; 1997) called 'latent pathogens'. These are plant states 'waiting to happen' and for which the safety systems are no longer available because of the progressive change of systems characteristics. Paradoxically, defence in depth and multiplexed safety systems may make it harder to foresee failures by concealing the existence of such states until a catastrophic failure occurs; at which time only human intervention may be able to devise work around methods to avoid serious consequences.

Maintenance personnel also remain in automated systems and they are often less well qualified than operators. There is good evidence that maintenance-induced faults occur and as hardware and software in advanced automation become ever more 'opaque', the ability of maintenance to detect unsafe states and to correct them becomes ever harder. Self-diagnosing automation again almost guarantees that if ever a failure does occur, it will be both obscure and difficult to diagnose, and complex to manage.

There is, at any stage of development, a limit to what engineers can automate. As Bainbridge (1983) pointed out, with increasing automation the operators' role becomes increasingly difficult, since what is not automated are those functions which are too difficult for engineers to implement. Even when advanced functions are successfully automated, as in aircraft such as the Airbus series, problems arise if the relation between human and machine is not explicitly considered.

If automation has great advantages, but humans must remain in automated systems, what is the role for human operators when a process is automated and how should the role be implemented? How should functions be allocated between human and machine?

7.2 Allocation of function

Between 1950 and about 1980 the contribution of ergonomics and human factors to the design of human–machine systems was dominated by what is sometimes called the MABA–MABA (Men [sic] Are Best At–Machines Are Best At) approach. The theory of design, seen from an ergonomic viewpoint, was that following a task analysis the characteristics of humans and machines should be reviewed to see which could perform which task better. Allocation of function could then be made on a rational basis. The comparative lists of abilities were called Fitts Lists (Fitts, 1951; Hancock and Scallen, 1998).

Even if there was once a time when such lists referred to real differences there is little evidence that they played a major role in human–machine systems design, and they are no longer appropriate, since if sufficient money is available almost any human ability can be at least approximated. Sensing and actions are mimicked by electronic sensors and closed loop controllers; perception is increasingly emulated by artificial pattern recognition and adaptive neural nets; and complex decisions can to some extent be mimicked by expert systems. Today therefore, the preferred approach emphasises co-operation of human and machine, combining human and machine capabilities, and aiming for a true symbiosis. The problem then is to decide how to realise such a combination.

Figure 7.1 shows several levels and modes in which humans and automation might work together (Sheridan, 1987). Recently, Parasuraman *et al.* (2000) have

1. The human does the whole job of planning, selecting options, and preparing the machine, up to the point of switching on the machine to carry out the actions.
2. The human does everything except to ask the computer to help in suggesting possible options.
3. Computer helps to select possible options and suggests to the human one that may be the best.
4. Computer selects a course of action, and it is up to the human to decide whether or not to do it.
5. Computer selects the action and carries it out if the human approves.
6. Computer selects the action and tells the human in plenty of time for the human to stop it if the human disapproves.
7. Computer does the whole job but necessarily tells the human what it has done.
8. Computer does the whole job and tells the human what it has done if it, the computer, decides that the human should know.
9. Computer does the whole job and tells the human only if the human asks what is going on.
10. The computer does everything, including deciding what should be done and deciding whether to do it, and deciding whether or not to tell the human.

Figure 7.1. Sheridan-Verplanck scale of human–machine function allocation in automated systems.

extended this scale, pointing out that it can be applied to data acquisition, to decision-making and to the choice of action separately. For example, high degree of freedom systems with many sensors might use artificial pattern recognition and filtering to provide the human supervisor with a coherent subset of the sensor data, leave diagnosis and decisions to the operator, and exercise closed-loop control to implement the goals chosen by the operator. Thus data acquisition might be performed at Level 8, decision-making at Level 2 and control at Level 9.

7.3 Levels of automation

Surprisingly little research has been done on optimising the level and mode of human-automation co-operation, and indeed there may be no universal 'best' level that is independent of the task. Different systems may require different levels of human–machine co-operation. At low levels, the human has a heavy mental load: at high levels the autonomous system performs massive computation. In the middle of the scale there is genuine co-operation. The information processing abilities and intelligence of both agents are there used to their mutual benefit, but at a price. Each now must communicate and interpret information received from the other in addition to processing information from the system environment. The total workload may then be heavier, not lighter, as a result of automation, due to communication overhead. This can lead to well known problems (see, for example Wiener and Curry, 1980 on 'clumsy' automation, and Bainbridge, 1983). Pilots of highly automated aircraft complain that the automation is useful in long-haul flights (when nothing much happens), but that it causes severe overload during approach and landing. This is due to the need to reprogram the automation rapidly, and the difficulty they have in understanding what the automation is intending at a time when the pilot workload is at its highest.

7.3.1 Mutual understanding and authority

A fundamental problem in human–machine systems is that each agent (human or automation) must understand what the other is doing. Several crashes of highly automated aircraft have occurred because of mode errors when the automation adapted itself to task conditions, but did not signal clearly to the human what mode it was in. If the operator believes that the system is in one mode and it is actually in another, very hazardous conditions can arise. The need for clear feedback is vital, using auditory, kinaesthetic and tactile signals as well as visual feedback, or in highly automated systems operators may come to feel that they are controlling the interface rather than the underlying physical system. It is essential that as automation takes on more and more responsibility the human should still remain in contact with the causal dynamics of the plant.

Figure 7.2 shows an abstract representation of problems of human–machine interaction. If the operator uses sub-system A, the changes in plant state will be

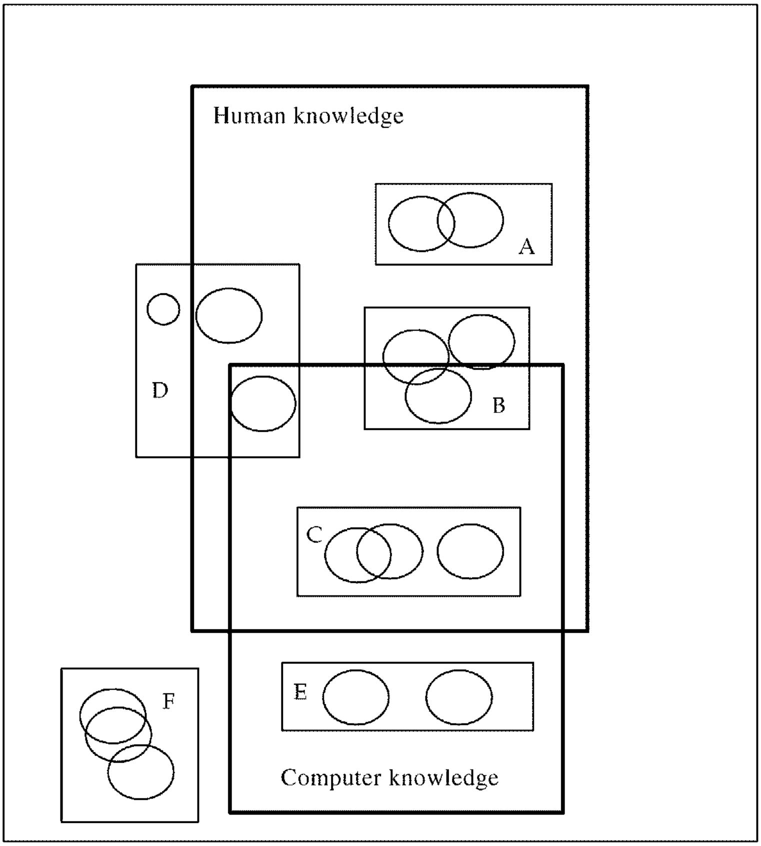


Figure 7.2. *A Venn diagram describing communication in a human–machine symbiosis. The two dark boxes contain all knowledge possessed by the human and the automation computer about the system's properties. The light boxes are quasi-independent subsystems. The ellipses are pieces of equipment such as pumps, valves, etc. Shared knowledge and understanding are indicated by the degree of overlap. The properties of subsystem A are known only to the human; of E only to the computer; of C to both. F is unknown to both. Some aspects of D are understood by both human and computer, some only to the human, and some are not known to either. For interpretation, see text.*

interpreted by the control computer as being disturbances, and it will try to counteract the changes. If the computer uses sub-system E the changes will be perceived by operators as disturbances and they will try to counteract them. Only sub-systems such as C, which are within the knowledge and mutual understanding of both human and control computer will be efficiently used. 'Add-on' equipment such as sub-system F which was not in the original design, and for which there are no standard operating procedures, or for which the control software has not been updated, will cause changes which will be resisted both by operators and by the computer. Further interpretations are left for the reader.

7.3.2 Responsibility

If humans and automation share control, who is ultimately responsible and who has the ultimate authority for action? The explicit definition of role and responsibility is a task for management, which must ensure a safety culture that is also productive. It is unreasonable to expect operators to behave creatively if they believe that the automation is meant to be in charge (Zuboff, 1988). Failure to make explicit this issue may lead to alienation of the workforce and to hesitation by operators to intervene in time-critical safety situations. If they intervene too soon or too often they reduce productivity, but if they fail to intervene when necessary, then too they will be held responsible for accidents. It should be noted, in times of globalisation, that allocation of responsibility and response to different patterns of hierarchical organisation may have very different effects depending on the culture (society, country) in which they are implemented (Moray, 2000). Socio-technical systems must be optimised for culture, not just for engineering criteria, and particularly in high hazard systems.

Currently, it is widely held that operators should have the final authority at all times (Woods, 1988; Woods *et al.*, 1994); but there are conditions in which authority must lie with automation. Some special cases have been analysed by Inagaki (1993; 1995) and Moray *et al.* (2000), and involve particularly time-critical situations where human decision-making time is long compared with the bandwidth of system events. In such cases final initiative and authority must reside in the automation, and there is then an absolute demand on the designer to provide fail-safe automated fault management systems which are more reliable than the process itself.

Operator intervention should usually be minimised when the system is efficient, but is required when the human is able to operate the system better than the automation. It is unlikely that operating procedures can be written to take account of all possible cases. 'Rule-based behaviour' (Rasmussen, 1986; Rasmussen *et al.*, 1995) is only effective as long as the situation is normal and fits the rules which have been developed by the designers. If creative intervention beyond the rules specified by operating procedure is to be possible, care in the design of interfaces is of paramount importance.

7.4 Designing for interaction

Operators must be able at all times to identify the state of the system. Therefore displays should match both the engineering requirements of the system and the cognitive properties of operators, and should encourage the optimal distribution of attention over the interface. Where the capacities of human attention are exceeded, appropriate alarms and warnings must be provided. Strong quantitative models for monitoring behaviour now exist, and in some cases have been validated in laboratory and field studies (Moray and Inagaki, 2001a). Further work in this direction using real industrial systems would be of great value to systems designers, providing a rational model to predict the time to respond to expected and unexpected events (Moray, 1986; 1999a). Operators seldom sample displays at a rate exceeding two per second and usually, particularly when computer displays requiring paging are used, the rate is likely to be far slower. Hence, unless advanced displays (e.g. predictor displays) are implemented, the bandwidth at which humans can monitor and intervene must be limited, and may be exceeded by fast transients.

In recent years the ready availability of powerful computation has given rise to a variety of advanced displays. In addition to predictor displays, which are known to make it far easier for operators to control systems with high order transfer functions, many new approaches have been proposed. For example, it has been suggested that since different operators may have different mental models of a given system, and may also have different tactics for controlling a multi-degree-of-freedom plant, operators should be able to re-configure the displays to suite their individual styles of control. This is an attractive notion, but computer displays that are adaptive, or that can be reconfigured by operators, are known to be potential sources of error and accident and should be implemented only with exceptional care, and only with appropriate training and operating rules.

Some of the problems with adaptive interfaces are due to the fact that most complex systems are operated by teams rather than individuals. In addition to designing for single operators, interfaces should support communication among members of the team, and very little design work has been done on this aspect of human–machine systems. Finally, what we may call the social relationship of operators to their equipment must be optimised. We now consider the role of displays in human–machine relationships.

7.5 The design of displays

Even if operators are supervising automation rather than exercising manual control, their task remains a control task and hence they must be able to identify accurately the system state at all times. The logically primary tasks of a supervisor are monitoring, pattern recognition and diagnosis, and one of the most interesting developments in recent years has been in the domain of display design.

An important distinction is between data and information. Data are simply numbers, but information is a meaningful combination of data that indicates system state with a minimum of mental calculation and manipulation of data. Bennett and Flach (1992) showed how display design can change the balance between data and information in a simple system (see Figure 7.3). They make an important distinction between values that are shown explicitly and values that must be deduced or calculated by the observer during state identification. Problems can arise due to the large number of displayed variables if raw sensor data are displayed. On the other hand, if high order encodings of data are presented as information, the system may become opaque to human cognition when faults occur. Bennett and Flach suggested some interesting approaches to finding displays which provide direct perception of relations among variables rather than leaving their

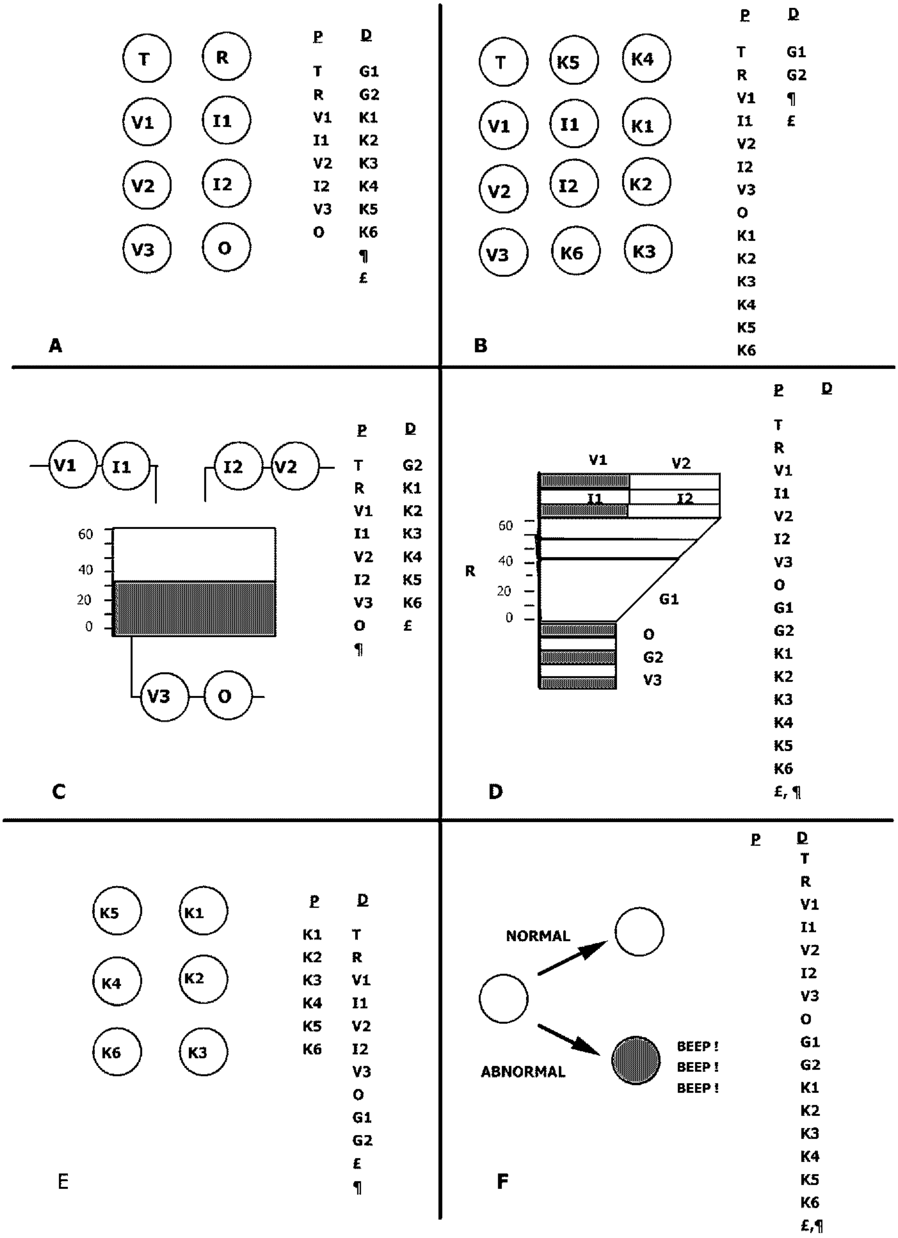


Figure 7.3. *Various methods of encoding data as information. The values of variables in columns labelled P can be directly perceived in the display. Those in the columns labelled D must be derived from the displayed variables. The K(·) are the values of equations describing relations among variables.*

discovery as a task for the operator. They considered how to represent both the values of variables, the relations among them, and the relation of the plant state to overall goals for a simple thermal-hydraulic system whose mimic diagram is shown in Figure 7.3C. The goal is to achieve a defined steady-state outflow rate through valve V3. The setting of each valve is indicated by V(•), and the true flow rate through the valves by I(•).

The most interesting of their displays is Figure 7.3D, where the geometry of the display contains inherent diagnostic properties. The following account of this display is taken from Moray (1999a: 233).

> The form of the display directly presents the values of all the system state variables, inputs and outputs, and the relations among them. The geometry shows the system state in the shapes of the components of the display... (As Simon says) 'solving a problem simply means representing it so as to make the solution transparent.'... Consider how the display supports the perception that the output goal is satisfied. By suitable scaling, the length of the bars O, V3 and G2 can be made equal when the goal G2 is satisfied. Then the vertical straight line bounding the end of O, V3, and G2 is a directly perceptible indicator that the output flow goal is satisfied regardless of quantitative values. Similarly we can scale all the variables and relations to produce a display in which all information is P, and none D. A system running normally is then represented by a vertical rectangle, and any departure from normality by a geometrical deformation of the rectangle... (and) the distortions indicate their cause. In Example D of Figure 7.3 the reservoir is filling (the inflow to the reservoir is greater than the outflow), and the shape of the trapezoid resembles something which is increasing (filling) because of the slope of the line on the right hand end of the trapezoid.... This form of display allows an assessment of system state without (any) calculations, using perception not cognition.

Such a display is similar to what have come to be called 'ecological displays'. Ecological displays show the position of operators in the constraint space that defines the operation of the automated system, rather than only showing the values of variables. Operators can see directly how close they are to the permitted boundaries of operation. The displays show the underlying physics and constraints of the system, and these do not have to be decoded from data. Among recent examples of ecological or near ecological displays are the Rankine cycle display for nuclear power plant steam generation (Beltracchi, 1987; Rasmussen *et al.*, 1995; Vicente *et al.*, 1996) the DURESS display (Vicente, 1999) and a display to show the relation between critical variables in the power output of a reactor (Jones *et al.*, 1992).

Ecological displays are at present experimental and few have been implemented in real systems. But the evidence is strong that they can greatly facilitate the supervisory control. Vicente (1999: 55) summarised the philosophy of ecological interface design (EID) as applied to large-scale systems:

> The approach... is based on a ... perspective which demands that system design begin with, and be driven by, an explicit analysis of the constraints that the work environment imposes on behaviour. Only by adopting such a perspective can designers ensure that the content and structure of the interface is compatible with the constraints that actually govern the work domain ... The (aim) is to ensure that workers will acquire a veridical mental model of the work domain, so that their understanding corresponds, as closely as possible, to the actual behaviour of the system with which they interacting ... The argument for the ecological

> perspective ... generalises meaningfully to any application domain where there is an external reality – outside of the person and the computer – that imposes dynamic, goal-relevant constraints on meaningful behaviour ... workers' mental models should correspond with this external reality. The only way to accomplish this objective efficiently is to adopt an ecological perspective by starting and driving the system design process by conducting an analysis of work constraints.

Even ecological displays require lengthy operator training. Moreover, one should note the importance of cultural differences in the context of the global market place for industrial systems, in the design of displays, operating procedures, and of control room teams and hierarchies. A control panel that is natural and easy to use in one culture may be counterintuitive in another. A switch that is 'up' is 'on' in the US, but 'off' in Europe, while in Japan switches move horizontally, 'on' to the right and 'off' to the left. Under the pressure of time critical actions, inappropriate mapping to cultural stereotypes will provoke errors. (Note that such errors are the fault of the designer, not of the operator.) It is also interesting that many different organisational hierarchies are found in roughly similar nuclear power plants round the world (Moray, 1999b). Can they all be optimal? Or is each optimal for a particular socio-cultural setting?

7.6 Trust between operators and machines

Sheridan (1976) noted many years ago that automation changes the role of human operators from controllers to supervisors. As in the supervision of humans by one another, trust is central to the supervisory control of automation by humans. As machines become more adaptable and intelligent, it becomes harder for human operators to understand what the machines are doing and it becomes increasingly important that they can trust the automation. People have a tendency to develop human-like relations with machines; and recently there has been considerable research into the nature of trust between humans and machines. Muir (1994), Lee and Moray (1992; 1994) and Muir and Moray (1996) have examined operator trust and self-confidence in a simulated process control environment, and Hiskes (1994) and Moray *et al.* (1994) extended the work to discrete manufacturing. Other relevant references will be found in the series of meetings held on humans and automation, organised by Mouloua and others (Mouloua and Parasuraman, 1994; Mouloua and Koonce, 1997).

Operators' decisions to intervene appear to depend on the relation of their trust in the automation to their self-confidence in their own manual control ability. Lee and Moray (1992; 1994) developed quantitative models which predicted what proportion of time operators would spend using the automation. They proposed the following relations:

$$\begin{aligned} T_n &= c_1T_{n-1} + c_2\text{Productivity}_n + c_3\text{Productivity}_{n-1} + c_4\text{FaultSize}_n \\ &\quad + c_5\text{FaultSize}_{n-1} + v \end{aligned}$$

$$\%AU_n = k_1\%AU_{n-1} + k_2\%AU_{n-2} + k_3(T - SC)_n + k_4\text{FaultSize}_n + k_5\text{Bias} + v'$$

where T_i is the subjective trust in the automation on trial i, SC_i the self confidence of operators in their ability to exercise manual control, and %AU the proportion of time spent in automatic control. Lower case letters are weighting coefficients specific to the particular task, and v, v' random noise terms. Similar equations were developed by Moray *et al.* (2000) for a different system. Tan and Lewandowsky (1996) showed that interaction with what was thought to be a human assistant was different from that towards automation, even when in fact both assistants were automated. This suggests that there are subtle dynamics between humans and machines that may need to be carefully optimised for best performance.

The fact that intervention is related to subjective trust, which itself is related to physical events in the plant, suggests other important relations. Muir and Moray (1996) found that the frequency of monitoring a display was directly related to the magnitude of faults that occurred in the displayed values (due to sensor noise or control noise), and inversely related to the degree to which the display was trusted. It follows that operators will spend more time monitoring a faulty component and less time in monitoring sub-systems that they trust.

The latter finding supports to some extent the fear that operators will become 'complacent' and cease to monitor highly reliable sub-systems, so that if a fault ever occurs it will not be noticed (Parasuraman *et al.*, 1993; Parasuraman and Riley, 1997). Such failure to monitor is probably more important than the classical problem of vigilance in complex industrial, transportation or commercial tasks: in particular Sarter *et al.* (1997) have suggested that several aircraft accidents have been caused by 'complacency'.

A re-analysis of these findings does not show over-reliance and complacency but rather a correct choice of strategy for operators who are overloaded by the management of large and complex systems. If a sub-system is 100 per cent reliable and has never been known to fail, a rational model suggests that it does not need to be monitored, precisely because it is reliable. It is easy to show that in a large system with tens or hundreds of conventional displays or pages of computerised displays, even a mathematically optimal model of human monitoring cannot detect all faults (Moray and Parasuraman, 1999; Moray and Inagaki, 2001b). If an operator is responsible for the supervision of a plant with many degrees of freedom, many sensors and many displays, no amount of training can guarantee the timely discovery of abnormal plant conditions unless alarms and warnings are provided. Further, these must be far more reliable than the sub-system whose state they indicate. In such systems the operator may not be able, by adaptive behaviour, to 'complete the design' for the designer (to use a phrase of Rasmussen's (personal communication)).

7.7 Human error

The ultimate irony of advanced automation is perhaps exemplified in a remark made to the author by the pilot of an Airbus A340. Asked how he liked flying such an advanced aircraft, he replied that it is a wonderful aircraft to fly, but only the most experienced pilots should be allowed to fly it. This was because the automation performed very well most of the time, but the human–machine interface in the cockpit made it difficult to determine the state of the aircraft and what it was trying to do. The high level of automation, designed to ensure safe flight by relatively inexperienced pilots, can only be understood by the very best pilots. (The occurrence of 'mode errors', in which the aircraft has misled pilots as to which of several control modes it is using, are well known in this series of aircraft.)

Although data suggest that 'human error' is a cause of a high proportion of system failures in automated systems, we should note that the phrase 'human error' often conceals the real cause of accidents. As Reason (1990; 1997) has shown and as described in Chapter 1, many faults and accidents are due to design or organisational and managerial errors which put operators in situations where they are forced to make errors, even if fully trained and working as efficiently as possible. Among accidents where human error is more correctly ascribed to management than to operators are Chernobyl, Bhopal, *The Herald of Free Enterprise*, and the *Challenger* disasters; while the accident at Three Mile Island was made worse by such poor control room design that well-trained operators could not determine the plant state. Saying that an accident was caused by 'human error', or 'operator error' should always be regarded as the starting point, not the end, of an investigation: we must ask what was it about the system design or organisation that made it difficult for the operator to act correctly?

7.8 Conclusions

In the last 20 years there have been great advances in our understanding of human cognition in relation to industrial systems. We have quantitative models of humans as manual controllers in systems with bandwidths between about 0.1 Hz and 1 Hz. We know that above 1 Hz machine aiding and automation are required, while at very low bandwidths people do not behave like continuous controllers. We have begun to understand how to represent information in very rich environments, with many degrees of freedom, and how to predict monitoring behaviour. We are beginning to understand how displays support complex decisions and know that many real-time decisions are based on pattern recognition not reasoning (Klein *et al.*, 1993). We are beginning to understand how to model quantitatively and qualitatively the psychological relations between humans and the machines with which they interact. Many principles of human–machine function allocation are now far better understood than 20 years ago. But it must be said that, with regard to translating those principles into design practice and transforming basic research into design practice, much remains to be done. Such translation requires far closer co-operation between engineers, and human factors and ergonomics specialists.

In conclusion, it is the best of times for cognitive engineers, human factors experts and engineers to work together, and with computer scientists, to develop more efficient, more productive and safer large scale socio-technical systems;

systems with which it will be a pleasure for workers to co-operate and which will serve society in safer and more efficient ways.

References

1 BAINBRIDGE, L.: 'Ironies of automation', *Automatica*, 1983, **19**, pp. 755–79

2 BELTRACCHI, L.: 'A direct manipulation interface for water-based Rankine cycle heat engines', *IEEE Transactions of Systems, Man and Cybernetics, SMC-17*, 1987, pp. 478–87

3 BENNETT, K. B., and FLACH, J. M.: 'Graphical displays: implications for divided attention, focused attention and problem solving', *Human Factors*, 1992, **34**, (5), pp. 513–33

4 EDWARDS, E., and LEES, F. P.: 'Man and computer in process control' (I. Chem. E., London, 1973)

5 EDWARDS, E., and LEES, F.: 'The human operator in process control' (Taylor and Francis, London, 1974)

6 FITTS, P. M. (Ed.): 'Human engineering for an effective air navigation and traffic control system' (National Research Council, Washington, DC, 1951)

7 HANCOCK, P. A., and SCALLEN, S. F.: 'Allocating function in human–machine systems', *in* HOFFMAN, R. R., SHERRICK, M. R., and WARM, J. S. (Eds.): 'Viewing psychology as a whole: the integrative science of William N. Dember' (American Psychological Association, Washington, DC, 1998)

8 HIGGINS, P. G.: 'Job shop scheduling: hybrid intelligent human-computer Paradigm'. Unpublished Ph.D. Thesis. University of Melbourne Department of Mechanical and Manufacturing Engineering, 1999

9 HISKES, D.: 'Trust and operator intervention in a simulated discrete manufacturing process'. Unpublished Master of Science Thesis in the Department of Mechanical and Industrial Engineering, Engineering Psychology Research Laboratory, University of Illinois, USA, 1994

10 INAGAKI, T.: 'Situation adaptive degree of automation for system safety'. Proceedings of the IEEE International Workshop on *Robot and Human Communication*, 1993, pp. 231–6

11 INAGAKI, T.: 'Situation adaptive responsibility allocation for human-centered automation', *Trans. SICE of Japan*, 1995, **31**, (3), pp. 292–8

12 JONES, B. G., SHAHEEN, S., MORAY, N., REISING, D. V., SANDERSON, P. M., and RASMUSSEN, J.: 'Development of advanced direct perception displays for nuclear power plants to enhance monitoring, control and fault management'. Technical Report DE-FG02-92ER75781, Department of Energy, Washington DC, 1992

13 KLEIN, G. A., ORASANU, J., CALDERWOOD, R., and ZSAMBOK, C. E.: 'Decision making in action: models and methods' (Ablex, Norwood, NJ, 1993)

14 LEE, J. D., and MORAY, N.: 'Trust, control strategies and allocation of function in human–machine systems', *Ergonomics*, 1992, **35**, pp. 1243–70

15 LEE, J. D., and MORAY, N.: 'Trust, self confidence and operators' adaptation to automation', *International Journal of Human-Computer Studies*, 1994, **40**, pp. 153–84

16 MORAY, N.: ‘Monitoring behavior and supervisory control’, *in* BOFF, K. R., KAUFMAN, L., and THOMAS, J. P. (Eds.): ‘Handbook of perception and human performance’ (Wiley, New York, 1986) Chapter 45

17 MORAY, N.: ‘The cognitive psychology and cognitive engineering of industrial systems’, *in* DURSO, F. T., NICKERSON, R. S., SCHVANEVELDT, R. W., DUMAIS, S. T., LINDSAY, D. S., and CHI, M. T. H. (Eds.): ‘Handbook of applied cognition’ (Wiley, New York, 1999) p. 233

18 MORAY, N.: ‘Advanced displays, cultural stereotypes and organisational characteristics of a control room’, *in* MISUMI, J., WILPERT, B., and MILLER, R. (Eds.): ‘Nuclear safety: a human factors perspective’ (Taylor and Francis, London, 1999) pp. 97–112

19 MORAY, N.: ‘Culture, politics and ergonomics’. Proceedings of the HFES/IEA International Meeting, San Diego, CA, 2000, pp. 858–68

20 MORAY, N., and INAGAKI, T.: ‘A quantitative model of dynamic visual attention’ (Forthcoming, 2001a)

21 MORAY, N., and INAGAKI, T.: ‘Attention and complacency’, *Theoretical Issues in Ergonomics*, forthcoming, 2001b

22 MORAY, N., and PARASURAMAN, R.: ‘Complacency, trust and dynamic attention’. Proceedings of Conference on *Attention Management*, Office of Naval Research, Washington, DC, 1999

23 MORAY, N., HISKES, D., and LEE, J. D.: ‘Quantitative models of human interaction with automated systems’. Proceedings of the 1st conference on *Automation and human factors*, Washington, DC, April 7, 1994, pp. 15–22

24 MORAY, N., INAGAKI, T., and ITOH, M.: ‘Situation adaptive automation, trust and self-confidence in fault management of time-critical tasks’, *Journal of Experimental Psychology: Applied*, 2000

25 MOULOUA, M., and KOONCE, J. M.: ‘Human-automation interaction’ (Lawrence Erlbaum Associates, Mahwah, NJ, 1997)

26 MOULOUA, M., and PARASURAMAN, R. (Eds.): ‘Human performance in automated systems: recent research and trends’ (Lawrence Erlbaum Associates, Hillsdale, NJ, 1994)

27 MUIR, B. M.: ‘Trust in automation: part 1 – theoretical issues in the study of trust and human intervention in automated systems’, *Ergonomics*, 1994, **37**, (11), pp. 1905–23

28 MUIR, B. M., and MORAY, N.: ‘Trust in automation: part II – experimental studies of trust and human intervention in a process control simulation’, *Ergonomics*, 1996, **39**, (3), pp. 429–61

29 PARASURAMAN, R., and RILEY, V.: ‘Humans and automation: use, misuse, disuse, abuse’, *Human Factors*, 1997, **39**, (2), pp. 230–53

30 PARASURAMAN, R., MOLLOY, R., and SINGH, I. L.: ‘Performance consequences of automation-induced ‘‘complacency’’’, *International Journal of Aviation Psychology*, 1993, **3**, (1), pp. 1–23

31 PARASURAMAN, R., SHERIDAN, T. B., and WICKENS, C. D.: ‘A model for types and levels of human interaction with automation’, *IEEE Transactions on Systems, Man, and Cybernetics, SMC-30*, 2000, pp. 286–97

32 RASMUSSEN, J.: ‘Information processing and human–machine interaction: an approach to cognitive engineering’ (North-Holland, Amsterdam, 1986)

33 RASMUSSEN, J., PEDERESEN, A-M., and GOODSTEIN, L.: ‘Cognitive engineering: concepts and applications’ (Wiley, New York, 1995)

34 REASON, J.: 'Human error' (Cambridge University Press, Cambridge, 1990)

35 REASON, J.: 'Managing risks of organizational accidents' (Gower, Ashgate, 1997)

36 SANDERSON, P. M.: 'The human planning and scheduling role in advanced manufacturing systems: an emerging human factors role', *Human Factors*, 1989, **31**, (6), pp. 635–66

37 SARTER, N. B., WOODS, D. D., and BILLINGS, C. E.: 'Automation surprises', *in* SALVENDY, G., (Ed.): 'Handbook of human factors' (Wiley, New York, 1997) pp. 1026–1943

38 SHERIDAN, T. B.: 'Towards a general model of supervisory control', *in* SHERIDAN, T. B., and JOHANNSEN, G. (Eds.): 'Monitoring behavior and supervisory control' (Plenum Press, New York, 1976) pp. 271–82

39 SHERIDAN, T. B.: 'Supervisory control', *in* SALVENDY, G. (Ed.): 'Handbook of human factors' (Wiley, New York, 1987)

40 SINGALEWITCH, H.: 'Learning strategies in a computer integrated manufacturing simulation'. M.Sc. Thesis, Engineering Psychology research Laboratory, Department of Mechanical and Industrial Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA, 1992

41 TAN, G., and LEWANDOWSKY, S.: 'A comparison of operator trust in humans versus machines'. INTERNET: CybErg International Electronic Conference. http://www.curtin.edu.au/conference/, 1996

42 VICENTE, K. J.: 'Cognitive work analysis: towards safe, productive and healthy computer-based work' (Lawrence Erlbaum Associates, Marwah, NJ, 1999)

43 VICENTE, K. J., MORAY, N., LEE, J. D., RASMUSSEN, J. D., JONES, B. G., BROCK, R., and DJEMIL, T.: 'Evaluation of a Rankine cycle display for nuclear power plant monitoring and diagnosis', *Human Factors*, 1996, **33**, pp. 506–23

44 WIENER, E. L., and CURRY, R. E.: 'Flight-deck automation: promises and problems', *Ergonomics*, 1980, **23**, (10), pp. 995–1011

45 WIERS, V. C. S.: 'Human-computer interaction in production scheduling' (BETA Central Secretary, University of Eindhoven, 1997)

46 WOODS, D. D.: 'Coping with complexity: the psychology of human behavior in complex systems', *in* GOODSTEIN, L. P., ANDERSEN, H. B., and OLSEN, S. E. (Eds.): 'Tasks, errors and mental models' (Taylor and Francis, New York, 1988)

47 WOODS, D. D., JOHANNSEN, L. J., COOK, R. I., and SARTER, N. B.: 'Behind human error: cognitive systems, computers, and hindsight' (CSERIAC SOAR Reports, Wright Patterson AFB, 1994)

48 ZUBOFF, S.: 'In the age of the smart machine' (Basic Books, New York, 1988)

Chapter 8
Task analysis

Les Ainsworth

8.1 Introduction

Task analysis is widely used (and sometimes misused) to examine task performance and to assess particular features of a work situation that may have an impact on performance, such as interfaces, job aids, procedures, work organisation or training. The tasks themselves can vary considerably – from short, simple routine tasks to much more complex team-based tasks that involve high-level cognitive reasoning and decision-making. Therefore, from the outset of any consideration of task analysis methods, it is important to appreciate the implications of attempting to examine several different potential issues over such a wide range of tasks. In particular, it is necessary to accept that it is unrealistic to expect that a specific method of investigation and analysis will suit all potential task analysis situations. Accordingly, the task analysis process should be seen as the selection and utilisation of a sub-set of potential methods that will be most appropriate for a particular application. Further, there are a number of practical issues associated with carrying out a task analysis and these will also be considered in this chapter.

8.1.1 Skills required for task analysis

Before proceeding to discuss how to tailor the selection and use of task analysis methods to specific applications, it is appropriate to consider the skills that are necessary to undertake an effective task analysis. Like any analysis process, much will depend upon the way that data have to be interpreted, and in some situations, relatively little interpretation may be required. For instance, if the main purpose of a task analysis is simply to develop a comprehensive set of task descriptions, basic data collection skills, such as the ability to extract, collate and organise data from different sources, will be sufficient to undertake the task analysis. However, in most situations, task descriptions only serve as stepping stones on the way to a more complete understanding of the tasks. This is because the purpose of the analysis is to assess the adequacy of particular task features, or to predict some aspect of performance. Such analyses require more detailed interpretation of the data, and so it is important that the analyst has sufficient knowledge and experience to make these interpretations. This may require some understanding of ergonomics, psychological or training principles. For example, merely knowing that a particular parameter is displayed at an interface is not sufficient to determine whether that display is adequate. To assess the display's adequacy an analyst must make judgements such as:

- Can the display be read with sufficient accuracy from all likely viewing positions?
- Is the information presented in an appropriate form for an operator to use it effectively?
- Can the display be located easily?
- Could the display be confused with other nearby information?
- Does processing the information impose an unacceptable memory load?
- Is other information, such as past history or limit values, required in order to interpret the display? *Etc.*

For simple situations, an understanding of ergonomics and psychological principles could be developed from guidelines. However, in more complex cases, there will be no substitute for more detailed understanding and experience. The importance of the analyst's background and experience is highlighted by the finding from a review of over 60 task analyses by Ainsworth and Marshall (1998). This study showed that without exception all the analyses judged as being more detailed and insightful were undertaken by analysts with an ergonomics or psychological background.

8.1.2 A model of the task analysis process

Inevitably, task analysis is an iterative process, in which the approach taken by an analyst should be constantly modified to optimise the data collection and analysis processes. This may mean that more data have to be collected to resolve ambiguities or inconsistencies, or the analysis may have to be redirected towards those issues that are discovered to be most important. Nevertheless, despite the need to adopt an iterative approach, it is convenient to consider task analysis as comprising six broadly consecutive stages, and so the remainder of this discussion of task analysis is structured around these. An overview of the task analysis process based upon these stages is presented as Figure 8.1.

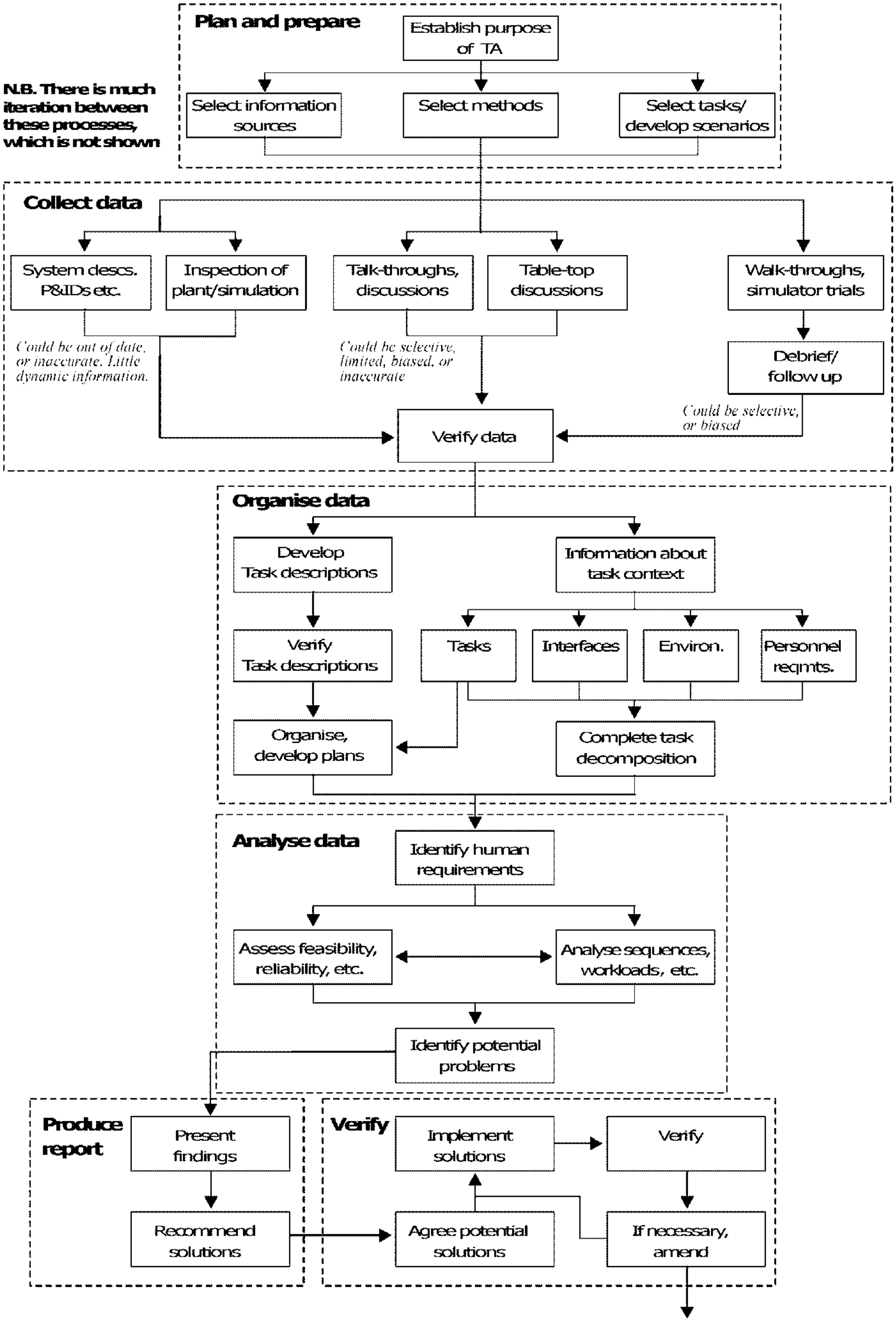


Figure 8.1. An overview of the task analysis process.

8.2 Planning and preparation

Every task analysis should start by defining the purpose of the analysis. It is then possible to plan how the analysis will be undertaken. The prime concern at this point will be to select the methods that will be used. It will then be possible to determine how to use each method, to identify an appropriate mix of data sources and, if necessary, to develop appropriate scenarios to examine particular issues.

8.2.1 General planning issues

At the outset of any task analysis it is necessary to consider the practical issues associated with undertaking that task analysis. Besides the normal project planning tasks, this should involve ensuring that adequate arrangements are made for organising and managing data collection. The aim of this is to predict potential data collection requirements early in the project and then to set up arrangements to ensure that these can be met. As a starting point, the analyst should establish a contact person on site who can assist in identifying the appropriate data sources or personnel to address particular issues. Then, throughout the project, the analyst should ensure that, as far as possible, arrangements are made in advance for personnel to be available and for rooms, simulators or other equipment to be prebooked.

Arrangements must also be made to ensure that all data that are collected are stored securely. There are two sides to this, because as well as providing back up copies of any vulnerable data, such as computer files, it is also vitally important that the anonymity of responses or comments from individuals is maintained.

From the outset it is also wise to make contingencies for obtaining additional data. For instance, it could be helpful to warn some individuals that further questions may arise and then to establish the best times that they can be contacted, so that they will be receptive if clarification has to be sought.

8.2.2 Selection of task analysis methods

There is a wide range of methods available for collecting and analysing task-related data. Most of these methods can be used for different design or assessment purposes, but they each have particular strengths and weaknesses. Therefore, the analyst should vary the use of the available methods according to the requirements of each project. Thus, the analyst should start by clearly defining the issues that are of interest and then methods that could be used to address these issues can be identified.

There are two main strategies for selecting task analysis methods. The first of these is known as Purpose-driven Methods Selection and involves selecting the methods on the basis of the main issues, such as training or interface design, that are to be investigated. The alternative to this is Performance-driven Methods

Selection, which defines the methods in terms of more specific aspects of task performance, such as the time taken, or stress experienced.

8.2.2.1 Purpose-driven methods selection

Most task analyses can be characterised as being undertaken for one or more of the following broadly defined purposes:

- Allocation of function.
- Interface design or assessment.
- Task and procedure design or assessment.
- Personnel selection.
- Operability and workload assessment.
- Human error identification or human reliability assessment.
- Training requirements or assessment.

For each of the above there are particular methods or approaches that are most appropriate. For instance, for task and procedure design an analyst must be able to combine procedural and technical information. Thus, written procedures will be a valuable source of the former information, but even in well-established procedures these must be checked for omissions and simplifications. Therefore, technical information, such as system descriptions or piping and instrumentation diagrams, is also needed. If possible, insights about the task context should be obtained by undertaking walk-throughs of representative tasks with typical operators in realistic settings, such as a simulator, a mock-up, or on the actual plant. If more specific information is required, this could be obtained by individual discussions with technical experts, or by table-top discussions with a selected group, based upon a specific procedure or scenario.

8.2.2.2 Performance-driven methods selection

This approach is only appropriate when, at an early stage, the analyst can determine those specific aspects of performance that are of particular interest. A typical example of this would occur if an analyst wished to determine task times, in order to ascertain whether the tasks could be completed within an acceptable time. This would require collecting data about task times in as realistic a setting as possible. Thus, an analyst might consider direct observations of task performance or uninterrupted walk-throughs in a simulator. The analyst might also wish to make a subjective assessment of workload.

Clearly, timelines would be an obvious way to analyse and present temporal data. If these indicated that the projected task completion time was very close to the permissible time, it would be appropriate for the analyst to verify the timings and the scenarios very meticulously. However, in many cases there will be sufficient margin to allow relatively crude timings for the individual steps.

Similarly, task analysis methods could be selected to provide data about other very specific aspects of task performance, such as decision-making, levels of task difficulty, or communications requirements.

8.2.3. Selection of data sources

At the start of any task analysis there will be some information already available about the tasks and the task interfaces, much of which will be in printed form (e.g. system descriptions, interface layout diagrams, incident reports or procedures). It may also be possible to inspect either the actual human–machine interfaces, or some realistic representation of them, such as a simulator. Much of this information will be useful for defining the context in which the tasks are undertaken, but in order to assess task performance and to identify features that could affect performance, it is likely that this will have to be supplemented with fresh information about task performance.

It is essential that the analyst selects and manages information carefully and in a manner that will be most effective for addressing the important issues. In particular, the analyst must make painstaking efforts to ensure that all data are as accurate as possible. This means that the analyst should try to use the most up-to-date, preferably controlled, copies of all design documentation and where possible the accuracy of all information should be verified by checks against an independent source. However, whilst this can identify incorrect or out-of-date information, some task variability is to be expected because of different personal strategies, or because of slightly different technical requirements. Therefore, it can be particularly useful to probe more deeply into task variability. For example, the author recalls an analysis in a process plant where a highly experienced operator explained how a new process train could be brought into service. However, it was only by asking a less experienced operator about the same task that it was discovered that another, more gradual strategy was also adopted. Further investigation revealed that the experienced operator had for several years, been, unwittingly, subjecting the system to excessive thermal stress by using this faster approach.

Different data sources provide various parts of the task analysis jigsaw and so it is important to focus upon sources that provide the specific information that is required. Some guidance on this can be provided by looking briefly at the characteristics of six potential types of task analysis data.

8.2.3.1 System records

These, such as system descriptions, screen layouts, alarm schedules or piping and instrumentation diagrams, can be useful for obtaining background information. Whilst flow charts and logic diagrams can be particularly useful for understanding the links between different sub-systems. However, it is also necessary to exercise some caution when using some of these sources. For instance, functional and system descriptions can be a useful source of information to identify high-level tasks and to define task goals, but there is also a risk that more detailed information about interfaces may be misleading. This is because the writers may make unwarranted assumptions about the interfaces, or else the descriptions can be written at such a high level that they can be interpreted ambiguously.

8.2.3.2 Task records

These provide an indication of how tasks should be done, or they record how they were reported to have been done on a specific occasion. These records include procedures, job descriptions, training documentation, logs and critical incident reports. Of these, procedures and training documentation are useful for developing task descriptions, but in both cases there is the risk that some steps may be covered in less detail or may be omitted altogether. Therefore, these sources should never be used as the sole source of information for developing task descriptions, because such task descriptions will merely mirror the procedures, rather than evaluating them. Other sources, such as logs, or operational feedback reports can be useful, but the analysis of these can be extremely resource-intensive.

8.2.3.3 Safety documentation

This includes safety reports, human reliability assessments and HAZard and OPerability Studies (HAZOPS). However, such documentation generally takes a normative view of how a task ought to be done and so the task information is likely to be incomplete. Another problem of safety documentation is that it often looks at very narrow aspects of tasks and so it is often better for describing abnormal situations than for describing the task itself in a more normal context. If human reliability analyses are used, the analyst should carefully check the underlying assumptions, which are often not fully described, and in some cases may be inaccurate.

8.2.3.4 Interface inspection

This provides a very effective way to examine interfaces. Where interfaces have been in use for some time, modifications that the operators have made (e.g. marking critical values on instruments) can provide valuable information for the analyst. Mock-ups, simulators or virtual reality displays can also be useful, but it is always necessary to check that they represent the current configuration of the controls and displays, or the latest design intent.

8.2.3.5 Subjective data

This ranges from relatively straightforward verbal descriptions of the tasks that are undertaken, through to subjective ratings about particular task features. Such data can be obtained by structured or semi-structured discussions with individuals or groups, or by administering questionnaires. However, questionnaires generally require an absolute minimum of 10 respondents. As well as selecting an appropriate method, it is also necessary to select those individuals who are most appropriate to provide the information required. For instance: experienced operators or maintainers to develop task descriptions of existing tasks; appropriate technical experts for technical issues; trainers or trainees to determine which tasks are most difficult.

At this juncture it should be noted that a common failing in task analysis is to rely too heavily upon highly experienced operators. This can sometimes lead to perpetuation of poor design features, because these operators have become so habituated to these features, that they now regard their adaptive coping strategies (which occur without conscious effort) as normal behaviour.

8.2.3.6 Task observations

Observations of walk-throughs, simulator trials, or the observation of actual tasks are particularly useful ways of identifying potential difficulties that an operator may experience. These observation sessions can be interrupted by the analyst seeking clarification at particular points, or else they can be undertaken in real-time with no interruptions. In the latter case, some form of debriefing session should be undertaken as soon after the actual session as possible.

8.2.4 Scenario selection and development

In large systems it may not be feasible, or even desirable, to undertake task analyses of all the possible tasks that might be undertaken. Therefore, it will be necessary to focus upon a limited sub-set of tasks that is sufficiently comprehensive and representative to cover the issues of concern. In this situation the analyst should develop a set of scenarios that can be used for the data collection and analysis.

Perhaps the most important requirement for a set of scenarios is that they should be comprehensive. This does not mean that every possible eventuality has to be covered, but that the scenario, or set of scenarios, will be sufficiently wide-ranging to cover all the main issues that are potentially of concern. For instance, if an operator has to monitor the automatic closure of several valves and to intervene if there are any failures, it would not be necessary to develop a scenario in which every single valve failed to close. It would be sufficient to postulate only one valve failure, provided that all the valve interfaces were similar. Thus, for an interface assessment, the scenarios should ensure that all the features of the interface to be assessed are included. This would include all the types of control and display objects and means of navigation between the major interfaces.

When a task analysis is being undertaken to examine particular procedures, it will normally be necessary to cover all the main instructions. It may also be necessary to explore different pathways through the procedures, but generally there will be no need to consider all the potential pathways within a set of scenarios. Nevertheless, it would still be prudent for an analyst to make his own check that any branches that are not tested practically do not present unforeseen difficulties.

For task analyses that are undertaken to assess a safety case, the scenario will generally be directly based upon the technical faults that are of significance to that safety case. Generally, maximum demands for human performance will be considered, because the requirements upon the operator during these tasks will ‘bound’ the required performance for less demanding tasks. Thus, scenarios that are developed to examine safety cases are typified by a maximum number of human actions, undertaken by a minimum staff complement. However, such assumptions must be carefully explored before they are accepted. For example, a minimum workload can, in some circumstances, also lead to sub-optimal performance.

To assess workload, or to ascertain the time required, it will generally be necessary to develop both a *maximum workload* or *longest task duration* scenario, and another scenario that includes the more likely situation. However, whilst the most difficult scenario will inevitably include some unlikely events, it should not involve such a mix of unlikely events that it becomes inconceivable. For completeness, it may also be appropriate to examine a *lowest workload*, or *shortest duration*, scenario. As well as systematically developing scenarios on the basis of the above criteria, it is also necessary to report the criteria that were used.

8.3 Data collection

The data collection process must be carefully managed in order to ensure that the information that is collected can be used to address the issues of concern. It is particularly important to ensure that the information is up-to-date and representative of the current situation, or the latest design proposals. Which means that, whenever possible, controlled copies of design documentation should be used. Even so, there are times when the design is modified after the task analysis data have been collected. Therefore, it is also necessary to record the versions of all data sources that are used in a task analysis and to verify the basis of all recommendations before they are implemented. Care must also be exercised during data collection to ensure that bias is eliminated. Finally, once data have been collected, they must be stored carefully so that they can be accessed easily during the analysis, without compromising the confidentiality of any comments that are provided. Data can be collected from existing sources, from discussions with users or other experts, or by specially designed exercises, such as simulator trials.

8.3.1 Using existing data sources

The documentation that is produced during the design of a complex control system serves many purposes and is written at different times during the development of that system. Thus, it should not be surprising to find some contradictions or ambiguities. Clearly, these could be caused because the design has changed since a document was written. However, it is also possible that the person writing this documentation may have limited information or understanding of the detailed operation of the interfaces, and may either make unwarranted assumptions, or write the interface information in an ambiguous manner. For instance, the person writing a system specification may assume that feedback of a particular control action will be provided by a flashing light, but unless this requirement has been specified, the actual instrument that is to be purchased may not incorporate any facilities for providing such feedback. Thus, the analyst must take great care when interpreting any system specifications. In particular the analyst must ensure that the specification does not make any unwarranted assumptions about the interfaces. Also, where possible, all data should be checked from an independent source. If any conflicts are discovered, the analyst will have to probe deeper, and this may involve some assessment of the relative accuracy of different data. One of the best indicators of accuracy is the recency of the data. However, it must be stressed that it is the date of the source material upon which documentation was based, rather than the date of the documentation itself, that is important. For instance, a fairly recent training guide may well be based upon much earlier information, and thus be less accurate than other, ostensibly less recent reports.

8.3.2 Verbal data

Verbal data obtained from discussions with operational personnel, potential users or technical experts can provide useful information about many aspects of tasks. These can involve direct reporting of tasks, or semi-structured discussion sessions in which an analyst directs the respondents through particular questions or issues of concern.

8.3.2.1 Direct reporting of tasks

It has been concluded from several reviews such as Ericsson and Simon (1980) that verbal data provide relatively accurate descriptions of activities when made by the person undertaking them. For accuracy, verbal reports should be taken during the execution of the task or shortly afterwards, and the verbalisations should be restricted to the description of actions rather than explanations of mental processing or decision-making. Furthermore, whilst it is possible to verbalise reasoning in complex problem solving tasks, such verbalisations appear to alter the approach that is taken to these tasks (Gagné and Smith, 1962). Therefore, during any reporting sessions it is essential to do everything possible to encourage verbalisation and to establish a rapport with the respondent. This can be achieved by clearly explaining the purpose of the reporting and assuring the respondents that all comments will be treated confidentially. Any contradiction of the respondents or comments on their performance should be strictly avoided. It can also be helpful to pre-warn the respondents before each session about the areas to be covered and, if possible, to arrange times that are convenient for them. If the reporting is being done on-line whilst actually undertaking a task, or some representation of it, the analyst should also try to avoid any interruptions. If any clarification is required, it is generally best to obtain this during a debriefing session that is held as soon as possible after task completion, or by undertaking an untimed session that can be interrupted. However, during a talk-through, or an untimed walk-through, interruptions should not prove too disruptive, provided that the analyst asks these questions at appropriate break points, such as when the respondent has completed their description of a particular step, or sequence of steps.

8.3.2.2 Semi-structured interviews

Semi-structured interviews or discussion sessions will be most effective if they are based upon particular issues, such as a procedure, a fault scenario, a specific interface, or if they are undertaken as a debriefing session. This should provide a clear framework through which it is possible to proceed logically, without jumps to seemingly unrelated items. Before each session the analyst should have prepared some questions, but during the discussions a flexible approach should be adopted so that, if it is appropriate, some issues can be probed in more detail, whilst others can be skipped. However, within this flexible approach, the analyst must remain vigilant and if the respondent moves away from the point, they should be gently returned to the issues of interest.

The two biggest problems in any task analysis discussions are bias from the respondent and misinterpretation by the analyst. These can be minimised by using more than one respondent, by avoiding leading questions and by asking different questions to obtain the same information. Also, if the participants make any hesitations or corrections, these should be probed by the analyst, because this often indicates that they are compensating for a sub-optimal interface. Finally, for complex task sequences it is often worth the analyst providing a verbal précis of their understanding of the task so that the respondent can indicate whether this interpretation of their own comments is correct.

8.3.2.3 Table-top discussions

Table-top discussions are group discussion sessions in which a selected group of experts discuss specific operational or technical issues, usually structured upon a specific procedure or scenario. Ideally, these groups consist of between four and eight members, with a chairperson who has some human factors or psychological experience. It is also helpful if at least one person has recent operational experience. The other group members should have expertise in an appropriate technical area and there should only be one representative for each area of expertise. If possible, each technical expert should be at roughly the same management level, otherwise, there is a risk that the more senior person's views will tend to prevail. If a number of table-top sessions are required it will be helpful, but not essential, for the same person to represent a particular area of expertise at all meetings.

In view of the specialist nature of table-top discussions, it is necessary to ensure that detailed preparations are made and that the discussions are carefully managed. The preparations involve pre-warning the group members about the procedures or scenarios that will be discussed, and arranging an appropriate meeting time, a meeting room and making available any documents or drawings. During the table-top discussion, the chairperson must ensure that each issue is properly discussed and that the opinions of all the group members are elicited. A corollary of this is that, if it is suspected that much of the meeting will be spent with one person, this information should be elicited in a discussion prior to the table-top meeting.

8.3.3 Observing performance

The third way of collecting task analysis data is by directly observing the performance of a task, or some representation of it. Inevitably, this will be linked with some verbal data collection, either whilst the task is being undertaken, or as a debriefing session afterwards and so the issues about verbal data in the above section also apply.

Data about task performance can be recorded by written records, computer logs of the operators' actions, alarms and the values of key parameters, or by audio or video-recording. Of these, the most useful is usually for the analyst to record the operators' actions and any problems that are identified as a written record during the task. Computer logging can be useful for accurately measuring the times of key control actions, especially for complex control sequences. Audio-recording of an on-line commentary by the analyst can also be very effective, but it can prove disruptive if the commentary can be heard by the operators. Video-recordings can take some time to analyse, but they can be particularly useful for the analyst to obtain clarification later in the analysis, or as a memory aid that can be played back to the operators during a debrief. However, it must be noted that there are professional codes of conduct relating to the use of audio or video-recordings by members of particular professional organisations and these should be adhered to. These require that audio or video-recordings should not be made unless all the participants have agreed to this, and that these recordings should only be used by the analysts themselves (not management) for a particular task analysis.

Other specific requirements for observing performance are that the participants should be allowed some time to get used to the presence of observers and that the observers must not disrupt the task that they are observing.

8.4 Data organisation

It is necessary to organise task information to improve understanding of the constituent tasks and to ensure that important data are not missed. This will generally involve breaking down a complex task into a comprehensive set of subtasks that are each summarised in a short task description. Then, for each of the lowest-level tasks, more detailed information is obtained about specific features of the task that are of interest, such as the time taken, the information displayed, or the training requirements. The task breakdown is undertaken by hierarchical task analysis (HTA) and then a task decomposition is used to focus upon specific issues for each task description.

8.4.1 Hierarchical task analysis (HTA)

HTA requires the analyst to define the main goal of a task, which is then broken down into a small, but comprehensive, set of subordinate goals. This process is known as *redescription*. Each of the subordinate goals is then further redescribed and this process continues until the analyst is satisfied that no greater detail is warranted. Thus, at the end of this process, the analyst has a comprehensive set of task descriptions that describe all the constituent sub-tasks that might be necessary to satisfy the main goal.

8.4.1.1 Redescription strategies and stopping rules

One of the main purposes of redescription is to ensure that no task elements are missed. Therefore, at each stage of redescription, the analyst should try to break down the immediate parent task into no more than seven sub-tasks. If the number of redescribed tasks exceeds this, it becomes difficult for the analyst to conceptualise them together and so there is a risk that a task element may be missed. At each redescription, it is essential that the goal of the immediate parent task can be completely satisfied on every occasion by satisfying the sub-goals of each of the redescribed tasks.

There are many ways that tasks can be redescribed that satisfy the above criterion for being comprehensive. However, developing the redescription directly from the hierarchical structure of a written procedure should be avoided.

One of the most useful starting points for redescription is to break tasks down into their main functions and then sub-functions. However, at the lower levels, it often becomes more appropriate to use different systems, equipment or interfaces as the basis for redescription. In multi-person tasks it can often be helpful to develop the redescription so that tasks done by the same person, or at the same location, are kept together. After each stage of redescription it can be useful to re-arrange the set of redescriptions into the sequence in which they would normally be undertaken or encountered.

The redescription process should continue to the point at which further redescription would not be fruitful. In practice, analysts usually have a good feel for the point at which redescription should be stopped. However, typical stopping rules would include: stopping when the main interface items can be identified; continuing to the point at which the workload associated with a task can be assessed; or stopping when the underlying skills and knowledge that are required can be defined.

8.4.1.2 Use of separate HTAs

In most systems, a comprehensive HTA of all the potential tasks would be extremely unwieldy. Therefore, it is appropriate to provide some guidance as to when separate HTAs should be produced. Definitely, different HTAs should be produced for different scenarios, because these are usually mutually exclusive. Similarly, it is often useful to undertake separate HTAs for different procedures, though this may not be appropriate for procedures that are closely linked together.

Tasks that are relatively self-contained can also be analysed as separate HTAs. Thus, operational and maintenance tasks could be treated separately, provided that they are undertaken independently. For other tasks that are undertaken by different persons, it can be appropriate to develop separate HTAs, but this should only be done when they involve very little communication with others, otherwise important interactions could be missed.

8.4.1.3 Plans

Once an analyst has produced all the task descriptions for an HTA, it is necessary to define the interrelations between them. This involves specifying how the sub-tasks at each level would be undertaken in order to ensure that the superordinate goal was satisfied, and is known as a *plan*. A simple plan would be a linear sequence (e.g. Do A, then B, then C), but it is also possible to have more complex plans that include conditional statements, branching and looping. For example, do A until valve opens, then do B if pressure above 90 psi, or else do C.

8.4.2 Task decomposition

Task decomposition involves systematically obtaining additional information about specific task features for each of the lowest-level tasks. This is usually achieved by filling out a tabulated display in which the first column lists all the task descriptions while the remaining columns list different decomposition categories. The information that is collected in this way can cover any aspect of the task or task performance. For example, it could be simple descriptive information such as the controls used or the feedback that is given. However, it is also possible to use a direct measure of performance, such as time taken, or even to make an assessment of specific ergonomics/psychological issues, such as an assessment of knowledge and skill requirements, or a qualitative or quantitative assessment of potential human errors.

Focusing upon specific issues in this way appears to be a great aid to the task analysis process. For instance, Ainsworth and Marshall (1998) found that analyses with six or more decomposition categories were generally judged to be much more useful than analyses that either had fewer decomposition categories, or had not undertaken any formal task decomposition. This showed up in the level of detail shown in the analyses, with the studies that had made more use of decomposition being typified by giving several detailed recommendations for improvements, whilst the other task analyses tended to limit their recommendations to a few very generic points.

Task decomposition categories should be defined early in a study so that it is not necessary to repeat earlier work in order to obtain data, though some iteration is almost inevitable. Similarly, if more than one analyst is involved, there should be close liaison between the analysts from the start to ensure that all analysts are recording similar information.

8.5 Data analysis

By the time that data have been collected and organised the analyst will already have identified many of the potential mismatches between human requirements and the facilities that are provided by the system. Similarly, it will also be evident where there will not be problems. However, it may also be helpful for the analyst to undertake some further analysis of the data. There are several techniques available for this, and some guidance for selecting them is given below. Readers who require further information about any of these techniques are referred to more detailed descriptions in Kirwan and Ainsworth (1992).

8.5.1 Analysis of informational requirements

This information can be used to help define the interface requirements or the knowledge requirements for a task. Graphical techniques, such as information flow diagrams or functional flow diagrams are appropriate analysis methods for this purpose.

8.5.2 Representation of task sequence

Link analysis and operational sequence diagrams can provide an effective visual way of representing task sequence and decision-action diagrams can also be useful.

8.5.3 Temporal requirements and workload

Timeline techniques can be used to ascertain whether operators have sufficient time to undertake all the necessary actions and to identify potential conflicts when undertaking a complex procedure. Once the data have been plotted in this form, it is relatively easy to assess the impact of various modifications to the way that the task is organised.

Various measures of mental workload can also be applied to these data to assess whether a task is within a person's capabilities. However, in this respect it is important to note that both overload and underload must be considered.

8.5.4 Errors

Task steps where there is a potential for human error are often investigated by using either event trees or fault trees that show the logical and temporal dependencies between human interactions. Another useful technique for identifying potential human errors is a human HAZOP. This is a table-top method that requires experts to make judgements for each task step about the likelihood of particular types of human errors from a prescribed list.

8.6 Report preparation

It is essential that task analysis reports convey all the information about a task analysis exercise in a clear and accurate way. The following guidelines apply throughout task analysis reports:

- Explain how all data were obtained, so that readers can assess how much confidence they should place upon the findings. In particular, it is necessary to describe the experience and background of all personnel who have contributed to a study.
- Clearly explain all task analysis conventions, so that readers who are not familiar with task analysis can fully appreciate the data.
- Establish meaningful conventions for presenting data and use these consistently throughout the report.
- Avoid ambiguous, or conflicting, wording in task descriptions.
- For diagrams that extend over several pages, clearly mark the entry and exit points.
- Present specific recommendations and also explain the underlying reason for them, so that it is possible for alternative approaches to be taken that may satisfy the same requirement if the specific recommendation is not practical.

8.7 Verification

Often recommendations from task analyses are either not implemented, or they are implemented incorrectly, because the person responsible for implementing them does not properly understand the purpose of the recommendation. Whilst much can be done to avoid this by paying assiduous attention to explaining the background to all recommendations within a task analysis report, it is also appropriate (if it is feasible) for the analyst to check the implementation of any recommendations. Clearly it is best for this to be done by examining the actual implementation, but it can also be achieved by an exchange of letters. This would involve the organisation explaining how each recommendation will be implemented and then the analyst would make a response about the adequacy of this implementation.

References

1 AINSWORTH, L., and MARSHALL, E.: 'Quality and practicability in task analysis: preliminary results from two surveys', *Ergonomics*, 1998, **11**, pp. 1607–17

2 ERICSSON, K. A., and SIMON, H. A.: 'Verbal reports as data', *Psychology Review*, 1980, **87**, pp. 215–51

3 GAGNÉ, R. H., and SMITH, E. G.: 'A study of the effects of verbalisation on problem solving', *Journal of Experimental Psychology*, 1962, **63**, pp. 12–18

4 KIRWAN, B., and AINSWORTH, L. K. (Eds.): 'A guide to task analysis' (Taylor and Francis, London, 1992)

Chapter 9
Training and technology for teams

Raphael Pascual, Matthew Mills and Simon Henderson

9.1 Introduction

This chapter discusses two approaches for enhancing control room teamwork. The first part of the chapter discusses a number of contemporary techniques which have been utilised successfully for team training. The second part of the chapter highlights design imperatives for building effective team support technology.

The rapid advances made in team training over the past 15 years have resulted primarily from the explicit recognition of the contribution that teamworking can make to the effectiveness with which complex tasks are undertaken. Traditional approaches to team training are often founded on the assumption that if every team member does their own job well, then the 'team' as a whole will function effectively (Taylor, 1911). Such approaches are consequently geared towards training a collection of individuals to do the tasks specifically allocated to their role – the focus thus being on training taskwork. This approach can work well where individuals' tasks are tightly defined, stable, and self contained, e.g. for workers on a manufacturing production line. Perhaps unsurprisingly, this approach also renders team members poorly equipped to deal with ill-defined, complex, shifting tasks, which require significant co-ordination with other team members in order for the tasks to be achieved effectively. Such conditions tend to be the norm for teams operating within control rooms.

Recent advances in team training have focused on training the members of a team together as a composite unit, with implicit and explicit interactions between team members being targeted specifically for practice and enhancement. Increasingly, such training is accounting for the complexities of the naturalistic environments in which teams must operate. This shift of emphasis requires a different way of thinking about how tasks are undertaken by a collection of individuals. In addition, it also requires that individual task demands are considered within the broader context of the interpersonal infrastructure, communication and co-ordination requirements of the team as a whole. Thus, the focus of these more recent approaches is on training *teamwork*.

Such an approach is the 'systems approach to training'. This is based around the 'training cycle', a generic variant of which is shown in Figure 9.1 (adapted from Rugg-Gunn *et al.*, 1999).

The training cycle comprises the following stages:

- **Analysis**: Identifying the training requirements (what needs to be trained).
- **Design**: Identifying how to train it (the training strategy).
- **Conduct**: Delivering the team training itself.
- **Evaluation**: Evaluating the team training.

Each of these stages will now be considered.

9.1.1 Analysis: Identifying what needs to be trained

Individuals are generally trained in order to be able to 'do' particular jobs to a desired standard. Training is thus designed to take an individual's existing capacity to do a job and increase it to a desired level via education, practice, and derived

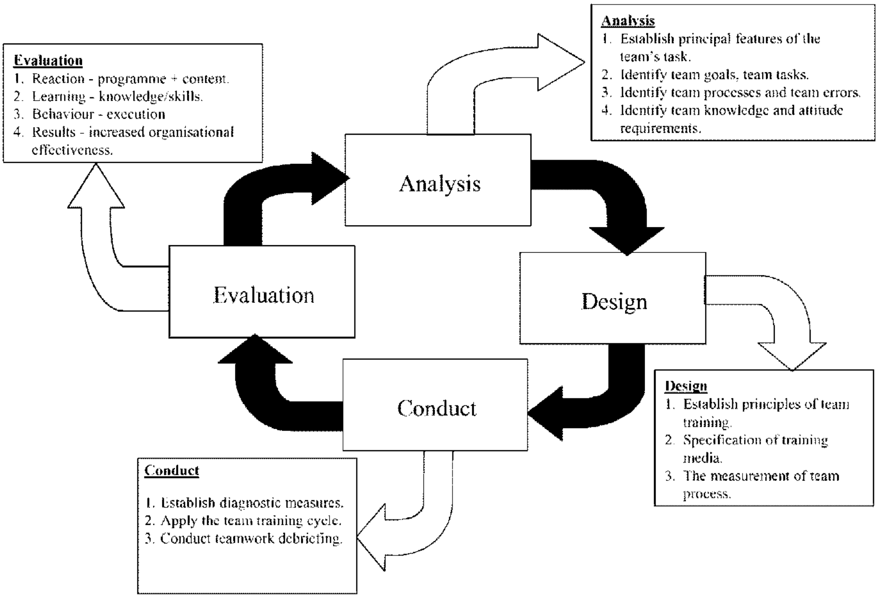


Figure 9.1. The training cycle (Adapted from Given (1993) and Rugg-Gunn et al. (1999)).

understanding. In order to assess whether a trainee has reached the desired level of capability, it is necessary to establish whether s/he has achieved a desired level of competency in the range of skills required to do the job. Such competencies can be specified as requirements for both taskwork and teamwork. In addition, the notion of a competency can also be applied at an aggregate level to the whole team (i.e. does the team exhibit the desired level of competency required to conduct its tasks effectively?).

The processes of both analysing the requirements, and designing the training can thus, simplistically, be viewed as 'starting at the end and working backwards'. In other words, defining where one wants to get to is prerequisite to identifying how to get there. By defining the competencies required to do a job, it is possible to identify what needs to be trained (the 'training requirements') in order to raise a trainee's (or indeed practitioner's) competencies to the desired level. This training requirement can then itself be used to identify an appropriate strategy for implementing this training.

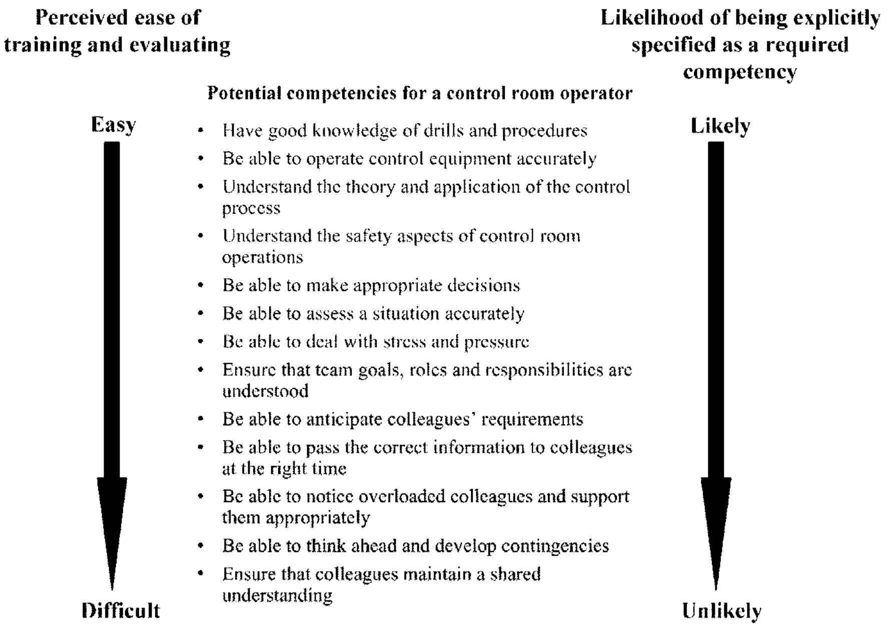
The key question upon which the effectiveness of team training hinges is, therefore, how to identify what team and task competencies are required in order for the team to conduct their task(s) effectively. Fortunately, contemporary research offers a range of tools for achieving this (Klein *et al.*, 1989; McIntyre and Salas, 1995). Such tools largely consist of interview, observational or verbal rehearsal techniques that are designed to elicit and capture an expert's thinking in relation to a task. The processes inherent in these approaches unlock many of the team interaction activities required to accomplish successfully critical tasks. The approaches can also be supplemented by drawing from generic teamwork taxonomies (i.e. definitions of non-task-grounded behaviours engaged in by successful teams, e.g. McIntyre and Salas, 1995).

Cannon-Bowers *et al.* (1995) have suggested that teamwork competencies comprise three elements (these have been discussed more fully in Chapter 5):

- The requisite **knowledge**, principles and concepts underlying the team's effective task performance.
- The repertoire of required **skills** and behaviours necessary to perform the team task effectively.
- The appropriate **attitudes** on the part of the team members (about themselves and the team) that foster effective team performance.

Figure 9.2 shows a set of hypothetical generic competencies for a team member operating within a control room, together with a generalised perception of the ease of training and evaluating such competencies, and the subsequent likelihood of each competency being explicitly specified as a requirement.

It is probably difficult to prioritise these competencies with any confidence, as all are causally related – in other words, 'a chain is only as strong as its weakest link'. However, unless they can be defined clearly and assessed precisely, it is likely that they will remain implicit and thus can only be subjectively assessed.



Potential competencies for a control room operator

- Have good knowledge of drills and procedures
- Be able to operate control equipment accurately
- Understand the theory and application of the control process
- Understand the safety aspects of control room operations
- Be able to make appropriate decisions
- Be able to assess a situation accurately
- Be able to deal with stress and pressure
- Ensure that team goals, roles and responsibilities are understood
- Be able to anticipate colleagues' requirements
- Be able to pass the correct information to colleagues at the right time
- Be able to notice overloaded colleagues and support them appropriately
- Be able to think ahead and develop contingencies
- Ensure that colleagues maintain a shared understanding

Figure 9.2. Potential competencies required for a control room operator.

9.1.2 Design: Identifying how to train team competencies

Once training competencies have been identified, consideration can then be given to the following activities:

- **Establishment of team training principles**. Identify what stage of training is being designed for (e.g. within the context of other training which is being delivered); and also identify the experience level of the trainees. The issues guide the form of training that should be implemented.
- **Specification of training media**. Match the training delivery media with the requirements of the trainees. Rugg-Gunn *et al.* (1999) exemplified this process using the example of high-fidelity and desktop simulators. They suggested that high-fidelity simulators are inappropriate media for novices, as they are designed to provide practice opportunities for more experienced operators who have undergone basic training. Simple desktop simulations, on the other hand, which would be appropriate for beginners, are not suited to experienced operators (who need to practise in complex task environments with a full range of operational scenarios).
- **Exercise trigger events**. Once trainees have acquired basic skills and knowledge, event-based training can be provided to enable trainees to practise skills and behaviours, eliminate errors and eventually move such processes

towards automaticity. In order to acquire the appropriate competencies (and thus achieve the training objectives) exercise trigger events are designed to expose trainees to situations in which such competencies are required to achieve a successful outcome. Methods which address the design and implementation of team stressors within a training context have been discussed in detail by Pascual *et al.* (2000).

- **Performance measurement tools**. These are then designed in order to collect and assess trainee performance in dealing with the events, to identify whether required competencies are being utilised.
- **Evaluation and feedback**. Methods for providing feedback are designed such that trainee performance feedback can be given back to the trainees within the context of the training tasks they have undertaken.

9.1.3 Conduct: Training practice

It is clear that trainees learn far more effectively through practice than by simply 'being told' (Salas *et al.*, 1997). By doing, trainees have the opportunity to:

- put theory into practice;
- see the consequences of their actions;
- understand the nature of causality within their domain of operation (e.g. derive target-response pairing)
- engage in a process of self-learning and discovery;
- acquire the competencies (knowledge, skills and attitudes) required to do the job for which they are being trained.

Figure 9.3 shows the 'practice' part of the training cycle and exemplifies how a team is coached by an instructor to learn from the experience of practice. Each of these stages will now be described briefly.

- **Clarify plans/goals (pre-brief)**. This stage is concerned with providing the team with a briefing as to what it is expected to achieve. Trainees' attention should be drawn to what is being taught, so those trainees are able to interpret and classify their experiences as they take place;
- **Perform/practice**. This stage involves exposing the trainees to the scenario events containing the triggers and stressors to teach and evaluate required team and task competencies. The practice should be 'hands-on', and can be real or simulated;
- **Diagnose performance**. This stage involves monitoring a trainee's individual and the team's collective performance in relation to the scenario triggers, using a range of diagnostic tools. Diagnostic tools (and their related measures) could include an assessment of the factors including a team's information passing, leadership, backing-up behaviour and communications.

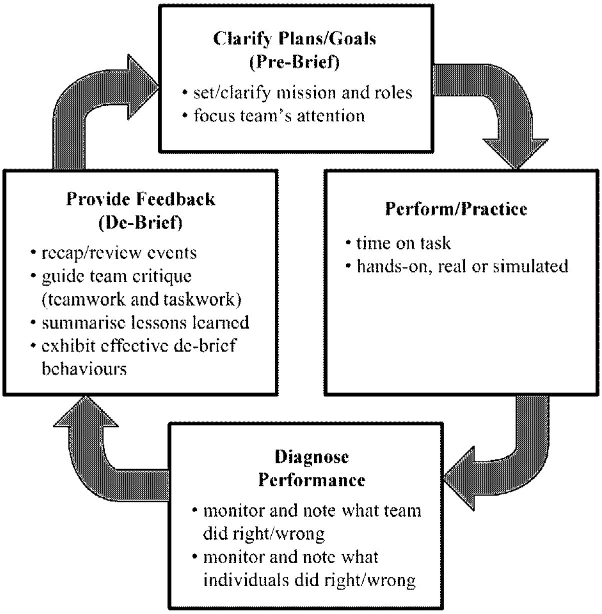


Figure 9.3. Delivering team training (NAWC-TSD, 1997).

- **Provide feedback (debrief).** This stage concerns the provision of feedback to the trainees on their performance. The debrief should follow the plan depicted in Figure 9.4.

This training delivery process thus works to expose trainees to situations in which team competencies are required to achieve a desirable outcome. It provides an objective assessment of their performance against a series of measures and provides specific, grounded feedback to trainees on how to develop their teamworking skills in support of their taskwork.

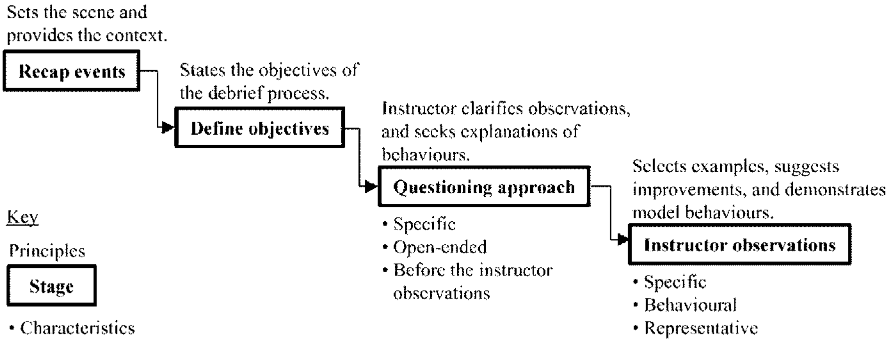


Figure 9.4. The de-brief process (NAWC-TSD, 1997).

9.1.4 Evaluation: Evaluating the effectiveness and impact of the training

Saari *et al.* (1988) summarised results from a large-scale survey of management development in 611 organisations, and concluded that few conduct any evaluation of training effectiveness at all. (In other words, assessing whether the training has actually delivered increased organisational effectiveness.) One potential reason underpinning this situation is the fact that team training theory and practice has only come to fruition over the past 15 years. From a theoretical perspective, Kirkpatrick's (1976) typology is still employed. This proposes four axes for evaluation of team training, assessed in terms of trainees' *reactions* to the programme and its content; *learning* – the acquisition of knowledge and/or skills; *behaviour* – changes in the extent to which trainees can execute desired training-related behaviours; and *results* – the extent to which job behaviours change and yield increased organisational effectiveness.

It is worth noting that various researchers have suggested extensions to this typology, including consideration of pre- and post-training trainee motivation, and the broader context of organisational development (see, Mathieu and Martineau, 1997).

In addition, Fernall *et al.* (1995) have considered the development of generic system performance and effectiveness measurement approaches, which can assess the organisational impact of any form of intervention, including training. It is suggested that such approaches are likely to be critical in the evaluation of team training regimes for control room operations, where performance relates directly to issues of safety, efficiency (and thus profitability) and the associated audit and assessment processes.

9.1.5 What are the implications of team training for control room operations?

The training of effective teamwork is critical to the successful management of control room operations. This is particularly true in relation to the conduct of tasks that require effective co-ordination amongst a number of operators. Some of the team training implications which are likely to warrant detailed consideration in relation to control room operations are now considered.

9.1.5.1 Safety

Increasingly, legislation is forcing system designers to consider safety as a fundamental design imperative; indeed, many control rooms are now designed primarily around the management of safety. Safety is considered critical in any system where an unintended sequence of events that causes death, injury, environmental or material damage (MoD, 1996). Whilst safety is clearly a consideration in the system design process, it is also a critical consideration in the design of team training for the operations staff. Reason (1997) suggested that the hazards facing control room operators arise from both technical and human failures. Further, the human contribution to hazards can equally arise from the inevitable errors made by operators at the 'sharp-end' of the system or from latent conditions introduced by system developers. Consequently, for complex systems in dynamic environments, an operator must attend to multiple sources of information, including sensors *and other operators* in order to acquire an awareness of the situation in question. Recent work on 'shared mental models' (Pascual *et al.*, 2000) is beginning to shed light on the information that should be shared amongst team members (together with *how* it should be shared) in order for a team to be effective when solving complex problems requiring high-levels of interaction. This research has spawned training guidelines which are geared towards enhancing the degree of overlap of critical mental concepts shared between team members, relating to both the experiential and dynamic mental models of the task, and also the team undertaking the tasks. These guidelines have been successfully trialed and subsequently adopted by Police armed-response teams (Pascual *et al.*, 1999) and it is suggested that these approaches have high utility for control room settings.

9.1.5.2 Emergency response

Flin (1996: 185) states that:

> the job of an incident command group is to gather and process information in order to aid the commander's decisions to direct, co-ordinate and monitor the actions of the crews engaged in emergency response, damage control, rescue and recovery duties. Therefore, an important facet of the incident commander's remit is to effectively manage the command team, to ensure that they execute their duties, share an understanding of the evolving incident, and are working on the same plan towards common goals.

Once an incident has occurred, critical activities for a control room team include gathering information about the situation; diagnosing what the situation is, employing experientially derived recognitional strategies to identify an appropriate course of action, and then implementing and monitoring the course of action. The activities are all undertaken within the context of applying and adapting organisational norms, procedures and drills. These are activities that clearly require high degrees of team interaction, particularly in the sharing and management of information (both situational and experiential). Also, key to the management of emergency situations, is the explicit role of the leader in managing the team resources most effectively. This can include drawing on individual sources of expertise, briefing the team, ensuring that critical information is shared between relevant team members, tasking team members to conduct response tasks and achieve goals, and managing stress and fear amongst the team. Such activities are critical in the management of emergency response situations, and should thus be specified as key team competencies, in order that appropriate training can be administered.

9.1.5.3 Shift handovers and maintenance of situation awareness

One of the most significant disrupters of situation awareness within teams is the process of shift-handover (i.e. the process of one team handing-on its responsibilities and activities to another team). This process poses many problems for the management of situation awareness. Some of the problems can be reduced by staggering the handover process (in other words, only part of the current team is replaced by incoming staff) as this approach maintains residual situation awareness within the newly formed team. In addition, the briefing process for the incoming team can be enhanced. For example, the incoming team should be made aware not only of what the current situation is, but also how the situation came to be (i.e. how it has changed over time). In addition they should understand where the situation is expected to go in the future (including a specification of the cues which would indicate that the situation is developing as expected, or not – in other words, the situational expectancies). Such processes should be considered explicitly as part of the team training process, and effective briefing processes should thus be specified as a teamwork competency.

9.1.5.4 Increased efficiency (and profit)

Control rooms often provide the venue for controlling and directing the generation of an organisation's profit. For example, the efficiency with which an oil refinery's cracking process is managed impacts directly on wastage, and thus on profitability. Consequently, the effectiveness of a team in managing an objective function can, for some control rooms, form a key training concern. Key to achieving such efficiency is the effective, and explicit, co-ordination of knowledge relating system state to objective function. This should be accompanied by a rigourous understanding and sharing of knowledge about the activities in which other team members are engaged. The fundamental training issues here are similar to those of incident management, and have been addressed by Pascual *et al.* (2000). However, co-ordination against an objective function requires careful consideration to achieve increased vigilance and monitoring process, in order to ensure that tracking effectiveness is maximised, and delays in response to system state changes are minimised.

9.1.5.5 Team self-review and organisational learning

One of the key emergent findings from the shared mental models research cited previously, is the criticality of team self review as an organisational learning process. Pascual *et al.* (2000) studied police armed response teams conducting their debriefing at the end of both computer-based simulation exercises, and high-fidelity field exercises. They identified that the debriefing processes employed tended to focus on the taskwork that the teams had undertaken and largely ignored the teamwork that had taken place. It was consequently noted that on the occasions where challenges to teamwork were observed, such issues might be omitted from the debrief process. Potential issues include: the lack of shared understanding about team goals leading to unnecessary co-ordination overheads; differing situational expectancies resulting in mismatched contingency planning across team members; and the introduction of ambiguity into the messages communicated down the command hierarchy. The nature of the existing debriefs did not easily lend themselves to the identification and discussion of such issues. These findings led to the development of team self review debriefing guidelines (Pascual *et al.*, 1999) that were implemented successfully within the organisation. Such approaches should perhaps play a key role in enhancing the organisational learning which takes place within control rooms, where a focus purely on task conduct might overlook some of the real lessons which could be learned at both a team, and at an organisational level.

9.1.5.6 Multi-agency teamwork

Often, control room teams must act as one component within a much larger, often geographically distributed organisation, comprising many different agencies and associated teams. For example, in an emergency situation a power plant control team may be required to interact with police, ambulance, fire, environmental, media and (potentially) military teams. Each of these teams is likely to have their own procedures and norms, their own terminology, their own shared experiences from which to derive an assessment of the situation, their own equipment, their own communications systems, their own leaders and their own chains of command. Consequently, the power plant team may need to co-ordinate its knowledge, situational understanding and processes with these other organisations. Effective teamwork processes will be highly critical in such circumstances and many, if not all of the teamworking knowledge, skills and attitudes described in Chapter 5 will be key to a successful outcome. The possibility of conducting teamwork within such a multi-agency organisation thus highlights the criticality of identifying, specifying, and training effective teamworking.

Thus far, this chapter has reviewed challenges and approaches to the development of effective teamwork training. The remainder of this chapter provides a similar perspective on the development of effective teamwork technology support.

9.2 Technology for teams

During the last 20 years, the world has witnessed an ever-increasing availability and use of communication and information sharing technologies. The power of these technologies for helping individuals to overcome the barriers of time and space to communicate, co-operate and co-ordinate has strongly come to the fore. Consequently, organisations that rely heavily on effective control room functioning are keen to better understand the functionality and impact of these technologies in the workplace for supporting collocated and distributed teamworking.

Communication and information sharing technologies are typically categorised under the broad banner of 'GroupWare'. The term was originally coined by Johnson-Lenz in the early 1980s (Grudin, 1994) and is commonly defined as the technology (software and hardware) that people use to work together (Baecker, 1993). Technologies often classified under the GroupWare banner include; electronic mail, workflow tools, textual conferencing (synchronous and asynchronous), shared whiteboarding and audio-video conferencing tools (Mills *et al.*, 1999).

The term 'Computer Supported Co-operative Working' (CSCW) is often used synonymously with the term GroupWare. However, many are keen to note that this is done in error (Grudin, 1994). CSCW is a term originally coined by Grief (1988) and specifically refers to the field of scientific enquiry dedicated to investigating (a) how teams work and (b) how technology influences the way teams work. Key questions of interest to the CSCW research community include: What factors influence teamwork? How do you identify the collaborative requirements of teams? How does technology help or hinder teamwork? What is best practice in designing and evaluating GroupWare? To emphasise CSCW as an area of research, many authors are keen to differentiate the term from GroupWare. To illustrate, Traunmueller and Gross (1996) have defined CSCW as a discipline concerned with the study of human behaviour in groups and the technology designed to support it (GroupWare). In short, CSCW is viewed by many as the science that motivates and validates GroupWare design.

Today, CSCW is a thriving research domain and continues to attract the interest of psychologists, sociologists, technologists and business managers. It is envisaged that emergent CSCW research findings will continue to enhance understanding of the relationship between technology and teamwork and improve the suitability and impact of GroupWare in the workplace.

9.2.1 GroupWare design considerations

A number of researchers (e.g. Hill *et al.*, 1994; Mandiviwalla and Olfman, 1994) have suggested that software development approaches have primarily evolved from a focus on eliciting and supporting single user, not team, requirements. Attention has tended to concentrate on the single user operating the system to achieve his or her own specific task goals (the majority of which have been 'technical' or 'data processing' oriented, and *not* 'team' or 'communication' oriented). Typical usability criteria for software design and evaluation focus on factors such as learnability; navigation and control, help functionality (e.g. help facilities, prompts and instructions, *etc.*), performance and effectiveness (e.g. speed, timeliness, accuracy, robustness and reliability, *etc.*) and attractiveness (e.g. colours, shapes, sizes, *etc.*). While these issues remain critical in design, GroupWare technologies require additional attention in a number of other key areas often neglected in design. In a review of GroupWare design and evaluation, Palmen and Ireland (2000) identified nine GroupWare design considerations suggested as being critical for success. It is envisaged that these considerations are also relevant when designing and/or assessing GroupWare for control room environments. In summary, the nine design considerations are:

1. **Public and private information requirements.** The designer needs to identify which data should be public (accessible to all) and which should be private (accessible by a specific user/sub-set of users).

2. **Personalised views vs. consistency.** The designer needs to examine and identify when and where it is permissible for users to customise their interface views to suit their own working styles, and when and where it is better to enforce a consistent view that is similarly used by all.
3. **Replication accuracy and timeliness.** The designer needs to identify the time and accuracy critical requirements for replicated data in the system. When users are drawing from the same centralised data source they typically need to feel confident that the data they are using is current, accurate and similarly shared/ viewed by all.
4. **Floor control.** The designer needs to consider the need for a manager/ chairperson/facilitator role, and the functionality/support from the system this role requires. In GroupWare systems with five or more users there is often the requirement for this role to focus and direct team effort.
5. **Fellow user/team member awareness.** The designer needs to consider the extent to which users require to monitor and have awareness of the presence, availability and activities of other users. Establishing the balance between the benefits of user awareness and the disadvantages of creating a 'big brother' syndrome need to be addressed and investigated.
6. **Temporal and geographical operation.** The designer needs to consider requirements for supporting team interaction across synchronous and asynchronous (time), and distributed and collocated (geographical) contexts. In other words, the designer needs to understand the requirements for information exchange in real time and staggered time environments, as well as considering the extent to which the system will be used to support distributed and collocated working (e.g. same location brainstorming).
7. **Role differentiation and associated requirements.** The designer needs to identify the roles of users within the team context and their associated task, information and communication requirements. More specifically, the designer will need to determine where role requirements overlap and differ, and the extent to which differentiated role requirements can and will be supported. The designer will need to draw a clear boundary specifying what the system is (and is not) being designed to support.
8. **Collective memory.** The designer needs to consider the requirements for storing, indexing and accessing team data within an archive/store. Configuration control and the ability to determine an audit trail for data are likely to be critical factors for the co-ordination and success of larger scale teams (i.e. 10 people or above).
9. **Design sensitive to teamworking culture.** The designer needs to appreciate established ways of teamworking in the organisation. This should be done to ensure that the GroupWare system does not conflict or contradict with established procedural and cultural understandings and methods of working that are widely deemed to be effective.

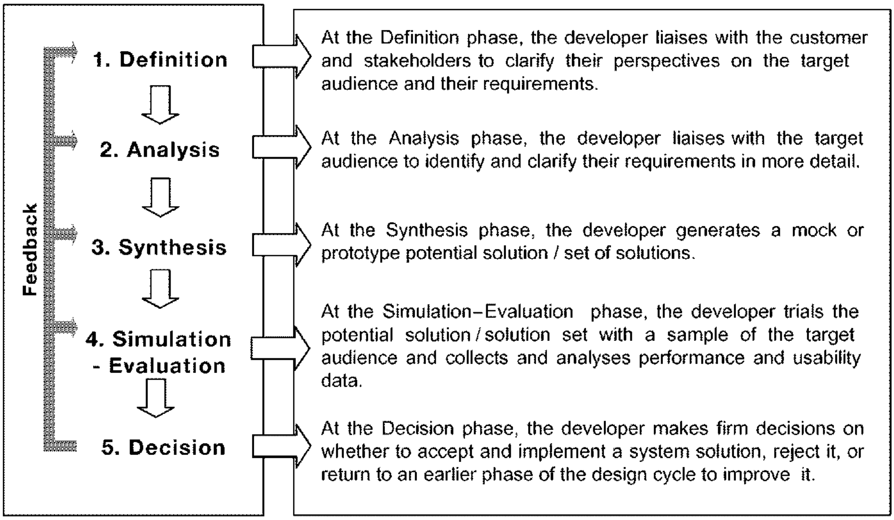
It is widely believed that well-designed and integrated GroupWare systems have much potential for supporting, maintaining and enhancing teamwork both in co-located and distributed settings. Designed with careful attention to collaborative and teamworking requirements, GroupWare can increase the timeliness and quality of information exchange, as well as promote levels of shared understanding across team members. Designed with scant attention to teamworking requirements and driven by unqualified assumptions, GroupWare can, and has been known to, hinder significantly team effectiveness (Henderson, 2000). To illustrate this, Henderson has described a list of recurring problems witnessed when evaluating technological support designed for military command and control teams. Common problems cited include:

- User frustration at feeling 'tethered' to their workstations.
- Users tending to operate digital systems in parallel with the paper systems they are supposed to replace.
- Automated position plotting on digital maps reducing user engagement, and thus situation awareness.
- Users' lack of understanding about where digital information comes from and how it changes over time.
- Users spending their time 'driving' these systems, which significantly reduces the amount of time they spend talking about the problems with which they are dealing.
- Teams having difficulty in brainstorming around a digital support system, i.e. focus being directed towards the technology (typically a screen) rather than each other.
- Teams having difficulties in using digital systems for situation and mission briefings, e.g. when panning, zooming and conducting other system operations.

9.3 Designing for teams: An integrated approach

As illustrated in Chapter 5, team science is currently evolving and continually improving our understanding of team phenomena through enhanced team process models and theories. A number of practitioners are already exploiting such models and theories to derive techniques for designing and evaluating team oriented training and technology. It is envisaged that many theoretical recommendations can both be applied to specific contexts in their own right and be linked to specific development cycles that analysts, developers, trainers, *etc.* are familiar/experienced with using. To exemplify the latter, a generic system's development cycle is presented. This illustrates where and how attention can be focused on specifically identifying and supporting team processes.

Figure 9.5 presents a generic system's development cycle that is characterised by consecutive phases and feedback loops. It is envisaged that this generic system's development cycle can be applied when designing for control room teams. Each phase is now considered in turn with reference to attention on identifying, capturing and supporting team processes.



At the Definition phase, the developer liaises with the customer and stakeholders to clarify their perspectives on the target audience and their requirements.

At the Analysis phase, the developer liaises with the target audience to identify and clarify their requirements in more detail.

At the Synthesis phase, the developer generates a mock or prototype potential solution / set of solutions.

At the Simulation–Evaluation phase, the developer trials the potential solution / solution set with a sample of the target audience and collects and analyses performance and usability data.

At the Decision phase, the developer makes firm decisions on whether to accept and implement a system solution, reject it, or return to an earlier phase of the design cycle to improve it.

Figure 9.5. Generic system's development cycle.

9.3.1 Definition phase

Consultation with customers and stakeholders.

- Here the developer should specifically focus questions/issues on team factors, such as:
  - What are the customers' views on teams to target?
  - Where is the boundary – i.e. which teams/ 'team of teams'?
  - Why should these teams be targeted?
  - What is the teamwork performance gap?
  - How, where and why should teamwork be improved?
- The developer should use theories concerning principles of effective teamwork (as described previously) to focus customers' perspectives on potential teamwork requirements/areas for improvement.

9.3.2 Analysis phase

Identifying the requirements of the target audience/teams.

- Here, the developer should focus attention on eliciting teamwork requirements as well as taskwork requirements (as described previously).

- Again teamwork principles can provide the focus at this more detailed level of investigation, e.g. to address requirements for communication, monitoring, feedback, situation awareness, planning, leadership, *etc.*
- Team Self Review (as described previously) can provide an additional effective mechanism for capturing teamwork requirements (i.e. via guidance on team observation and one-one/team interviews).
- GroupWare design and evaluation guidelines (as described previously) provide a valuable mechanism for assessing the effectiveness of current technology used to support target teams. These sorts of guidelines are particularly valuable when a potential solution under consideration is technology focused.
- If the potential solution is training focused, the developer can use team training guidelines to develop a prototype solution (this phase corresponds with the 'analysis' training phase described previously).
- Although outside the scope of this chapter, the designer may wish to consider issues concerning teamwork workspace ergonomics (see Ives and Widdowson, 1999).

9.3.3 Synthesis phase

Develop a prototype solution/solution set.

- If the potential solution is technology focused, the developer can draw on GroupWare design guidelines and other CSCW research outputs to assist the development of a requirements focused solution.
- If the potential solution is training focused, the developer can use team training guidelines to develop a prototype solution (this phase corresponds with the 'design' training phase described previously).
- Again, the designer may wish to consider issues concerning teamwork and workspace ergonomics (see Ives and Widdowson, 1999).

9.3.4 Simulation-evaluation phase

Trial the mock solution/solution set with a target team/sample of teams.

- Here the developer can use a team scenario to generate and incorporate team stressors (including situation awareness challenges, communication demands, monitoring requirements, *etc.*).
- Team self review can provide the mechanism for capturing and reviewing dynamic teamwork data and evaluating performance.
- If the potential solution is training focused, the developer can use team training guidelines to develop a prototype solution (this phase corresponds with the 'evaluation' training phase described previously).

- Other techniques that can also be used to generate team scenarios and capture process data include teamwork focused Critical Incident Analysis (Klein *et al.*, 1989) and the TARGETS Methodology (Salas *et al.*, 1997).

9.3.5 Decision phase

Build or return to earlier phase.

- Here the developer decides whether to build and implement a solution/solution set or return to an earlier phase of development. If attention has been paid to team factors throughout the development phases, the developer can justify his/her decision with rationale focused on the adequacy of solution support to both taskwork and teamwork processes.

9.4 Conclusions

It is envisaged that training teamwork competence explicitly in conjunction with taskwork competence promotes cohesion and effective co-ordinated response to critical incidents. This chapter has addressed means of identifying teamwork training requirements and subsequently delivering against them. In addition, it has described approaches to evaluating the effectiveness of team training. The implications and potential benefits of team training for control room teams are seen to be many.

In recent years, communication and information sharing technologies have become a key area of interest for a number of organisational science authors who have been using the term ‘network organisations’ to describe visions of effective organisational forms of the future. Network organisations are typically described as flatter and leaner, decentralised, organised around project team or capability areas, flexible and responsive (Symon, 1999). The key drivers that are envisaged as enabling network organisations are ‘effective teamwork’ and ‘GroupWare’. Many hypothesise that GroupWare will support teams by promoting effective information sharing, providing effective communication media, allowing the provision of detailed and timely information, providing access to key sources of information and generally supporting and enhancing team co-operation and co-ordination.

Although GroupWare systems may well possess the functionality to enhance team performance, the GroupWare industry continues to produce these technologies with a limited appreciation of (a) how teams work, and (b) how technology impacts on the way teams work. Academics and practitioners working in the relatively young discipline of CSCW continue to research and debate these issues. Although the domain remains one peppered with more questions than answers, useful outputs for enhancing the design and evaluation of GroupWare technologies are beginning to emerge. For example, although far from the 'definitive' answer, the design considerations described in this chapter provide initial and useful high level guidance and should be exploited by those who are responsible for introducing or designing GroupWare for control room contexts and beyond.

Finally, this chapter has described at a high level how systems developers can begin to think about incorporating practical techniques within their development cycles for targeting, identifying and supporting team requirements. The development cycle described in the chapter was presented as an exemplar and pitched at an intentionally generic level. However, it is envisaged that with a degree of time and effort, developers working to specific systems engineering and development cycles could tailor such to incorporate methods (as described here) to enhance the assessment and support of team, as well as task, requirements.

References

1 BAECKER, R.: 'Readings in groupware and computer supported cooperative work' (Morgan Kaufmann, San Mateo, CA, 1993)

2 CANNON-BOWERS, J. A., SALAS, E., and CONVERSE, S. A.: 'Cognitive psychology and team training: training shared mental models and complex systems', *Human Factors Society Bulletin*, 1990, **33**, pp. 1–4

3 CANNON-BOWERS, J. A., TANNENBAUM, S. I., SALAS, E., and VOLPE, C. E.: 'Defining competencies and establishing team training requirements', *in* GUZZO, R. A., and SALAS, E. (Eds.) 'Team effectiveness and decision-making in organisations' (Jossey-Bass, San Francisco, 1995), pp. 333–80

4 FERNALL, R. P., HENDERSON, S., HUNT, M. B., and PASCUAL, R. G.: 'Measuring and explaining system performance and effectiveness'. Unpublished DERA Report, 1995

5 FLIN, R.: 'Sitting in the hot seat' (Wiley, Chichester, 1996)

6 GIVEN, K. C.: 'The evaluation of training', *in* TELFER, R. (Ed.): 'Aviation instruction and training' (Ashgate, Cambridge, 1993)

7 GRIEF, I.: 'Computer-supported cooperative work: a book of readings' (Morgan Kaufman, San Mateo, CA, 1988)

8 GRUDIN, J.: 'CSCW: history and focus', *IEEE Computer*, 1994, **27**, (5), pp. 19–26

9 HENDERSON, S.: 'Human factors pitfalls in the automation of command', *Journal of Defence Science*, 2000, **5**, (3), pp. 350–6

10 HILL, R. D., BRINCK, T., ROHALL, S. L., PATTERSON, J. F., and WILNER, W.: 'The rendezvous architecture and language for constructing multi-user applications', *ACM Transactions on Computer-human Interaction*, 1994, **1**, (2), pp. 81–125

11 IVES, C. J., and WIDDOWSON, A. J.: 'Command teams and technologies'. Unpublished DERA Report, 1999

12 KIRKPATRICK, D. L.: 'Evaluation', *in* CRAIG, R. D. (Ed.): 'Training and development handbook' (McGraw-Hill, New York, 1976) pp. 301–19

13 KLEIN, G. A., CALDERWOOD, R., and MACGREGOR, D.: 'Critical decision method for eliciting knowledge', *IEEE Transactions, Systems, Man and Cybernetics*, **19**, (3), pp. 462–79

14 MANDIVIWALLA, M., and OLFMAN, L.: 'What do groups need? A proposed set of generic groupware requirements', *ACM Transactions on Computer-human Interaction*, 1994, **1**, (3), pp. 245–68

15 MATHIEU, J. E., and MARTINEUA, J. W.: 'Individual and situational influences on training motivation', *in* FORD, J. K. *et al.* (Eds.): 'Improving training effectiveness in work organisations' (Lawrence Erlbaum Associates Inc., New Jersey, 1997) pp. 193–221

16 MCINTYRE, R. M., and SALAS, E.: 'Measuring and managing for team performance: lessons from complex environments', *in* GUZZO, R. A., and SALAS, E. (Eds.): 'Team effectiveness and decision-making in organisations' (Jossey-Bass, San Francisco, 1995) pp. 9–45

17 MILLS, M. C., BLENDELL, C., HENDERSON, S. M., and RODDEN, T.: 'Communication technologies and PJHQ planning: human factors issues for consideration'. Unpublished DERA CHS Report, 1999

18 MOD.: 'Safety management requirements for defence systems'. Defence Standard 00–56 (Part 1)/Issue 2, December, 1996

19 NAWC-TSD: 'Team training series'. NAWC (Naval Air Warfare Centre), Training Systems Division, 1997

20 PALMEN, H., and IRELAND, J.: 'Human factors influencing distributed planning'. Unpublished DERA Report, 2000

21 PASCUAL, R. G., MILLS, M. C., BLENDELL, C., and MOLLOY, J. J.: 'Understanding and supporting team cognition; enhancing teamwork through team self-review'. Unpublished DERA Report, 1999

22 PASCUAL, R. G., MILLS, M. C., BLENDELL, C., and MOLLOY, J. J.: 'The role of mental models in team effectiveness: final report'. Unpublished DERA Report, 2000

23 REASON, J.: 'Managing the risks of organizational accidents' (Ashgate, Gower, 1997)

24 RUGG-GUNN, M., CUNNINGHAM, T., GRIMSHAW, T., and LAWRENCE, S.: 'Team TNA methods for the procurement and selection of team trainers'. Unpublished DERA Report, November, 1999

25 SAARI, L. M., JOHNSON, T. R., MCLAUGHLIN, S. D., and ZIMMERLE, D. M.: 'A survey of management training and education practices in U.S. companies', *Personnel Psychology*, 1988, **41**, pp. 731–43

26 SALAS, E., CANNON-BOWERS, J. A., and KOZLOWSKI, S. W. J.: 'The science and practice of training – current trends and emerging themes', *in* FORD, J. K. *et al.* (Eds.): 'Improving training effectiveness in work organisations' (Lawrence Erlbaum Associates Inc, New Jersey, 1997), pp. 357–67

27 SYMON, G.: 'New communication technologies and the network organisation'. Paper presented at the Annual Occupational Psychology Conference, 1999

28 TAYLOR, F. W.: 'The principles of scientific management' (Harper and Row, New York, 1911)

29 TRAUNMUELLER, R., and GROSS, T.: 'CSCW, Groupware und Workflow', Leitbilder der Verwaltung praegen die Entwicklung. *Verwaltung & Management*, 1996, **2**, (4), pp. 202–8

Chapter 10

Naturalistic analysis of control room activities

Paul Luff and Christian Heath

10.1 Introduction

It would seem relevant and appropriate to observe the activities of personnel in control rooms prior to designing new technologies to be placed in them, altering the configuration of equipment and layout, or changing the assignment of staff responsibilities and tasks. Numerous approaches have been outlined for organising and shaping these observations, most noticeably task analysis (see Kirwan and Ainsworth, 1992). These can support the analyst to delineate the tasks and functions of individuals, to examine the role of new technologies in the accomplishment of activities and to identify what information is available, where, and how this information should be communicated. Typically, these approaches focus on the formal tasks and activities, or at least, what can be made explicit. In this chapter the tacit and social practices on which control room work relies will be considered. Observations of work and interaction in one particular setting, station control rooms in London Underground, will be reported. This contributes to the growing number of naturalistic studies of control rooms and similar settings. The details of the practices revealed by the analysis could contribute to the development of technologies for the control room and more generic kinds of systems. In particular, we seek to re-examine how individuals ‘monitor’ the surrounding domain and the activities of their colleagues. We look at how the personnel utilise the available technologies that are a combination of audio, visual and computational systems and how, in concert with colleagues, they monitor the happenings in the world beyond the control room. Finally, the ways in which such an analysis can provide insights into how prototypes technologies could best be deployed in organisational settings will be examined. The chapter will conclude by briefly outlining some recent technological developments that aim to support the work of control room personnel.

10.2 Background

Recently, studies of work in control rooms for such diverse domains as transportation control, telecommunications management and power plants have highlighted how activities in these settings are reliant on a range of tacit social practices (e.g. Harper *et al.*, 1989; Heath and Luff, 1991; Suchman, 1993; Goodwin and Goodwin, 1996). Of particular interest to developers of advanced technologies has been conduct characterised as 'monitoring' or as 'peripheral awareness' where individuals appear to be sensitive to other activities and events whilst often engaged in other tasks. These studies have suggested requirements for collaborative technologies like media spaces and collaborative virtual environments, particularly how those technologies could provide resources for supporting collaboration whilst not intruding on the ongoing activities of participants (Bly *et al.*, 1992; Dourish and Bly, 1992; Gaver *et al.*, 1992; Benford and Fahlén, 1993). However, it is unclear how such unobtrusive support can be provided (e.g. Heath and Luff, 1992b; Heath *et al.*, 1995; Hindmarsh *et al.*, 1998). Characterisations of the practices can appear vague and ambiguous, 'monitoring' and other similar concepts like situation awareness being typified as general, unmotivated and usually passive. Moreover, associated with these terms there tends to be an emphasis on supporting the 'peripheral' and spatial nature of co-participation, glossing over how such practices are embedded within the work and interaction of the participants.

In this chapter we examine a setting which *prima facie* requires the individuals within it to be aware of events occurring around it. The control (operations or 'ops') rooms of stations on the London Underground provide a range of technologies through which the operators (typically station supervisors) can survey the local setting. Drawing on audio-visual recordings made in the operation rooms and extensive fieldwork, we suggest how such practices are organised with respect to the organisational responsibilities of the staff. In addition, we will consider how a practical geography of the events that can occur and the sequential import of happenings which are seen through the technology are oriented to. In particular, we examine the work of supervisors who, amongst other things, monitor stations on the London Underground through several digital systems, including a bank of closed circuit television (CCTV) monitors. The world they are presented with is not complete, it is disjointed and fragmented, transformed by the technology available to them. The supervisors have to make use of what is available to make sense of the remote environment and then act accordingly. Despite the almost pervasive nature of these surveillance technologies, and the substantial literature on the perceptual skills and cognitive processes required to scan complex, dynamic images (e.g. Monty and Senders, 1976; Senders *et al.*, 1978; Fisher *et al.*, 1981; Rensink *et al.*, 1997), there appears to be few studies of how individuals 'interact' with and through these systems in real-world environments.

The analysis in this chapter draws upon an orientation informed by ethnomethodology and interaction analysis (Garfinkel, 1967; Sacks, 1992), and discussed and developed through a range of studies of social interaction, institutional conduct and workplaces studies (e.g. Goodwin, 1981; Heath, 1986; Drew and Heritage, 1992; Luff *et al.*, 2000). There are also now a number of texts which discuss these analytic orientations, their conceptual underpinnings and the methodological resources they draw upon (e.g. Heritage, 1984; Heath and Luff, 1992a; Lynch, 1993). It should be noted, however, that this body of work adopts a similar orientation to considering a range of diverse materials rather than explicitly adopting a common 'method' or 'framework' for analysis. This orientation is best exemplified by the case studies of everyday talk and social action in a range of domains, including ordinary conversation, institutional interaction and collaborative activities in the workplace (e.g. Atkinson and Heritage, 1984; Drew and Heritage, 1992; Heath and Luff, 2000). These studies are typically qualitative, involve detailed analysis of audio or visual materials and can draw upon accompanying fieldwork. So in the case here the materials are drawn from an ongoing study of stations which involves in-depth fieldwork and audio-visual recordings undertaken in the major stations of London Underground (known as 'Section 12' stations). The total corpus of video recordings consists, at present, of over 300 hours of material. The analysis emerges iteratively, gradually developing analytic themes and issues through viewings of the recordings and consideration of the fieldwork observations. It is therefore distinctive from studies that draw from a pre-specified set of analytic categories, either to provide an initial quantification for the analysis (cf. analyses of the number and direction of eye movements), or to circumscribe the ways actions and activities are characterised (cf. a hierarchy or structure of goals, operators or plans).

10.3 Tracking events through a fragmented environment

The station supervisors who work within station control rooms (or 'ops rooms') of London Underground stations have a wide range of responsibilities concerning the safe and smooth running of the station. They have to cope with a variety of routine problems and complaints from passengers, ranging from ticket and travel queries to dealing with missing persons and property. They have to manage the moment-to-moment operations of station staff on the gates and platforms, informing them of delays on the line and any critical incidents that have occurred on the network and managing their disposition to cope with emerging problems on the station. The supervisors have to deal with critical incidents which may occur, including extreme station or platform congestion, violent and problematic behaviour from passengers and evacuations due to fire and suspect package alarms. They also have to manage a wide range of more mundane concerns such as staff rosters, the state and maintenance of equipment on the station, and coping with buskers, beggars, pickpockets, fare dodgers and smokers.

To assist them they have a wide range of communication and information technologies (see Figure 10.1). These include radio, passenger announcement systems, train information, emergency control and alarm systems and various phone systems. However, in the main ops rooms, perhaps the most notable technology is the bank of between 4 and 12 CCTV displays, and these provide the focus in this chapter.

The CCTV monitors display images from a large number of cameras, (e.g. Leicester Square has 80), situated around the station. There are typically cameras located in all public areas, at various places along the platforms, along corridors and passageways, above stairs and escalators and in the various circulating areas of the stations. The bank of CCTV screens displays several kinds of image: there are a number of fixed displays of platforms, there is usually a display that switches automatically through all the cameras, and one or two other displays can be selected by the operator. The operator can exploit these different camera positions to track passengers that are considered to be of concern. On seeing an individual of interest from one camera the station supervisor may, after a few moments, select a camera further along the individual's prospective route through the station. So, in one case several seconds after seeing a passenger entering the gates in the foyer of the station, the supervisor selects an image of an underground corridor through which the same passenger then passes, onto another showing the passenger moving from the end of the corridor to another passageway and then one of the platform, where the passenger is seen to enter, walk along to a seat and then sit on a bench.

The station supervisor utilises the technology to accomplish a smooth tracking from image to image. The cameras are identified by a unique number and the supervisor selects from the 80 or so which are available, the one which will be the next most appropriate image. Amongst these cameras are several pan-tilt-zoom, or 'Omniscan' cameras which can be manipulated to move horizontally, vertically and provide close ups. In the above case, the supervisor uses these cameras to follow

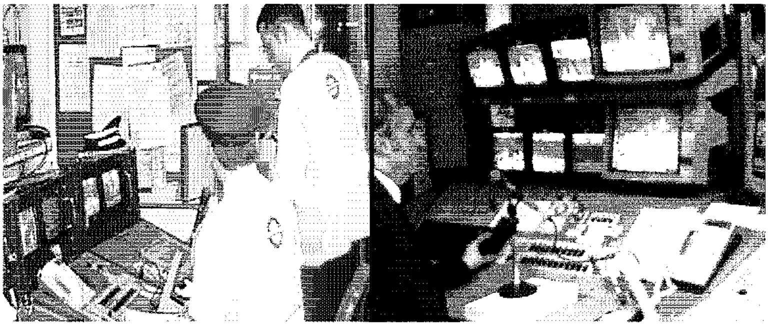


Figure 10.1. The technologies in Piccadilly Circus (left) and Leicester Square (right) station operation rooms. As well as CCTV monitors, visible amongst the artefacts in these settings are textual displays of trains on the line, control panels for the CCTV systems, the passenger announcement (PA) system including microphones for the PA and radio system for staff announcements.

the passenger from the end of one corridor to the beginning of the next and along the platform – the places from where it is least clear where the passenger will go next. The station supervisor can thus use the technology to anticipate possible future locations of passengers, their trajectories through the station. Moreover, in the course of this tracking they can also select the most appropriate way of observing the scene, whether this is the selection of the image, the display on which it is presented or choosing the devices that will allow them to track passenger behaviour when this is most ambivalent. Using the Ominscan can therefore not only offer a means of tracking a passenger as they move through a particular area, but can be selected prospectively for those locations where passengers can take different paths and trajectories. Indeed, through the various distinct images available to them and the different devices through which these are presented the station supervisors manage to produce a coherent sequence of views to follow incidents and track problems around the station as they emerge.

Throughout the day station supervisors can be seen occasionally to undertake such switching of cameras to monitor scenes and individuals. However, their active 'use' of the technology more typically arises due to some incident emerging on the station. So, in the following case a station assistant (SA) comes into the control room to inform the supervisor (SS) of complaints about a passenger with 'some megaphone'.

Fragment 1 [LSQ 15/8/96 12:24] Transcript 1

SA Mr (Surname)
(0.4)
SS Yes mate
(0.7)
SA There's a passenger just complained. There's somebody down the sou- platform downstairs, southbound platform. He got some megaphone like, (running up) to people and (blowing) it at them and (things) frightening them and like.

When the SA says 'He got some megaphone like', the gist of his utterance, the supervisor turns from the large CCTV display to look at a smaller monitor; the display showing the image of the busy southbound Northern Line platform. Some moments later he calls on the radio, another station supervisor (the mobile station supervisor, MSS) who is out and about on the station:

Fragment 1 [LSQ 15/8/96 12:24] Transcript 2

SS Yeah on the southbound northern there's a bloke with a megaphone, we've had complaints, he's setting his megaphone off and frightening passengers. Can you have a word with him please
MSS Yeah I am on my way

Some three minutes later, as the mobile station supervisor can be seen on the CCTV images to arrive at one end of the crowded southbound platform, the SS calls him on the radio and guides him to the problem.

Fragment 1 [LSQ 15/8/96 12:24] Transcript 3

SS Yes he's at the north end of the platform I can see him now he has got a black coat on, right at the far end (1.0) he's got a (mega-) here you are (.) he is dashing onto the train now (3.5) (jumping about)

The SS, therefore, using the CCTV system not only manages to locate the troublesome passenger on the platform, but also to guide his colleague to the location of the problem. This is not straightforward. In the time taken for the mobile station supervisor to get down to the platform several 'waves' of passengers have accumulated on the platform and then dispersed on the trains. At times the platform appears full of indistinguishable bodies. Moreover, the troublesome passenger can be seen to move away, along other passageways and onto another platform, before returning. The supervisor switches between various cameras, in particular using the Omniscan on the platform, moving it up, down and zooming in on the crowded platform. Even as the supervisor is describing features of the passenger and his location, the passenger appears to jump aboard a train. Through a crowd of people trying to board the train and from some distance away, the mobile supervisor manages to locate the passenger and some nine seconds later the mobile radios in for the supervisor to get assistance from the British Transport Police (BTP) to deal with the passenger.

Although having individuals bellow instructions through megaphones to their fellow passengers may not be a routine occurrence on the London Underground, such cases reveal some of the routine ways operators use the video technology to accomplish their practical tasks and activities. In order to deal with a wide variety of incidents, staff not only choose from a large number of possible camera angles, but choose those cameras that will best allow them to undertake future activities. As we see in the case above, the use of video technologies used in concert with other devices, like radio, form a critical resource for collaborating with colleagues and developing a co-ordinated response to an emerging incident.

Needless to say the use of the technology relies on a detailed understanding of the domain in question, not only the names of locations and the corresponding numbers of the cameras, but also the typical locations where incidents occur and the ways in which passengers move around the scene. However, this is not just an awareness of the fixed layout of the domain it requires control room personnel to be continually engaged in the emerging activities occurring in the domain.

It may seem unusual that the station supervisor seems to notice the passenger so quickly amongst all the people on the platform. However, as part of their work in front of the displays, station supervisors appear to be particularly sensitive to unusual passengers and odd passenger behaviour. They frequently note to colleagues or visitors, passengers seen on the screen or through the window who are of a strange disposition, ones who are over- or under-dressed for the time of year, or strangely attired for the local area. They can note people carrying large or unusual packages, people trying to get bicycles or push-chairs through the underground or individuals who are using sticks to walk. They also note passengers who may dawdle on platforms after several trains have been through, who stand at the extreme end of a platform, or who stand too close to the entrance of the tunnel. They note these not because they have some prurient interest in the curious and strange, but that such appearances can have potential organisational significance for them and the running of the station. Such fleeting images can suggest potential problematic activity. For example, individuals carrying large amounts of baggage up flights of stairs often leave some behind at the bottom – which can in turn invoke sightings by other passengers of 'suspect packages'. Similarly, bicycles and other awkward items can cause congestion in passageways and platforms. People hanging around platforms may be potential suicides or accidents. Hence, 'keeping an eye' on potentially problematic passengers from the normal ebb and flow may help deal with troubles later.

Some of these practices rely on noticing the 'strange' from apparently 'normal appearances' (cf. Sacks (1972)). Control room staff practically discriminate apparently routine behaviour for organisational purposes. For example, two people, even within a crowd, who walk noticeably close to each other up towards a gate can be seen as potential 'doublers' – passengers who try and squeeze in two at a time through a gate. Or, passengers who are noticeably not looking where others are looking, particularly if they are looking down at the midriff of other passengers, can appear as potential pickpockets.

Station supervisors and other control room personnel, then, do not passively monitor the displays waiting for 'something to happen' or just glance at the screens to gain a general 'awareness' of the activities occurring within the station. They are actively undertaking analyses of the conduct they see on the screens and through the window of the control room. These may be tentative and highly contingent accounts of what they see, but they are critical for making sense of the complex behaviour and images made available through the technology. To make temporary, fragmentary and disjoint views of the world coherent and useful, relies on routine categorisations and discriminations made with respect to organisational purposes, of the human conduct they see before them.

What perhaps is remarkable about these sightings is that the images are not always very clear. The very environment of the London Underground can mean images are poorly lit, camera lenses and screens dirty and monitors kept on the same view for hours on end can be 'burnt out'. Hence it can be hard to distinguish even gross features, e.g. it can be hard to tell whether a crowd is stationary or moving – a feature which can only be distinguished by whether the heads of the passengers are bobbing or not. Moreover, even with a large number of cameras there are gaps in coverage. The larger stations on the London Underground are very complex environments. Over a hundred years of gradual development, extension and transformation has meant that these stations are built on many levels with irregularly shaped passageways, corridors and stairways interconnecting them. This means that there are blind spots and discontinuities between camera images. A corridor, passageway or stairwell may only be covered by one camera making the area directly beneath it invisible. There may be 'nooks and crannies' which are similarly inaccessible, even for Omniscans. One of the skills of the station supervisor is knowing where these blind spots are and dealing with them. So tracking a passenger requires not only coping with the different locations and various kinds of technology, but also the potential problems that arise when using it. Station supervisors develop a body of practices, ways of seeing the world before them on blurred and flickering images that make features of that screen organisationally accountable.

10.4 The world beyond the image

As mentioned, amongst the problems supervisors face in dealing with the fragmentary images are the blind spots which are inevitable in the video coverage of a large and complex station. Many of the blind spots occur just where problems emerge. So, for example, it is often difficult to get full coverage right to the ends of the platform; however, this is the place where the train enters the station and where particular incidents, like passengers falling under a train, typically occur.

Certain regular visitors also become aware of the blind spots. Beggars and buskers will frequently place themselves in locations where they are not directly in view of a camera. To cope with this the supervisors again draw on routine ways of surveying the scene. Particular areas of the station, at junctions, at the top of stairs, at locations down a corridor, for example, are popular not only because they are invisible to CCTV, but because they are on the regular routes of passengers. The supervisors are aware of these locations and can utilise passengers' behaviour in these locales to infer what may be happening in the invisible world, the world beyond the image. So, for example, they can note 'unusual' swerves in the movements of passengers as they avoid the individuals out of sight. They can also utilise passengers' own lookings 'out of the scene' to infer that there might be something further to investigate.

The world available on video therefore is not restricted to the domain circumscribed by the images. Personnel can use what they see as a way of determining activities and events beyond the image. Such inferences rely upon their abilities to recognise particular patterns of conduct and the ways in which they are embedded in particular activities; activities which may not themselves be fully available on the screen.

Whilst personnel in ops rooms select particular views, e.g. to take a closer look at a particular scene or to deal with a particular incident, they tend to configure the views of the domain in routine and regular ways. The provision, and the selection of these views, is organised with regard to their responsibilities and the activities in which they engage. There is a geography to activities and events in the domain, and surveillance is organised with regard to what happens, or might happen, in particular locales. So, for example, the line of gates is considered an area where violent incidents can occur, particularly disputes concerning the validity of tickets. Maintaining a view of these not only allows personnel the possibility of seeing such incidents, but can facilitate staff to make recordings as a problematic encounter develops. A geography of events allows personnel to view particular scenes with regard to specific types of problem, or even certain types of trouble maker.

Individuals who use surveillance equipment, therefore, are themselves sensitive to their views on the world beyond, and orient to the relationship between their view and the domain itself. They know their views, the limitations and possibilities, by virtue of their familiarity with the 'actual' world beyond. Their sensitivity to the constraints of the views feature in how they perceive and handle events, and the ways in which they need to collaborate with others. Ops room and other staff organise their surveillance activities with respect to a geography of events that occurs at particular stations at certain times. This may be in terms of specific areas of the station, different camera views and popular locations for various kinds of problems. Supervisors may configure the technology with respect to these locales, selecting images of places where serious incidents may occur in the future or which provide access, in some way, to a world beyond the image. But these are not idiomatic selections, they are chosen with regard to the organisational concerns and practical problems that staff may face. Personnel build up a practical geography of the domain; a geography that is developed by managing the moment-to-moment difficulties that arise within the setting, a geography that shapes the configuration and use of the surveillance technology.

10.5 Seeing and sequentiality

Switching cameras and finding an identifiably problematic individual, object or event through the use of the technology will typically be followed by some organisationally appropriate next activity, e.g. getting a colleague to investigate further, escorting them off the premises or calling the police. Looking and seeing an incident, object or individual usually serves to engender a particular activity; an activity through which the supervisor accountably fulfils one of his organisational responsibilities. The CCTV displays provide personnel with access to the scene, but as we have begun to see, their so-called 'monitoring or awareness' does not involve some blanket or stable sensitivity to the scene. Rather, their perception of the scenes, their occasioned meanings and intelligibility, is organised with regard to a limited set of interests and concerns, which derive not simply from their organisational responsibilities, but the routine actions and activities in which they engage. Their perception of the scene, its discrimination, is embedded and inseparable from particular activities; activities which are sequentially relevant by virtue of the occurrence of particular events. Prospective and sequentially relevant actions inform the very ways in which personnel discriminate and make sense of the scene.

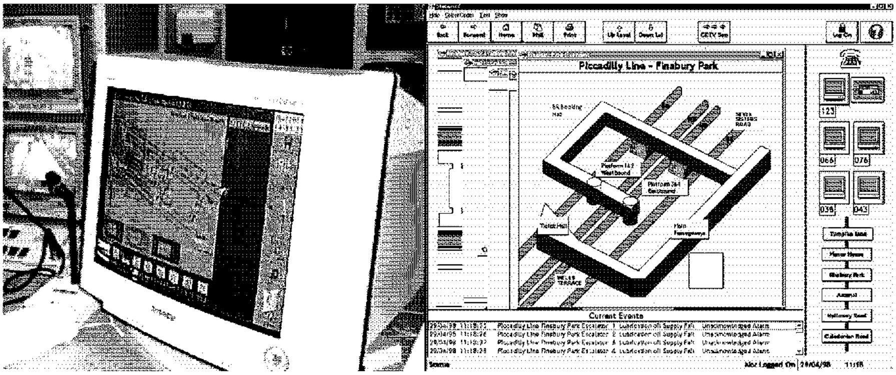
Specific courses of action therefore are linked to, and dependent upon, particular events. Becoming a competent station supervisor involves learning to undertake a range of actions with regard to the occurrence of particular events, and learning how to recognise, identify and categorise these. It is not simply a matter of contrasting some event or appearance with what might ordinarily be expected to happen, but rather learning to see particular events, activities, people or objects in particular ways. Pickpockets, buskers, fare dodgers and the like behave in routine and organised ways, and it is learning to become familiar with their practices, their favoured locales and the methods for dealing with them, which allows such people to be spotted and managed. All sorts of other matters, that to lay people would appear curious, even suspect, pass unnoticed by personnel. It is not however, simply the application of a generalised set of categories and procedures which informs how the scenes are discriminated and managed, but rather the specifics of particular events at the moments in which they are constituted as relevant by staff.

As the brief examples given here suggest, monitoring and surveying the scenes in the station control rooms of London Underground are neither passive nor unmotivated activities. Rather than some general activity through which staff, using individual perceptive skills and reasoning, make sense of the scenes, their surveillance, even if undertaken whilst engaged within some other activity, is thoroughly immersed within the organisational contexture of the setting. The control room personnel are not just trying to maintain 'awareness' of the happenings around the station, their monitoring of the screens is tied to the local contingencies, the moment-to-moment demands and sequential import of their everyday activities. The very seeing, the surveillance of the scene, its meaning and intelligibility, derives from and is produced through the individual's orientation to the sequential relationship between particular events and activities. This interdependence between the perception of the scene, events and the activities of personnel forms the foundation to the staff's surveillance and their highly organised discrimination of the views before them.

10.6 General and design issues

What may be noticeable in the few examples given in this chapter is how station supervisors manage a range of heterogeneous technologies to co-ordinate their response to events and incidents occurring on stations. As well as the CCTV, radio, alarm, PA and phone systems utilised in the examples, other technologies provide resources of relevance. These include systems for monitoring the gates, the ticket machines and the traffic on the line. In concert with the CCTV system these can, for example, provide accounts of congestion on particular escalators, platforms and public areas, help in responding to passenger queries and dealing with a range of other problems which are the responsibility of staff in station control rooms. As in most other organisational settings, station control rooms are the site of various developments aimed at providing a more integrated technological infrastructure. Rather than the collection of different systems they typically have available at present, integrated station management systems are beginning to be deployed that provide access to a range of functionality through one system (see Figure 10.2). Hence, the various controls for passenger announcements, alarms and CCTV are being replaced with personal computers and touch-screen displays to provide access to all these resources.

These systems no doubt have technical advantages for maintaining the system. At the very least they no longer require different controls, connections and information handling for each device. In the future, it should be possible to extend the system more easily, integrate different information sources and make relevant station information accessible in remote locations. Station management systems may also be easier to use for newcomers to the settings. For example, for the CCTV system the operator no longer needs to select cameras by number, but rather can touch icons on graphical maps of the stations displayed on the screen. However, it appears that the developers of these systems, whilst concentrating on the technical



29/04/98 11:18:21 Piccadilly Line Finsbury Park Escalator 1 Lubrication oil Supply Fail Unacknowledged Alarm
29/04/98 11:18:26 Piccadilly Line Finsbury Park Escalator 2 Lubrication oil Supply Fail Unacknowledged Alarm
29/04/98 11:18:32 Piccadilly Line Finsbury Park Escalator 3 Lubrication oil Supply Fail Unacknowledged Alarm
29/04/98 11:18:28 Piccadilly Line Finsbury Park Escalator 4 Lubrication oil Supply Fail Unacknowledged Alarm

Figure 10.2 *Recent integrated station management systems (the GTIE Optech on the left, the ASCOM system on the right). These systems provide a range of capabilities for CCTV selection, alarm control and monitoring and passenger announcements through a graphical user interface. The principal display is typically a graphical representation of the station; subsequent operations are performed on other screens and through intermediate menus.*

characteristics of the systems, have missed an opportunity to take into account how such technologies could be used within the settings in which they are deployed. So, in dividing the capabilities of the system so that only one kind of operation can be performed at a time, e.g. making an announcement, selecting a camera or checking the status of an alarm, the systems no longer allows such activities to be carried out so easily in parallel. Whilst they are preparing an announcement, supervisors can no longer so easily track an individual from image to image, or follow the people towards which the announcement is to be directed. Moreover, introducing a computer system to mediate the operation of the technology adds a further domain for the users to manage. Rather than selecting physical keys and controls on a panel in front of them, users of these systems have to turn to another monitor and select representations of camera locations to display on the bank of the monitors. Integrating a range of diverse resources into a single system, even if mediated through a direct manipulation graphical interface, may undermine the abilities of users to juxtapose and utilise these resources in parallel. On the other hand, it introduces another locale which further fragments the domains the users have to manage (cf. Greatbatch *et al.*, 1993).

With the new technology, operators still have to deal with surveying a domain that is made accessible to them through a set of disjoint images. The practices we have outlined in this chapter reveal how, despite this, supervisors draw on a range of resources in order to gain a wider access to the domain around them and get the technology to work to meet the organisational demands of the domain. It may be that in future developments the new technologies could be shaped to help support the everyday practices of the operators. For example, the technology could be configured to facilitate the selection of sequences of images, so that incidents can be tracked and monitored or, given the organisational consequences of particular viewings, support the invocation of appropriate 'next actions'.

It may be that more sophisticated technological support could be provided that is sensitive to the demands of the setting and the activities of the participants could be provided in a setting such as surveillance operations centres. Such support is already being considered, although some of the ways in which this is being approached has been more in terms of automating surveillance than supporting awareness (Norris *et al.*, 1998). The techniques being utilised have been developed from analysis of simpler CCTV images, such as of moving road vehicles, motivated by a concern that only a small proportion of the available images are being monitored at any time. The algorithms can detect certain attributes of crowds, e.g. whether they are static or in motion, and even track particular individuals within a scene (e.g. Velastin *et al.*, 1993; 1994; Indyk and Velastin, 1994; Davies *et al.*, 1995; Velastin, 1995; Marana *et al.*, 1998). However, it is becoming increasingly recognised that the identification of incidents is not simply a matter of detecting particular patterns in the image. The events on the screen have to be made sense of with respect to activities that are not visible and with regard to the organisational demands of the settings. Thus, technologies are now being considered in terms of supporting surveillance operators rather than automation. It may be possible to inform such technological support from studies such as the one reported here. This is the objective of a recent project called PRISMATICA (GRD 2000–2001) which aims to draw from studies of surveillance operators in various CCTV operations centres in Europe to develop algorithms, interfaces and systems to support staff in these domains. Findings such as those above can, for example, suggest individual objects and features for the algorithms to detect, combinations of features to identify and trajectories of actions to track, preferences for the analysis of particular images at different times and locations, ways in which the analyses can be presented to staff; and how the support could be integrated with other technologies and prospective activities. In collaboration with colleagues at King's College, we are considering how image recognition systems algorithms can identify lone passengers who stand at the end of a platforms for a 'significant' period (potential 'one unders'), distinguish different types of crowd congestion with respect to other occurrences on the station and network (e.g. gate closures, escalator breakdowns and train delays) and how to display the results of the image analysis that facilitates how they oversee the domain.

Rather than attempting to automate the activities of surveillance operators, such a focus seeks to support the ways in which they monitor and maintain awareness of a complex setting. It should be noted that the utility of such capabilities would rest on being tied to the everyday tasks, responsibilities and concerns of the staff. They manage to make sense of the complex space presented through the technologies with respect to the moment-to-moment organisational demands placed upon them. These provide resources to allow them to make sense of particular happenings on the screens and also for these happenings to inform their developing courses of activities in the domain. The video-recordings of the setting, such as those presented here, accompanied by analysis, may provide a valuable resource in maintaining this sensitivity to the conduct of participants in the domain. This may be not only when considering the requirements for such systems, but throughout the development process. This suggests a novel way of providing access into the domain for designers as well as preserving a trace back to the requirements as the system design emerges (cf. Jirotka *et al.*, 1993).

By drawing on the materials we can identify particular organisational demands placed on users of technology, the importance and priorities for identifying particular incidents, for example, as well as those occasions and circumstances when these may not be seen as consequential. The analysis can also suggest ways of integrating, or relating, the recognition algorithms with particular events signalled by other technologies, as well as how the technology could be configured to be used alongside other devices, particularly those used for communicating to staff not inside the control room. Most algorithmic surveillance systems have focused on identifying events and objects from single, fixed images. The analysis above reveals ways in which staff draw upon images from many cameras; pan, tilt and zoom in and out of the views; and juxtapose these images with other resources from the domain. Maintaining awareness of the domain, and events and incidents occurring within it, relies on being able to draw together resources from fragmented views, from distinct resources and by being able to transform what is being viewed in the light of rapid, but tentative, analyses of the scene. Although initial identification of problematic occurrences may come from a simple glance at a monitor, these analyses can be stored and then manipulated. They can also be shaped by the juxtaposition with other views and resources and through transforming the scene that is available. The awareness of events and activities is constituted by staff through their use and integration of these diverse resources. If algorithmic image recognition systems are to offer more than the generation of numerous, if ranked, candidate incidents, they will need to be sensitive to these practices.

Algorithmic surveillance systems is one technology which has been considered as offering support for situation awareness. These technologies are diverse in nature, although typically focusing on the capabilities required for real-time monitoring of complex and dynamic environments. However, even in these systems, as in most current CSCW systems, awareness is typically considered as a general, blanket activity which can be supported merely by providing wider access to a large number of individuals, spaces and locations. It may be that we need to consider more focused support; support that is tied to the particular organisational and collaborative demands of the practices of the participants. Drawing on the sequential and interactional practices, of the kind considered here, may then provide quite a distinctive resource for examining how technologies can support awareness in innovative digital environments.

10.7 Conclusions

Recent studies of work in 'centres of co-ordination' in such diverse domains as financial trading rooms, newsrooms, transportation control, telecommunications management and power plants have highlighted how activities in these settings are reliant on a range of tacit, social practices (e.g. Harper *et al.*, 1989; Suchman, 1993). Of particular interest to developers of advanced technologies has been conduct that has been characterised as 'monitoring' or as 'peripheral awareness' where individuals appear sensitive to activities and events whilst often engaged in other tasks. Such studies have suggested requirements for collaborative technologies like media spaces, collaborative virtual environments and others that are meant to support the provision of information in digital environments. In particular, it is critical how those technologies could provide resources for supporting collaboration whilst not intruding upon the ongoing activities of participants (e.g. Dourish and Bly, 1992; Benford and Fahlén, 1993). However, it is unclear how such unobtrusive support can be provided (e.g. Heath *et al.*, 1995; Hindmarsh *et al.*, 1998). Characterisations of the practices can appear vague and ambiguous, 'monitoring' and other cognate terms being typified as general, unmotivated and usually passive. Moreover, there tends to be an emphasis on supporting the 'peripheral' and spatial nature of co-participation which glosses over how such practices are thoroughly embedded within the work and interaction of the participants.

Consideration of what may seem a highly distinctive setting with specialised technology may have more general consequences for studies of how individuals 'interact' with technology. With respect to many systems currently proposed within human-computer interaction and CSCW, the particular systems available to the operators, and the devices through which they interact with them, would appear quite straightforward. However, in this chapter we have indicated how the use of such systems is thoroughly embedded within the many disparate activities of the personnel. Individuals draw on a body of common-sense resources to deal with the technology they are operating. They utilise the multiple images displayed in front of them in concert with other technologies and in collaboration with other personnel. A body of practice has emerged which infuses how they monitor the incidents displayed on the system, how they make sense of what is happening on the screens and what is not, and how they draw together the fragmented views available to them. More importantly, these resources are made sense of with respect to particular organisational demands of the setting. Viewings may engender subsequent actions, but the sequential implications of a viewing also inform how the images are viewed, what is noticed and selected as needing attention. What appears to be a peculiarly individual and mundane activity, looking at screens, is actually achieved through practices that are social and infused with the organisational constraints of the setting.

Acknowledgements

We would like to thank Jon Hindmarsh, Hubert Knoblauch and other members of the WIT Research Group for discussions relating to this chapter. We are grateful to the partners in the DTI/ESRC Link Portrait project for their collaboration; in particular, we would like to thank Eddy Goddard and Barry Hodges and the personnel of London Underground who are involved in this study, for their continued assistance and support.

References

1 ATKINSON, J. M., and HERITAGE, J. C. (Eds.): 'Structures of social action: studies in conversation analysis' (Cambridge University Press, Cambridge, 1984)

2 BENFORD, S., and FAHLÉN, L.: 'A spatial model of interaction in large visual environments'. Paper presented at the ECSCW, Milan, September 13–17, 1993

3 BLY, S. A., HARRISON, S., and IRWIN, S.: 'Media spaces: bringing people together in a video, audio and computing environment', Communications of the ACM, 1992, **36** (1), pp. 28–47

4 DAVIES, A. C., YIN, J. Y., and VELASTIN, S. A.: 'Crowd monitoring using image processing', *IEE Electronic and Communications Engineering Journal*, 1995, **7**, (1), pp. 37–47

5 DOURISH, P., and BLY, S.: 'Portholes: awareness in a distributed work group'. Paper delivered at the CHI '92, Monterey, CA, USA, 1992

6 DREW, P., and HERITAGE, J. C. (Eds.): 'Talk at work: interaction in institutional settings' (Cambridge University Press, Cambridge, 1992)

7 FISHER, D. F., MONTY, R. A., and SENDERS, J. W. (Eds.): 'Eye movements: cognition and visual perception' (Lawrence Erlbaum Associates, Hillsdale, NJ, 1981)

8 GARFINKEL, H.: 'Studies in ethnomethodology' (Prentice-Hall, Englewood Cliffs, NJ, 1967)

9 GAVER, W. W., MORAN, T., MACLEAN, A., LOVSTRAND, L., DOURISH, P., CARTER, K. A., and BUXTON, W.: 'Realizing a video environment: EuroPARC's RAVE system'. Paper presented at the CHI 92, Monterey, CA, May 3–7, 1992

10 GOODWIN, C.: 'Conversational organisation: interaction between speakers and hearers' (Academic Press, London, 1981)

11 GOODWIN, C., and GOODWIN, M. H.: 'Seeing as a situated activity: formulating planes', *in* ENGESTRÖM, Y., and MIDDLETON, D. (Eds.): 'Cognition and communication at work' (Cambridge University Press, Cambridge, 1996) pp. 61–95

12 GREATBATCH, D., LUFF, P., HEATH, C. C., and CAMPION, P.: 'Interpersonal communication and human-computer interaction: an examination of the use of computers in medical consultations', *Interacting With Computers*, 1993, **5**, (2), pp. 193–216

13 HARPER, R., HUGHES, J., and SHAPIRO, D.: 'Harmonious working and CSCW: computer technology and air traffic control'. Paper presented at the ECSCW '89, Amsterdam, 1989

14 HEATH, C. C.: 'Body movement and speech in medical interaction' (Cambridge University Press, Cambridge, 1986)

15 HEATH, C. C., KNOBLAUCH, H., and LUFF, P.: 'Technology and social interaction: the emergence of "workplace studies"', *British Journal of Sociology*, 2000, **51**, (2), pp. 299–320

16 HEATH, C. C., and LUFF, P.: 'Collaborative activity and technological design: task coordination in London Underground control rooms'. Paper presented at the ECSCW 1991, Amsterdam, Netherlands, 1991

17 HEATH, C. C., and LUFF, P.: 'Explicating face to face interaction', *in* GILBERT G. N. (Ed.): 'Researching social life' (Sage, London, 1992a)

18 HEATH, C. C., and LUFF, P. K.: 'Media space and communicative asymmetries: preliminary observations of video mediated interaction', *Human-Computer Interaction*, 1992b, **7**, pp. 315–46

19 HEATH, C. C., and LUFF, P. K.: 'Technology in action' (Cambridge University Press, Cambridge, 2000)

20 HEATH, C. C., LUFF, P. K., and SELLEN, A.: 'Reconsidering the virtual workplace: flexible support for collaborative activity'. Paper delivered at the ECSCW '95, Stockholm, Sweden, September 10–14, 1995

21 HERITAGE, J. C.: 'Garfinkel and ethnomethodology' (Polity Press, Cambridge, 1984)

22 HINDMARSH, J., FRASER, M., HEATH, C. C., BENFORD, S., and GREENHALGH, C.: 'Fragmented interaction: establishing mutual orientation in virtual environments'. Paper delivered at the CSCW '98, Seattle, WA, 1998

23 INDYK, D., and VELASTIN, S. A.: 'Survey of range vision systems', *Mechatronics*, 1994, **4**, (4), pp. 417–49

24 JIROTKA, M., LUFF, P., and HEATH, C.: 'Requirements for technology in complex environments: tasks and interaction in a city dealing room', *SIGOIS Bulletin* (Special Issue) Do users get what they want? (DUG '93), December 1993, **14**, (2), pp. 17–23

25 KIRWAN, B., and AINSWORTH, L. K. (Eds.): 'A guide to task analysis' (Taylor and Francis, London, 1992)

26 LUFF, P., HINDMARSH, J., and HEATH, C. (Eds.): 'Workplace studies: recovering work practice and informing system design' (Cambridge University Press, Cambridge, 2000)

27 LYNCH, M.: 'Scientific practice and ordinary action: ethnomethodology and social studies of science' (Cambridge University Press, Cambridge, 1993)

28 MARANA, A. N., VELASTIN, S. A., COSTA, L DA F, S.A., and LOTUFO, R. A.: 'Automatic estimation of crowd occupancy using texture and NN classification', *Safety Science*, 1998, **28**, (3), pp. 165–75

29 MONTY, R. A., and SENDERS, J. W. (Eds.): 'Eye movements and psychological processes' (Lawrence Erlbaum Associates, Hillsdale, NJ, 1976)

30 NORRIS, C., MORAN, J., and ARMSTRONG, G. (Eds.): 'Surveillance, closed circuit television and social control' (Ashgate, Aldershot, 1998)

31 RENSINK, R. A., O'REGAN, K., and CLARK, J. J.: 'To see or not to see: the need for attention to perceive changes in scenes', *Psychological Science*, 1997, **8**, pp. 368–73

32 SACKS, H.: 'Notes on police assessment of moral character', *in* SUDNOW, D. (Ed.): 'Studies in social interaction' (Free Press, New York, 1972) pp. 280–93

33 SACKS, H.: 'Lectures in conversation: volumes I and II' (Blackwell, Oxford, 1992)

34 SENDERS, J. W., FISHER, D. F., and MONTY, R. A. (Eds.): 'Eye movements and the higher psychological functions' (Lawrence Erlbaum Associates, Hillsdale, NJ, 1978)

35 SUCHMAN, L.: 'Technologies of accountability: on lizards and aeroplanes', *in* BUTTON, G. (Ed.): 'Technology in working order' (Routledge, London, 1993) pp. 113–26

36 VELASTIN, S. A.: 'Towards detection of dangerous situations in images of crowds in semi-confined spaces using image processing'. Paper presented at the Application of Optical and Electromagnetic Sensors to Obstacle and Intrusion Detection, Passenger Flow Analysis and Localisation, Villeneuve d'Asque, October 26, 1995

37 VELASTIN, S. A., YIN, J. H., VINCENCIO-SILVA, M. A., DAVIES, A. C., ALLSOP, R. E., and PENN, A.: 'Analysis of crowd movements and densities in built-up environments using image processing'. Paper presented at the IEE Colloquium on *Image Processing for Transport Applications*, London, December 9, 1993

38 VELASTIN, S. A., YIN, J. H., VINCENCIO-SILVA, M. A., DAVIES, A. C., ALLSOP, R. E., and PENN, A. 'Automated measurement of crowd density and motion using image processing'. Paper presented at the 7th International Conference on *Road Traffic Monitoring and Control*, London, April 26–28, 1994

Chapter 11

Development of a railway ergonomics control assessment package (RECAP)

Lucy Cordiner, Sarah Nichols and John Wilson

11.1 Introduction

Traditionally, control room design research has focused on the design of displays and controls (e.g. Kinkade and Anderson, 1984; Oborne, 1995; Wickens *et al.*, 1997), but more recently, interest has developed in the use of technologies such as closed circuit television (CCTV), different representations of geographical displays and the impact of the organisation of teams within control. As a result of control room redesign projects, a number of specific guidelines are available for particular applications or domains (e.g. MoD, 1988; NUREG-0700, 1996; MacKendrick, 1999). However, there is little (if any) published general advice available on how any company might fully evaluate all human factors aspects of its control room operations, from environmental conditions to workplace layout or cognitive demands of interface design. If such advice, supported by appropriate measurement tools, were available then current activities could be reliably assessed. These tools would have the added benefit of providing the opportunity to obtain a reliable benchmark against which the impact of any future changes or developments could be compared.

The aim of the work discussed here was to produce a framework for evaluation of the human factors (ergonomics) issues associated with the introduction of new, integrated control centres within the UK railway network. The framework took the form of a measurement package – Railway Ergonomics Control Assessment Package (RECAP) (Cordiner, 2000). The intention of RECAP was to enable the evaluation of:

- system performance in network control (including hardware, software, environment, systems and organisation);
- consequences for people involved in control operations;
- long term changes and benefits.

This chapter describes the process of developing RECAP, highlighting the different methodological approaches incorporated, and identifying the relative suitability of these approaches. In particular, the practical issues associated with implementing such a tool in a large, geographically dispersed organisation that currently uses a number of different system generations and types are discussed. A generic model of the process of developing such a tool is presented, and the way in which this model was applied in the railway control domain discussed. The chapter concludes with some practical guidelines derived from the experience of measuring and analysing human factors in railway control.

11.2 Development of RECAP

The development of RECAP was driven by the need to cover a number of key areas initially identified by the project sponsors (Railtrack PLC). These areas were:

- safety;
- network management capacity and capability;
- network performance;
- signalling environment;
- system effectiveness;
- role and workload redistribution;
- individual and team performance.

In order to gain a thorough understanding of the structure of railway control, how different control groups interacted with each other and with their customers (e.g. train operating companies) or support staff (e.g. maintenance staff), and the potential impact of the new integrated control centres, it was necessary to spend a large amount of time visiting a number of sites and talking to people at different levels of the systems hierarchy. After some analysis of both the overall railway system and the individual elements within this system, down to the level of critical elements of individuals' work tasks, three key groups were identified as being of particular interest for the initial administration of RECAP. These were:

- Electrical control room operators (ECROs): responsible for the supply of electrical power to tracks.
- Zone controllers: who monitor train movement in large geographical areas.
- Signallers: who are responsible for controlling train movements within a smaller geographical area.

In parallel with gathering knowledge about the industry, time was spent reviewing existing research that could be incorporated into RECAP. Areas where the development of new methods or techniques was required were also identified. It was vital that any methods that were chosen or developed were suitable for the diverse contexts (both current and future) in which RECAP was to be applied, ranging from office-like visual display unit (VDU) work-stations to small mechanical signal boxes. It should be noted that a fundamental consideration of this research was to use methods that could be applied in the field environment, as opposed to a simulator. In some cases, methods previously used in simulated research could be directly transferred to the field context, but in others a large amount of adaptation or development was required.

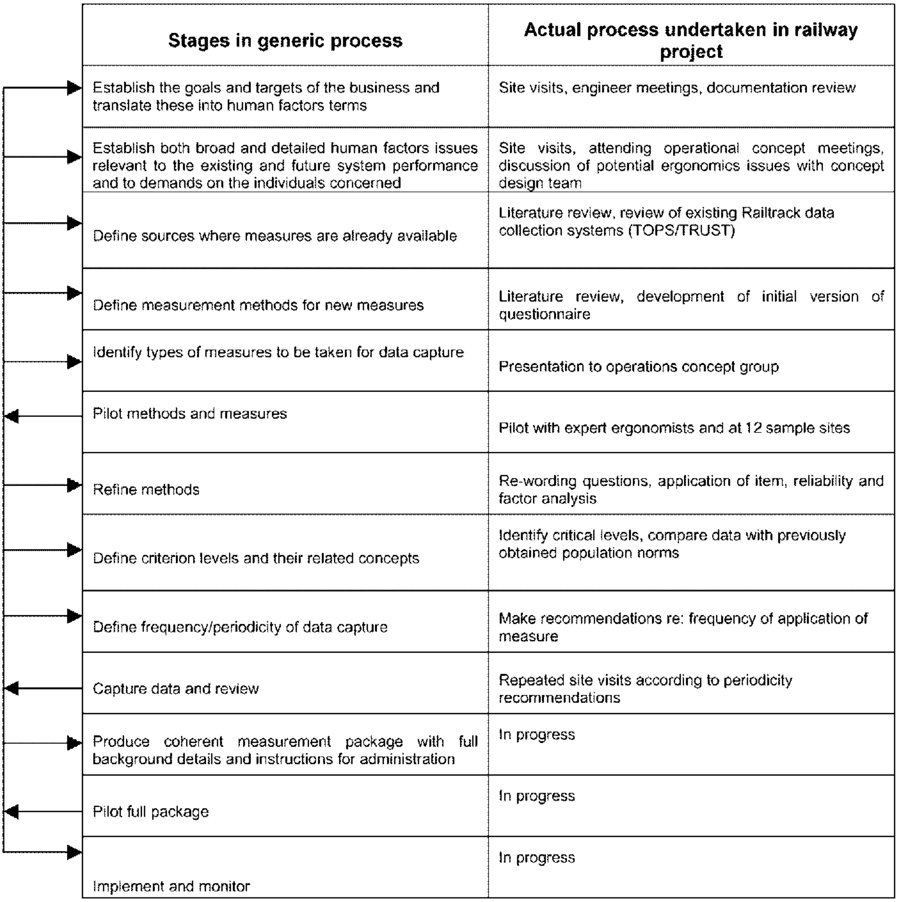
Figure 11.1 illustrates the development process identified as suitable for a human factors measurement package. The figure also describes how each stage was actually completed in the development of RECAP. Although it is implied that the stages should be completed sequentially, in practice the development of the measurement package involved a number of stages being undertaken concurrently. The development of the measurement package was an iterative process (which is still continuing) with the results of the measurement methods and data collected being reviewed and amended as necessary.

Key personnel involved in the RECAP development process were:

- Expert ergonomists.
- Railtrack management.
- Operational concept design group (made up of management, operational staff and ergonomists).
- ECROs, zone controllers and signallers.

Based on the overall business goals of Railtrack, and our gained understanding of the railway control environment, a matrix of relevant human factors issues was developed. This matrix defined the human factors issues that may be affected during system change, conditions of measurement and the aspects of human factors that need to be considered. The following 13 areas of human factors measurement were selected:

1. Health and Safety.
2. Workload.
3. Occupational stress.
4. Reliability.
5. Performance.
6. Work attitude.
7. Communication.
8. Workplace layout and environment.
9. Training.
10. Teamwork.
11. System usability.



Stages in generic process	Actual process undertaken in railway project
Establish the goals and targets of the business and translate these into human factors terms	Site visits, engineer meetings, documentation review
Establish both broad and detailed human factors issues relevant to the existing and future system performance and to demands on the individuals concerned	Site visits, attending operational concept meetings, discussion of potential ergonomics issues with concept design team
Define sources where measures are already available	Literature review, review of existing Railtrack data collection systems (TOPS/TRUST)
Define measurement methods for new measures	Literature review, development of initial version of questionnaire
Identify types of measures to be taken for data capture	Presentation to operations concept group
Pilot methods and measures	Pilot with expert ergonomists and at 12 sample sites
Refine methods	Re-wording questions, application of item, reliability and factor analysis
Define criterion levels and their related concepts	Identify critical levels, compare data with previously obtained population norms
Define frequency/periodicity of data capture	Make recommendations re: frequency of application of measure
Capture data and review	Repeated site visits according to periodicity recommendations
Produce coherent measurement package with full background details and instructions for administration	In progress
Pilot full package	In progress
Implement and monitor	In progress

Figure 11.1. Process of development of human factors measurement techniques.

12. Decision making.
13. Work organisation and management.

An extensive review of the ergonomics literature was undertaken to identify if any suitable measurement methods could be adapted from other industries or extracted from relevant ergonomics guidelines. In conjunction with this, a search for existing measurement methods already used within Railtrack was undertaken. A matrix of general human factors issues, specific issues associated with railway control and candidate measurement methods was then produced. This matrix is shown in Figure 11.2.

One of the outcomes of the production of this matrix was that several different key areas of interest emerged, for which separate tools were developed. In particular, a questionnaire tool (REQUEST) was developed (see Cordiner *et al.*,

General human factors issue	*Specific railway control issues*	*Measurement methods*
1 Health & Safety	Compliance with regulations Safety culture	Review of standards and regulations compliance Review of incident data, Subjective rating scales
2 Workload	Overload/underload Performance Human reliability	Subjective rating scales, Primary and secondary task performance Direct/retrospective video observation
3 Occupational stress	Influence on performance Relationship with work attitudes	Subjective rating scales Physiological measurement
4 Reliability	Performance of human–machine system System safety	Error analysis from railway reporting databases Near miss reporting system, Task analysis
5 Performance	Human error Level and type of intervention Performance efficiency Situation awareness	Subjective rating scales Direct/retrospective video observation Analysis of data from railway reporting databases Review of incident data
6 Work attitude	Job satisfaction and control Problem solving demand Responsibility/error cost Skill utilisation	Subjective rating scales Interviews Group discussion
7 Communication	Quality, amount, mode Communication network	Communication mappings Review of incident data, 360° interviews
8 Workplace layout and environment	Work-station design Physical discomfort Environmental conditions	Direct measurement Subjective rating scales Mock-up/modelling
9 Training	Training and re-training needs Training effectiveness	Subjective rating scales Training needs analysis Change in work performance, Rules/knowledge measurement
10 Teamwork	Social interaction Team/co-worker support Team cohesiveness, Team trust	Rating scales Workshops Observation/ethnography
11 System usability	Efficiency, likeability, control over system, learnability	Subjective rating scales Heuristic evaluation and expert walkthrough Error analysis and data logging
12 Decision making	Effectiveness, confidence People involved in decision making	Subjective rating scale Interview
13 Work organisation and management	Responsibilities, costs	Documentation analysis Observation, Task analysis

Figure 11.2. Matrix of measures.

2000) and a programme of signaller workload assessment (RELOAD) was implemented (see Nichols *et al.*, 2001). Other areas of research included a consideration of situation awareness in railway control (RESA) and error analysis (REPEAT). This process is described in more detail in Wilson (2001).

It should be noted that the methods selected in the matrix are not only traditional subjective assessment or observation methods, but also include examination of archival data, and use of existing databases used within the railway industry. Not all of the methods listed in the matrix were actually applied in the initial implementations of RECAP, but may well be used in future work.

This matrix was then reviewed by the operational concept design team. During this review the various methods were evaluated for their practicality, appropriateness and anticipated acceptance for use in the field. As a result of this, a number of the areas identified in the matrix were initially evaluated through the railways ergonomics questionnaire (REQUEST). This chapter discusses the measurement of three of the original 13 human factors issues identified in the matrix – workload, work attitude and communication.

11.3 Discussion of specific human factors issues

11.3.1 Workload

Workload was identified as one of the key human factors issues that would be affected by the introduction of the new integrated control centres. The type and extent of tasks for staff working in the new centres are likely to be changed, with staff becoming responsible for a mix of tasks currently performed by dedicated staff (e.g. signallers or electrical control room operators). It was also envisaged that the geographical area that an individual would control would become larger. Furthermore, the increasing use of computer-based signalling systems meant it was likely that the nature of the signalling task would develop into a predominantly monitoring and supervisory role. Therefore, any workload measurement undertaken needed to provide data on the ‘amount’ of work currently carried out – both mental and physical – and how this changed under different circumstances. It also needed to be in a form that could be used to inform the development of future signalling technology to allow the optimisation of operator workload and staff loading in future control centres.

A literature review identified several different techniques that could be used to measure workload and operator loading; these included subjective, performance and psychophysiological measures. Due to the difficulty in simulating a controller’s job (see Nichols *et al.*, 2001) it was decided to assess workload in the ‘live’ environment. As a result, intrusive methods such as those used to assess operators’ physiological responses, or the introduction of secondary task measures (Meshkati *et al.*, 1995) were felt not to be appropriate methods. Similarly, the use of performance measures was also judged not to be practical in a field situation due to the safety critical nature of the job. (However, these methods may of course be appropriate for use during simulations.) In the first instance, it was decided an estimate of current work levels would be obtained from subjective assessment and observation techniques.

11.3.1.1 Subjective self report methods

Subjective methods provide data on operators' perceived level of loading. The criteria for any subjective methods used in the assessment of railway control staff were that they had to:

- cause minimum interference to the operator doing their job;
- be quick and easy to use in the field;
- require minimum explanation before use;
- be applicable to all types of control environment;
- be understandable by operators and require no (or very little) training.

Two self reporting rating scales were identified as meeting the above criteria: the NASA Task Load Index (TLX) (Hart and Staveland, 1988) and Revised Air Force Flight Test Centre (AFFTC) (Ames and George, 1993). A pilot study using the NASA TLX was carried out with signallers and staff from a Zone control centre. However, this scale did not prove to be very successful because:

- TLX evaluates the workload of specific tasks. Railway control jobs do not involve discrete tasks – incidents are made up of many simultaneous tasks and incidents can be ongoing for a number of hours.
- Staff had some difficulty understanding the terminology used to describe the six constructs. These were revised, e.g. temporal demand was changed to time pressure, which did help improve understanding to some extent.
- Getting workload ratings on six scales (TLX is a multi-dimensional scale) took over two minutes and this length of time was not acceptable when staff were busy.

The AFFTC Revised Workload Estimate was then trialled. Controllers were asked every 10 minutes to rate their workload on a seven-point scale, with 1 being 'no system demands' to 7 'system unmanageable, essential tasks undone'. This scale, although uni-dimensional, was found to work better as it was less intrusive and could be completed in less than 30 seconds.

Staff were also asked to give an estimate of how much time they spent on different work tasks that made up their 'overall workload'. They were asked, through a questionnaire, to report the proportion of time spent on different task elements during different workload scenarios. Unfortunately this method of data collection did not prove to be particularly successful. Staff reported that it was very difficult to complete this section because they were not involved in managing all aspects of incidents. A number of respondents also commented that workload breakdown varied considerably depending on where and at what time the incident occurred, if any other incidents were ongoing at the same time, and how much help they received from colleagues. This meant staff found it difficult to generalise and only provided a rough estimate of the time they spent on different tasks. It is intended that this method of measurement will be modified as it is felt that this type of information may provide valuable details when defining job roles in the new control centres.

11.3.1.2 Direct observation

Additional information was obtained from direct observation of work activities. The frequency and range of tasks, such as telephone calls, route setting and interaction with other staff in the control box were noted. This provided valuable information on the types of tasks currently carried out and how they are sequenced. The techniques applied are discussed in more detail in Nichols *et al.* (2001) and Bristol and Nichols (2001).

11.3.1.3 Video observation

Observer Pro software (Noldus *et al.*, 2000) was used to retrospectively analyse tasks undertaken by signallers during several ten-minute periods. This helped build up a more detailed picture of time spent on different tasks and highlighted some of the more subtle behaviours that were missed during direct observation.

11.3.2 Work attitude

Mechanisms already existed within Railtrack PLC that could be used to measure the impact of integrated control centres on operational performance. For example, the number of minutes delay per train, time taken to apply emergency track possessions and number of trains regulated successfully were already being calculated by tracking systems. Therefore monitoring of these would highlight any effect of control integration on performance. However, no recording methods were currently being used to measure attitudes of controllers towards their work, and therefore there were no existing means of monitoring changes in work attitudes as a result of control integration.

The areas of particular interest relating to work attitudes were identified as job satisfaction, job complexity, timing control, method control, monitoring demand, repetitiveness/boredom, problem solving demand, responsibility/error cost and skill utilisation. It was decided that an appropriate way to measure work attitude was to incorporate questions relating to all of these issues within the REQUEST questionnaire tool.

An existing measure of job satisfaction (Mullarkey *et al.*, 1999) was applied. However, no suitable measures were found to exist to measure the other areas of interest, therefore a new scale – the Job Characteristics Scale (JCS) – was developed. The questions for this scale were produced by members of the research team and reviewed by expert ergonomists, human factors specialists within Railtrack and potential respondents. As a result of this review process, 32 items were derived, each of which were responded to on a five-point scale. Figure 11.3 shows example items from this initial version of the scale.

As a result of a factor analysis performed on the responses from the first issue of the questionnaire, six separate groups of questions were identified. Following the presentation of the question groups to a group of independent reviewers (Nichols *et al.*, 1999), these six factors were named:

- Method control.
- Work fulfilment.
- Consequences (potential).
- Task significance and variety.
- Skill requirements.
- Support and personal development.

Although this resulted in a reasonably large set of questions relating to job characteristics, it was felt that the importance of this area and the lack of any existing monitoring tools within existing railway control procedures justified the retention of such a large number of items.

11.3.3 Communication

It was recognised that accurate, timely and effective communication and decision-making would be key in the ability of any new centre to achieve and maintain operational activities safely and efficiently. To gain an understanding of how information was currently transmitted and verified, and to anticipate consequences of any decisions made, data was collected from staff involved in critical decision-making. A range of staff were asked to indicate how frequently they communicated with other members of staff during incidents, either to provide or receive information and who they believed was involved in critical decision-making during

Job Complexity					
How much is each of the following present in your work?					
The opportunity to use your abilities?	Not at all	Just a little	Moderate amount	Quite a lot	A great deal
Timing Control					
To what extent					
Do you set your own pace of working?	Not at all	Just a little	Moderate amount	Quite a lot	A great deal
Monitoring Demand					
To what extent					
Do you have to keep track of more than one thing at a time?	Not at all	Just a little	Moderate amount	Quite a lot	A great deal

Figure 11.3. Example items from Job Characteristics Scale (JCS).

different incidents. Fourteen different scenarios, each asking about communication during general information transfer and decision-making, were identified ranging from normal service to specific incidents (e.g. points failure). An example showing the information flow and decision-making during normal service is shown in Figure 11.4. The graph shows the mean rating, where '1 = never' and '5 = always', for each job type, by each of the four groups of respondents. The different lines indicate the different responses by categories of respondents.

The results from this process did indicate some differences in perceived responsibility within control. For example, signallers reported that they would frequently communicate with other signallers, whereas ECROs reported rarely communicating directly with signallers.

It was felt that the resulting graphs were a very useful way of summarising data; however, controllers reported that it was very difficult to estimate accurately the exact lines of communication and state who took decisions. An additional problem was that the number of scenarios (28 in total) took too long to consider. It is intended that this process, in a slightly revised format, will be re-applied at a later date as it is felt that it has the potential to provide very useful information that will assist in defining new roles for staff.

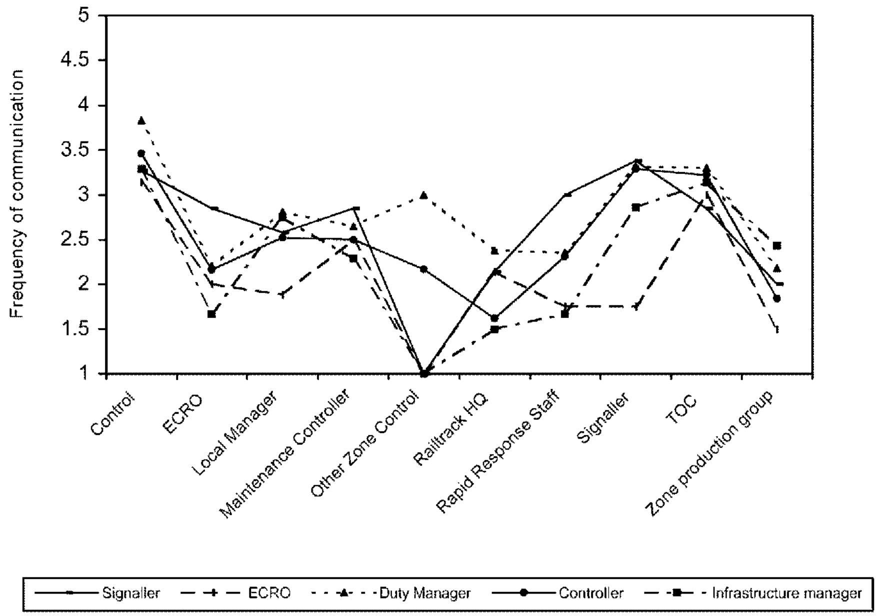


Figure 11.4. Normal service – information.

11.4 Feedback of results to respondents

One of the main criticisms made by controllers during site visits was that they were frequently asked for their opinions but then the data collected disappeared into a 'big black hole' and they never found out what the results were. Therefore, it was considered vital that to maintain controller co-operation, feedback was provided to all sites that participated in the study.

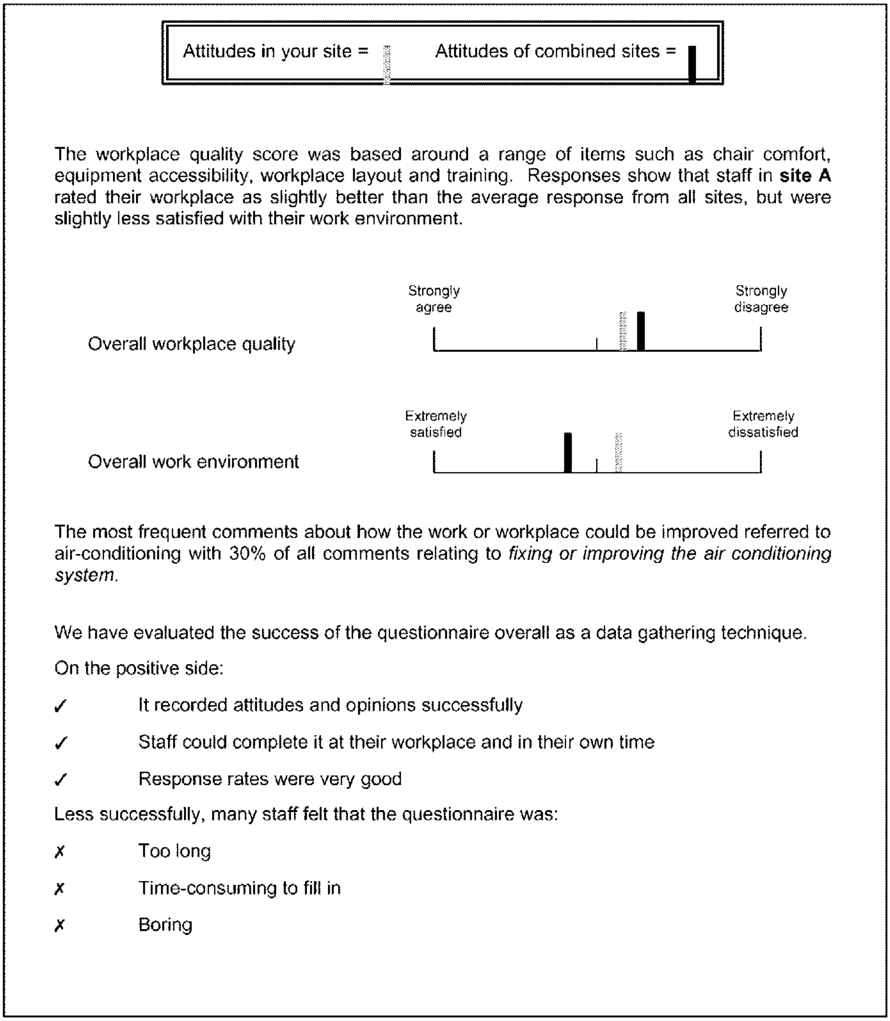
It was realised that it was not appropriate to provide feedback to the same level of detail as in management reports, due to both time constraints and the need to avoid possible direct comparison between sites. (Feedback to management provided information about each site but maintained individual response anonymity.) It was also recognised that the information had to be put into context; simply presenting numbers obtained from each scale would be meaningless. Therefore, a method of feedback was given in which each individual sites' details were compared to the average response obtain from across all the sites visited. Information was also provided on the measurement process itself. An extract from the feedback presented is shown in Figure 11.5.

11.5 General methodological issues

During the initial stage of development of RECAP, a number of methodological issues were encountered – some specific to the railway domain, some associated generally with the measurement of human factors issues in a field context. This section describes some of these methodological issues.

11.5.1 Reporting of cognitive processes via questionnaires

The problems encountered with the application of the retrospective subjective workload assessment measure within the initial administration of REQUEST illustrate the difficulty that participants can experience when they are asked to report on their cognitive processes (Smirnov, 1973). It is known that people tend to find it difficult to report on cognitive processes – indeed, elicitation methods such as verbal protocol specifically aim to overcome this difficulty (Ericsson and Simon, 1993). Therefore, in order to elicit information relating to cognitive processes – in this case, workload – concurrent measures were applied in a field context. However, the use of a field context also has implications for the type of data measured, primarily meaning that minimally intrusive methods should be employed.



Attitudes in your site = ▨ Attitudes of combined sites = ▮

The workplace quality score was based around a range of items such as chair comfort, equipment accessibility, workplace layout and training. Responses show that staff in **site A** rated their workplace as slightly better than the average response from all sites, but were slightly less satisfied with their work environment.

The most frequent comments about how the work or workplace could be improved referred to air-conditioning with 30% of all comments relating to *fixing or improving the air conditioning system.*

We have evaluated the success of the questionnaire overall as a data gathering technique.

On the positive side:

✓ It recorded attitudes and opinions successfully

✓ Staff could complete it at their workplace and in their own time

✓ Response rates were very good

Less successfully, many staff felt that the questionnaire was:

✗ Too long

✗ Time-consuming to fill in

✗ Boring

Figure 11.5. Example of feedback returned to all sites visited.

11.5.2 Field-based research vs. simulation

The differences between fieldwork and laboratory-based research have been well documented (see Robson, 1993). It was felt to be vital that the majority of tools applied within RECAP were field-based, due to the complexity, multiple inputs and outputs, and lack of predictability of all of the jobs examined within railway network control. However, there are obvious problems with examining issues within a field context.

A problem that was frequently encountered was the lack of ability to predict the level or type of work that is likely to be experienced at any particular time. In addition, despite efforts to introduce non-disruptive measures, when busy periods do occur, participants often, understandably, do not respond to verbal questioning, (exhibiting the same phenomenon as detected by a secondary task measure). This makes the collection of quantitative data in a field context problematic.

Another large consideration that had to be overcome was the need to ensure measures were as unobtrusive as possible. This placed restrictions on the study and meant that traditional measures such as secondary task measures are not suitable. However, observation methods, particularly combined with the available resources to use computer-based analysis tools, were felt to be successfully used.

11.5.3 Distinction between operator and control system

The initial approach taken within the RECAP research was to consider the operator and system as separate entities, with transfer of information between the two via inputs and outputs. However, several aspects of the work presented in this chapter have indicated that this is an approach that can place unnecessary restrictions on the way in which data is collected. For example, when a signaller receives a telephone call from a driver, the driver may have the original purpose of making a request to the signaller (i.e. the signaller has received an 'input'). However, during the telephone call, the signaller may inform the driver about a delay that s/he may encounter later on in their journey, such as delay due to traction failure. In this case, the signaller has also given an 'output' to another person in the work environment. Hollnagel (1998) described the concept of a joint cognitive system – an approach that emphasises the socio-technical context of a human action or decision rather than considering system and human inputs and outputs independently. This may well be a useful approach to adopt when considering the railway control environment.

11.5.4 Questionnaire fatigue

As with any large organisation, a number of questionnaires are routinely distributed within Railtrack. Understandably, this is likely to reduce the response rate especially if there is a feeling amongst respondents that the survey is not in fact confidential or that nothing is done with the information obtained. This potential problem of 'questionnaire fatigue' was highlighted to the authors early on during the pilot trial. Therefore, steps were taken to maximise the response rate. These included:

- making the questionnaire design distinct from routine mailings, but ensuring that it was clear to understand;
- using terminology that was appropriate and recognisable to the participants;
- splitting questions into manageable, related sections;

- issuing printed rather than photocopied questionnaires to maximise quality and legibility;
- making it clear that the questionnaire was being distributed by an independent consultancy and only summarised results would be given to management.

One main contributing factor to the high response rate (40 per cent) was that one of the research team personally visited the individual sites, distributing the questionnaire and explaining the purpose to all staff on at least one shift. This process was very time consuming but was judged to be worthwhile due to the high response rate and the number of positive comments received from participants. Participants said they appreciated that someone was taking the time to come a visit them and see for themselves the working conditions. (The benefit of this approach was further highlighted by the poor response obtained from the one site not visited personally.)

11.5.5 Issues associated with longitudinal studies

The aim of RECAP is to provide a benchmarking tool that will monitor the effects of system change and highlight any specific issues. As part of developing the tool decisions had to be made as to the periodicity of data capture. The frequency with which the different parts of the tool were to be re-administered varied, but the factors that decided the frequency of REQUEST administration are discussed here. A review of the literature revealed no standard or acceptable periodicity for administration of a questionnaire. After some discussion with fellow ergonomists it was decided that in the first instance it would be issued annually at the same time each year, but that this might extend to bi-annually in the future. This time was felt to:

- be sufficient to ensure that an adequate record of any impact of organisational or workplace change was obtained;
- limit any additional bias caused by time of year, weather, impact, or annual pay negotiations;
- provide a long enough period for any issues highlighted to have been addressed;
- provide a long enough period for attitudes to have changed and, quite importantly;
- be a long enough interval so controllers did not get bored completing the same questionnaire on repeated occasions.

It was realised that this could not take into account unexpected external events such as a major train crash, which are likely to have an overwhelming impact on workplace attitudes and workload but would ensure the opinions gathered were as compatible as possible.

11.6 Guidelines and recommendations

As a result of our experience in developing the RECAP tool, a number of general guidelines for conducting such field-based research have emerged. Of course, general guidelines for questionnaire design apply (see Oppenheim, 1992; Robson, 1993).

The following practical guidelines, relating to the tool development process, are recommended. Some of the proposed guidelines are likely to already be known by researchers, but are felt to be worth emphasising due to their particular importance in the context of developing field-based measures.

- **Learn about the company**. This enables the use of appropriate terminology within measures. Also, understanding of the nature of work (e.g. safety-critical) helps to determine the choice of suitable methods.
- **Visit sites personally**. This will increase the questionnaire response rates and also enable a rapport to be established with participants, who may then feel more comfortable participating in any field trials.
- **Pilot tools thoroughly**. This process should include the development of a number of iterations of any tools within the research team, and should also cover extensive piloting of tools in the field, before any reliable empirical data can be obtained. This is especially important if tools are to be used for benchmarking or long-term monitoring purposes.
- **Make sure you are not interfering with work tasks**. An inevitable effect of fieldwork is that participants' behaviour may be affected by the presence of an observer. This may have the effect of changing the way in which work tasks are performed. Completing a number of visits, so that participants become accustomed to the observation process, may minimise any interference that may occur.
- **Make presentation of questionnaires professional**. This standard piece of guidance within questionnaire design should be emphasised. Professional-looking questionnaires are more likely to be taken seriously; in addition, making the 'look' of a questionnaire distinct from standard company stationery has the advantage of highlighting the independent nature of the research.
- **Feedback results to participants**. The feedback of results from tool administration to participants is important to retain compliance with research, particularly with questionnaire respondents. If questionnaires are being presented to respondents for a second time, it is recommended that these are accompanied by feedback from earlier administrations of the tool.
- **Continually update and improve tools**. Inevitably researchers will wish to update later versions of tools, in order to probe issues highlighted in earlier administrations, or improve questionnaire wording, for example. It is important to maintain an appropriate compromise between continual improvement of the tools whilst retaining consistency between versions in order to allow longitudinal monitoring of change.

- **Find out about other surveys/evaluation projects being conducted within the company**. If possible, it is useful if direct conflict or repetition of measurement programmes can be avoided within an individual company. Researchers should try and find out what standard monitoring procedures are routinely implemented, and attempt to avoid administering similar tools within a short time period.
- **Consider context of work rather than isolated work elements**. Although it can be more straightforward to treat individual operators and systems separately, it may be appropriate to consider the entire work environment (which may in some cases include remote sites communicated with by telephone) when attempting to gain an overall understanding of operator's work.

The process described in this chapter is continuing, with the tools still being developed and routinely administered. It is hoped that the lessons learned from the development of such a global tool can be used by researchers in other contexts, and that the available resources within the human factors community can be developed to allow the impact of organisational and workplace change to be effectively monitored.

Acknowledgements

The work presented in this chapter was sponsored by Railtrack PLC. The authors would like to thank Theresa Clarke and Steve Roberts at Railtrack for their invaluable help and support.

References

1 AMES, L. L., and GEORGE, E. J.: 'Revision and verification of a seven-point workload scale'. AFFTC-TIM-93-01 (Air Force Flight Test Center, Edwards AFB, CA, 1993)
2 BRISTOL, N., and NICHOLS, S.: 'Impact of new technology on operator workload: development of a measure of underload', *in* HANSON, M. A. (Ed.): 'Contemporary ergonomics 2001' (Taylor & Francis, London, 2001) pp. 251–6
3 CORDINER, L., NICHOLS, S., and WILSON, J. R.: 'Development of a railway ergonomics control centre assessment package (RECAP)'. Proceedings of the HFES/IEA International Meeting, San Diego, CA, 2000
4 ERICSSON, K. A., and SIMON, H. A.: 'Protocol analysis: verbal reports as data' (MIT Press, Cambridge, MA, 1993)
5 HART, S. G., and STAVELAND, L. E.: 'Development of NASA-TLX (Task Load Index): results of empirical and theoretical research', *in* HANCOCK, P. A., and MESHKATI, N. (Eds.): 'Human mental workload' (North-Holland, Amsterdam, 1988)
6 HOLLNAGEL, E.: 'Cognitive reliability and error analysis method (CREAM)' (Elsevier, Oxford, 1998)

7 KINKADE, R. G., and ANDERSON, J.: 'Human factors guide for the nuclear power plant control room development'. EPRI report NP-3659 (Electric Power Research Institute: Palo Alto, CA, 1984)

8 MACKENDRICK, H.: 'R & D report: development of a human machine interface (HMI) Guidelines database for air traffic control centres' (National Air Traffic Services Ltd, London, April, 1999)

9 MESHKATI, N., HANCOCK, P. A., RAHIMI, M., and SAWES, S. M.: 'Techniques in mental workload assessment', *in* WILSON, J. R., and CORLETT, E. N. (Eds.): 'Evaluation of human work: a practical ergonomics methodology' (Taylor & Francis, London, 1995, 2nd edn.)

10 MOD 'DEF STAN 00-25 Human factors for designers for equipment' (Ministry of Defence, 1988)

11 MULLARKEY, S., WATT, T., WARR, P., CLEGG, C., and STRIDE, C.: 'Measure of job satisfaction, mental health and job-related well-being: a benchmarking manual' (Institute of Work Psychology, Sheffield, 1999)

12 NICHOLS, S., BRISTOL, N., and WILSON, J. R.: 'Workload assessment in railway control', *in* HARRIS, D. (Ed.): 'Engineering psychology & cognitive ergonomics' (forthcoming, 2001)

13 NICHOLS, S., CORDINER, L., and WILSON, J. R.: 'Statistical validation of human factors questionnaire for Railtrack'. Confidential IOE Report 99/005, 1999

14 NOLDUS, L. P. J. J., TRIENES, R. J. H., HENDRIKSEN, A. H. M., JANSEN, H., and JANSEN, R. G.: 'The observer video-pro: new software for the collection, a management and presentation of time-structured data from videotapes and digital media files', *Behaviour Research Methods, Instruments and Computers*, 2000, **32**, (1), pp. 197–206

15 NUREG-0700: 'Human-system interface design review guideline – process and guidelines'. Final Report Rev. 1, Vol. 1 & Vol. 2, Office of Nuclear Regulatory Research, Nuclear Regulatory Commission, 1996

16 OBORNE, D. J.: 'Ergonomics at work', (John Wiley & Sons, Chichester, 1995, 3rd edn.)

17 OPPENHEIM, A. N.: 'Questionnaire design, interviewing and attitude measurement' (Pinter Publishers, London, 1992)

18 ROBSON, C.: 'Real world research' (Blackwell, Oxford, 1993)

19 SMIRNOV, A. A.: 'Problems of the psychology of memory' (Plenum Press, New York, 1973)

20 WICKENS, C. D., GORDON, S. E., and LIU, Y.: 'An introduction to human factors engineering' (Longman, New York, 1997)

21 WILSON, J. R.: 'Ergonomics in railway network control', *in* HARRIS, D. (Ed.): 'Engineering psychology & cognitive ergonomics' (forthcoming, 2001)

Part Three
Control Room Design

Chapter 12
Control room design

John Wood

12.1 Introduction

In the developed and developing world control rooms directly influence much of our daily lives. Our supply of electricity and water, the control of our railways and the disposal of our waste, are all examples of services co-ordinated through control centres. Modern communications and computing power have been used to achieve economies of scale with individual control centres being the hub of ever-larger networks. This, coupled with a tendency to downsize staffing levels, places an increasing burden on these centres and the staff who operate in them.

It is these very changes that have brought the issue of 'ergonomics' – matching systems to people – to the forefront. Systems that rely on human intervention will fail if the demands they make on their operators lie outside their capability. On the other hand, systems that can run automatically will also 'fall over' if operators do not understand how to intervene when control systems go haywire. The risks of overall systems failure are now too great for the human factor to be ignored – systems designers need to take account of people-related issues and make sure that human factors are given as much attention as mechanical and electrical elements.

Whilst this chapter concentrates on the role of the operator in the control room there are other 'users' of the system whose ergonomic requirements need to be considered. Lack of attention to matters such as ease of equipment maintenance, field equipment design and training will all undermine overall performance just as surely as 'loose wires' and 'bugs' in software.

12.2 Lack of 'top down' approach

For many project teams designing control rooms, 'ergonomics' is something that is added towards the end of the programme. It is this failure to address the underlying human factor issues which will later come back to haunt the specifier and operator through festering problems that are not that easily resolved.

In common with other elements of control system development, human factors need to be approached systematically if abortive work is to be avoided. The approach taken is no different to that adopted by any systems engineer, namely, the initial consideration of general requirements followed by specific, detailed needs.

For the human factors engineer this top-down approach starts with an appreciation of the system, after which control room objectives are spelt-out – preferably in terms of performance measures. The role of the operator within the system is then determined using a baseline understanding of human limitations in such matters as short-term memory, the ability to absorb information simultaneously from multiple sources and maximum periods of concentration. This approach results in the design of meaningful jobs that can be successfully completed by control room staff.

The essence of the 'top-down' approach has been incorporated into the standard on the 'Ergonomic Design of Control Centres' (ISO 11064). A simplified version is presented in Figure 12.1.

12.3 Impact of increasing automation on control room operation

At first sight there are attractions to automating most of the functions, leaving the operator free to undertake 'more worthwhile' activities. Unfortunately the downside to such a solution is that lightly loaded operators are prone to get bored and even to falling asleep – it's in our nature to do this. The matrix of 'levels of automation' and human factor issues in Table 12.1 attempts to shed some light on some of the issues which arise from different levels of system automation.

12.4 Team working and job satisfaction

Providing a good working environment is often considered a way of smoothing over organisational issues such poor workflow, clashing job descriptions, weak lines of supervision and a lack of team spirit. It is my experience that such an approach is doomed to failure – however much is spent on interior design and 'potted plants'.

Most control room failures occur not because people are unable to read a piece of paper, or visual display unit (VDU) screens are washed-out with veiling reflections, but because of a lack of 'can do' spirit and goodwill. Goodwill will be eroded faster by a poor organisational structure than a poor working environment and it is essential that matters such as lines of responsibility, balanced distributions of tasks and robust job description are worked out as part of the implementation strategy.

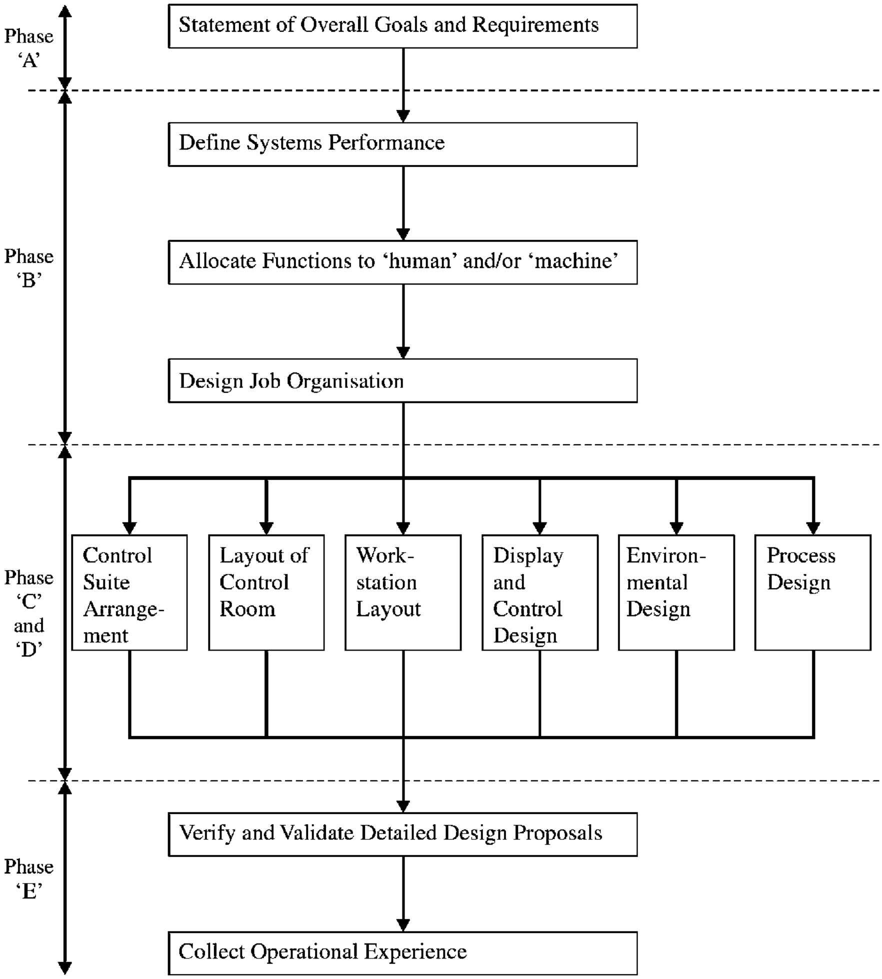


Figure 12.1. *Top down approach (After ISO 11064-1 – The Ergonomic Design of Control Centres).*

12.5 Why do we need a control room standard?

It is a sad fact of life that many operators have to work 12-hour shifts, coping with wildly varying levels of workload and in totally inappropriate conditions. We also expect that they will perform flawlessly. The importance of creating environmental conditions commensurate with the nature of the tasks being undertaken has been

Table 12.1. Human factors issues and levels of automation of systems.

	Level of automation/systems integration	*'People Issues'*
1	system 'start' and 'stop' buttons	– no operators involved, apart from starting and stopping
2	integrated automatic system – intervention by exception	– skills fade – boredom – supervision fade
3	semi-automated integrated system	– operator actively in control loop
4	discrete automated systems	– operator(s) needs to be able to transfer information between systems
5	discrete manual systems	– potential overload under perturbed conditions

recognised for some time in some service organisations, for example air traffic control. For others, the recognition of the importance of these factors has been belated – the emergency service and industrial plant control centres provide examples.

The standard now being written (ISO 11064) on the ergonomics of control rooms provides a base-line set of requirements for these rooms. It recognises that specialised tasks take place in control rooms and that the tasks and working environments need to be designed to support the activities being undertaken.

12.6 The structure of the control room ergonomics standard

Control rooms are to be found in process industries, military environments, emergency services, the security industry and the utilities – to name just a few applications. Such centres will range in size from a single operator to well over 200. Also, technological advance differs from sector to sector, with the public services being somewhat more cautious, perhaps, than their private industry counterparts.

The various parts of the new standard are shown in Table 12.2. In order to cope with this wide field of application some early decisions were made in structuring the standard. We de-coupled the relatively stable ergonomic material from that which might be subject to rapid change. Thus the first seven parts of the standard deal with general principles of good practice of control room ergonomics. The last part will deal with additional requirements that may need to be applied to particular types of centre such as nuclear reactor control rooms and railway signalling centres.

The following is a resume of some of the topics that are to be found in each part of the first seven parts of the standard.

Table 12.2. ISO 11064 – The ergonomic design of control centres: List of parts.

Part 1:	Principles for the design of control centres.
Part 2:	Principles of control suite arrangement.
Part 3:	Control room layout.
Part 4:	Work-station layout and dimensions.
Part 5:	Displays and controls.
Part 6:	Environmental requirements for control rooms.
Part 7:	Principles for the evaluation of control centres.
Part 8:	Ergonomic requirements for specific applications.

12.6.1 Part 1 – principles for the design of control centres

The first part presents a process whereby an ergonomic approach in the planning and design of a control centre is achieved, see Figure 12.1 for a version of the process.

Part 1 lists nine principles for the ergonomic design of control centres:

- **Principle 1**. The Application of a Human Centred Design Approach – essentially the need to consider human factors throughout the process following the process chart presented in Figure 12.1.
- **Principle 2**. The Integration of Ergonomics in Engineering Practice – which covers the organisation of projects and stipulates that ergonomics needs to be part of project managers' guidelines in which the ergonomic contribution is planned in for the entire project.
- **Principle 3**. The Improvement of Design through Iteration – this principle recognises that control centre design projects are dynamic, changing through the various stages of concept design, detailed design and implementation. The principle requires that there should be formal processes whereby human factors is an integral part of this iterative process.
- **Principle 4**. The Conduct of a 'Situational Analysis' – a requirement to examine existing or similar working environments. This principle overlaps somewhat with that of conducting a task analysis discussed next.
- **Principle 5**. The Conduct of a Task Analysis – a core requirement to examine in the tasks that operators are likely to perform by examining existing similar tasks or modelling future tasks.
- **Principle 6**. The Design of Error Tolerant Systems – a principle that starts with the premise that human error cannot totally be eliminated.
- **Principle 7**. Ensuring User Participation – this recognises that user participation is essential in the design process and that this input has to be properly structured.
- **Principle 8**. Forming an Interdisciplinary Team – reflecting the need for agreement with a number of stakeholders. This principle spells out the requirement to identify the appropriate team to co-ordinate the work of control centre design and agree 'sign-off'. This team will include user representatives.

- **Principle 9**. Documenting the Ergonomic Design – this requirement spells out the need to fully document the ergonomic basis of the final design so that any future changes are undertaken with a full knowledge of the original criteria used for the existing centre.

It will have been noted from Figure 12.1 that the standard is based around an ergonomic process starting with goal setting and finishing with detail design and evaluation.

- **Phase A** – 'Clarification', during which such aspects as the systems to be managed, project timescales, equipment constraints and project budgets are determined. It is within this context that a satisfactory ergonomic solution must be sought.
- **Phase B** – 'Analysis & Definition', during which the control centres function and performance requirements are spelt out and culminating in a preliminary functional allocation (between computer and operator) and outline job designs.
- **Phase C** – 'Conceptual Design', during which such aspects as control room layouts, furniture designs, display and control arrangements are developed.
- **Phase D** – 'Detailed Design', involving the development of detailed ergonomics and associated design specifications which can be used in the procurement of buildings, equipment and furniture. In the past this has usually been seen as the starting point for the ergonomics input.
- **Phase E** – 'Operational Feedback', during which post-commissioning reviews are conducted to identify successes and shortcoming in the design for use in future projects.

12.6.2 Part 2 – principles for the arrangement of control suites

Many control centres consist of a number of different rooms including the control room itself. Other functions to be provided for include equipment and maintenance zones, welfare areas and training facilities. Part 2 of the control room standard addresses the ergonomic issues involved in the layout and planning of such functionally related groups of rooms or spaces.

As with each part of this multi-part standard design processes are summarised in a flowchart. The format mimics that of Part 1 starting with a definition of operational objectives, followed by a task analysis, link analysis (see case study below for example) and culminating in control suite layout. The following examples illustrate the type of requirements covered in control suite layout.

12.6.2.1 'Traffic and routing'

Some control rooms also provide security and reception functions as well as process control. For these vehicle and pedestrian routes will play an important part in the location of the control room within the control suite complex. For other control rooms direct sight lines to plant, or in the case of airfield control towers, to runways, will dictate the location of the control room as well as the arrangement of windows.

12.6.2.2 'Environmental conditions'

The preferred location of a control area may be influenced by noisy processes, e.g. generators or by dust or heat gain and this may dictate the control suite layout adopted.

12.6.2.3 'Cleaning'

Operational staff need to feel proud of the installation they work in. Inability to clean areas properly can undermine this necessary pre-requisite for a first-rate installation. In designing for cleaning the standard requires that attention is paid to minimising the introduction of dirt and contaminates into the control suite. This requirement, for example, will be a consideration for the routing of staff from plant areas to control areas.

12.6.2.4 'Maintenance'

The requirements associated with maintenance reflect the standards' concern that maintenance engineers are also users and designing for their needs is an essential part of achieving a satisfactory ergonomic design. The standard includes requirements concerning ease of access, the provision of appropriate maintenance spaces and disruption to work-station users during maintenance activities.

12.6.2.5 'Visitors'

The standard addresses recognised conflicts that may occur between the need to maintain operator concentration and the wish to provide access to visitors. Requirements are included concerning both 'professional' and 'public' visitors and the control suite features which should be included to minimise their potential to disrupt control operations, for example by the appropriate design of viewing galleries and control of circulation routes.

12.6.3 Part 3 – control room layout

This Part presents the ergonomic principles that contribute to the good layout of control rooms. Its requirements relate to control room layout, work-station arrangements, off-work-station visual displays and control room maintenance ergonomics. The process presented for the design of control room layout is similar to the earlier parts, moving from a consideration of the needs to be met running through to the selection and testing of layouts. The standard takes account of the difference between the refurbishment of an existing control room and starting off with a 'clean sheet of paper'. Often refurbishment projects can be more challenging as there is no choice but to adopt a sub-optimal ergonomic solution.

Many control rooms suffer from a lack of space, especially towards the end of their planned lifespan. The standard recommends that control rooms allow for a 25 per cent increase in working positions and equipment during the expected life of the room.

The selection of an appropriate space is crucial to the success of achieving a satisfactory control room layout. The space needs to be big enough and of the appropriate shape. A heuristic value put forward by the standard for calculating space provision is to allow $9m^2$ to $15m^2$ per working position with a minimum of not less than $9m^2$. Where off-work-station, shared displays are used – such as mimics or projected overviews – a more generous space provision needs to be provided to take account of non-usable space close to the displays and additional maintenance areas. A space providing the necessary area on paper may still not be appropriate because of building constraints. Thus pillars, large glazed areas, clearances for doorways and unusable corners will all erode the actual usable space within which a satisfactory layout can be achieved. Also a long narrow room, even though of the same overall area as a square room, may not allow for a satisfactory layout to be achieved.

For some types of control rooms windows provide operational information – airport control towers, railway shunting yards. For others viewing the process may not be practical, such as with subway signalling systems or, alternatively, not required for the performance of the control task. However, the standard recommends the inclusion of windows wherever practical – not for the purposes of illumination but purely for psychological reasons.

In selecting spaces vertical dimensions may also be an important consideration. In many instances, particularly where control rooms are being installed in office buildings, the final finished floor to ceiling heights may constrain the use of optimum artificial lighting systems. The use of combined indirect/direct luminaires for control rooms has been found to minimise levels of glare and reflections off display screens. However, such installations typically require a minimum of three metres from floor to ceiling – preferably four metres. Sometimes the finished floor to ceiling height in a typical office block may be as low as 2.5 metres.

Having selected an appropriate space the remainder of this Part deals with the requirements associated with ergonomic layouts of work-stations. Thumbnail summaries of some of the topics covered are presented below:

- **'Links'** – achieving the necessary operational links, such as speech, direct voice communication or paper message passing, between operators.
- **'Normal vs. abnormal operations'** – ensuring that arrangements will support operations under all likely operating conditions including degraded equipment functionality.
- **'Maximum and minimum staffing levels'** – a requirement particularly relevant for larger centres where at different times of the control cycle there may be large fluctuations in the number of staff. The ergonomic solution will allow for the range of predicted staffing levels to be seamlessly accommodated.
- **'Interpersonal distances'** – a requirement which takes account of operational needs for proximity yet psychological barriers to working too closely together.

- **'Social contact'** – the need to allow informal as well as formal communication to occur. An issue of particular concern at night, when workloads are low, and operators are struggling to maintain vigilance.
- **'Supervision'** – the nature of supervision and requirements associated with eye-to eye contact in order to achieve this.
- **'Off-work-station shared displays'** – the use of large displays, typically wall-mounted, which are shared by a number of operators. Ergonomic requirements associated with viewing angles and direct sightlines are covered by the standard.

12.6.4 Part 4 – layout and dimensions of work-stations

This Part establishes general principles for the ergonomic design of work-stations found in control centres. It covers work-station design with particular emphasis on layout and dimensions. The scope includes seated and standing control work positions and maintenance work-stations. At the present time, September 2001, this Part has reached 'Draft International Standard (DIS)' stage.

An early step in the work-station design process is to establish what will be the user population considering such factors as nationality, male and female users and accommodating those who might be in wheelchairs.

The need to allow for 'real' postures, compared to those derived from 'laboratory-type' measurements is discussed. The 'slump factor', for example, is a correction introduced to data collected on erect postures to take account of the more natural relaxed ones found in practice. Control rooms are environments where teams of operators often share a single work-station. This places more critical demands on work-station layout – the standard lays down an approach to reconciling differences between tall and short users.

When considering visual tasks the draft identifies ways of maximising operators' abilities to detect and identify information presented on work-station displays. For the purposes of developing guidelines on this the document assumes four types of seated posture – 'bent forward', 'erect', 'reclined' and 'relaxed'. In the latter the operator is reclined back in their chair and has moved away from the front edge of the work-station. It is recognised that the work-station needs to accommodate this range of postures with key information on the screens still remaining visible. In a major difference with typical office tasks the operator in a control room is likely to adopt a viewing distance to the screen of between 750–1000mm; for office tasks this distance is estimated to be about 500–600mm. The reason for this is that a typical array of equipment on a control work-station is likely to be two to five large screens that need to be viewed in parallel. It is the size of this array, and the need to have an overview of all of them, which pushes the operator away from the screen surfaces – in contrast in a typical office environment the work task is concentrated on a single display. The increased viewing distances will have a direct impact on font size selection so that characters subtend the recommended angle of visual arc at the eye (minimum of 16 degrees of arc).

To establish the ergonomic arrangement of visual displays at a work-station the concept of the 'space of identification' is being developed by the standards writing team. In this approach the operator postures, sizes of displays, character sizes and user sizes are amongst the factors considered.

The geometrically defined space might be considered as looking into the interior of a spherical chamber. The interior surface of the space is made up of all those points in space at which characters are at their limits of acceptable viewing.

12.6.5 Part 5 – displays and controls

This Part of the standard will deal with the selection and use of displays and controls within a control centre environment. The scope includes off-work-station display design as well as the ergonomics of closed circuit television (CCTV) monitor use. In control rooms the range of display technology used at one time is likely to be greater than in other environments – it's not unusual for operators to be faced with VDU displays, liquid crystal display (LCD) screens, hard-wired lamps and switches and a requirement to look out of windows for feedback on plant running. Similarly for auditory displays, annunciators are likely to be fitted to all major systems and voice communication is often a key element in the operators' tasks.

At this stage the material has yet to be developed to Committee Draft stage. The Working Group preparing this draft recognised early on that there already exists material in other ergonomic standards (ISO 9241 – Ergonomic Requirements for Office Work with Visual Display Terminals), which may be directly applicable – navigation and screen layout principles for example.

12.6.6 Part 6 – environmental requirements for control centres

Part 6 of the standard deals with environmental matters such as noise levels, lighting, interior design and thermal environments. The 24-hour operation of most control rooms places special demands concerning temperature regulation, for example where it has been found advisable to offer user control to raise temperatures during the early hours.

The issue of windows has already been introduced. Shift systems often do not link in with 'work, rest and play' cycles for the rest of the family and community. The introduction of a window allows staff to retain a link with community rhythms, to which they will return, and thereby reduce the sense of isolation and loss often commented on by these operators.

This Part of the standard will provide a comprehensive framework for the design of control room interior environments including aesthetic treatment.

12.6.7 Part 7 – principles for the evaluation of control centres

The ergonomic evaluation of the control rooms, during both the design programme as well as at the end of the project, is seen as a key contributor to meeting the standard's more strategic requirements. In this Part, the preferred techniques to be adopted at each stage of the design programme are spelt out. At the time of writing, material is still being assembled for the preparation of a Committee Draft.

12.7 Case study: Railway signalling upgrade and control room ergonomics

The transfer of the signalling functions from a signal box located near Leeds, England railway station to another centre in York raised questions about workloads and team integration. Additionally, the move involved a transfer from 'panel-based' operation, where signals and points are controlled via discrete hard-wired lamps and switches, to screen-based control where the same functions are undertaken using a tracker ball and VDU screen. In seeking answers to the uncertainties a number of human factor techniques were used – the first of these was to carry out a 'task analysis'.

The underlying philosophy behind a task analysis can be explained in a single sentence

> find out what operators are doing now and then apply this information in designing new jobs and working environments.

Figure 12.2 illustrates graphically the type of data that can be derived from a task analysis. Here the components of mental and physical work have been determined for different times of the day. By careful interpolation judgements can be made about future workloads and these compared with known statistics on human performance at various workload levels. Through such an analysis, staffing levels can be matched to predicted workloads – in an auditable way, thereby relieving managers from basing their estimates on judgement only. Complementary task modelling techniques record the processes undertaken and the times required to complete them.

Measurements of physical workloads, particularly where they involve large expenditures of energy, are relatively simple. There are well-established connections between such physiological measures as oxygen consumption and heart rate to estimate the level of effort being expended and, more importantly, remaining capacity. Measurement of mental workload, in comparison, is much more difficult. When taken under 'operational' conditions, measurement techniques must be unobtrusive and not disrupt potentially safety-critical tasks. Direct observation may give some indication of the level of workload though this is, at best, still a 'judgement'. Operators can also be asked at various times during the working day to answer questions about their workloads or to complete crib sheets, although inevitably this will disrupt the very activities being measured. Alternatively, simulators can be used and more intrusive approaches to workload

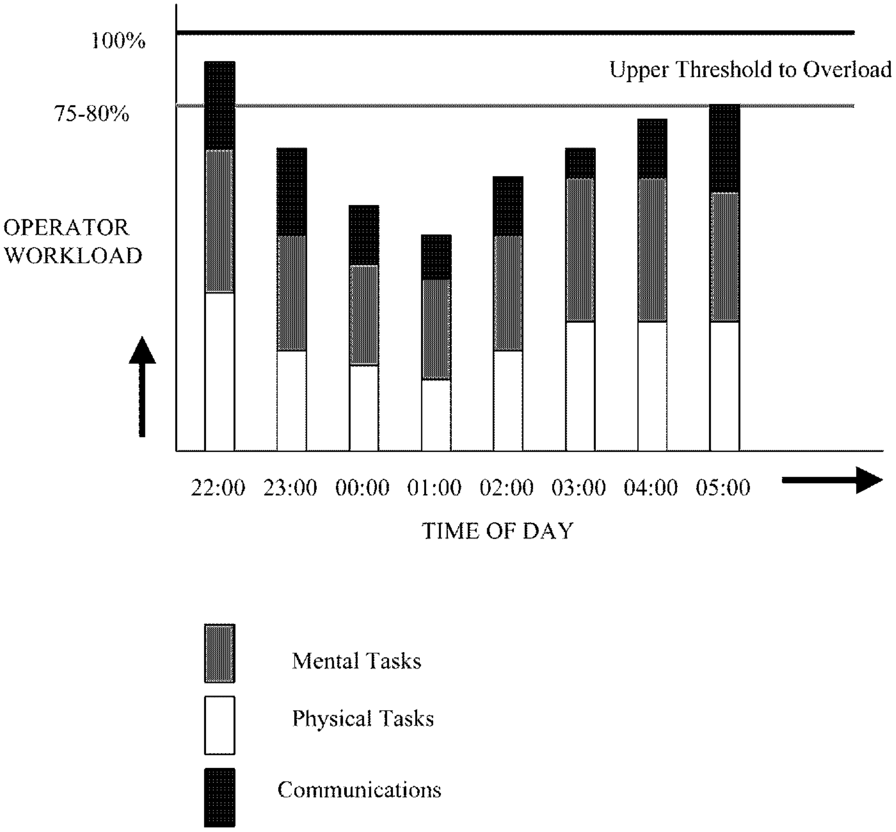


Figure 12.2. Workload modelling.

measurements introduced. At least the disruptive effects will not impact on a live system.

For this programme direct observations were made, at set intervals, and the signallers' activities grouped into pre-determined categories. These judgements were then entered directly into laptop computers for subsequent analysis. Mere classification and measurement of activities, however detailed, unfortunately does not directly provide an answer on potential overloading or underloading at any particular stages of the work-cycle. As a rule of thumb, CCD (the author's organisation) use a maximum acceptable, overall workload figure based on operators spending no more than 75–80 per cent of their time on time-sensitive tasks – the remaining 25 per cent spare capacity is for recovery and undertaking non-time critical tasks such as management and administration. This maximum figure is based on average levels of activity and allows for transient peaks where operators may exceed 100 per cent loadings with time critical tasks but only for short periods. Where continuously heavy workloads are expected, involving high levels of concentration, work periods will need to be carefully structured so they take account of fatigue effects and loss of performance.

Apart from workloads, task analysis also provides information on the way in which equipment will be used, such as task frequency and sequential operation. Predicted use of work-station equipment was used to develop ergonomic work-station layouts. By applying some simple rules – such as high priority items being placed close to the operator and equipment arrangements relating to underlying sequences of use – work-station layouts which would support the operator rather than hinder them in their work were developed. These were mocked up at full scale (see Figure 12.6) and users were asked to comment and criticise. By using mock-ups, alternative layouts are simply and quickly examined. Through this process, agreed work-station layouts were developed (Figure 12.3). This was repeated for each type of work-station.

The next stage of the process broadened the focus to look at how teams of operators might work together and how their work-stations might be grouped to achieve operating needs.

'Link analysis' is a technique used to determine the communication paths that occur between different members of a team. Observations during both busy and quiet periods was carried out to reveal the nature of these paths – direct speech, intercom, paper, visual contact – and the nature of the information being passed. Based on these groupings of work-stations, and options for the layout of the entire room that captured these links, Figure 12.4 presents an example of one of the 'bubble diagrams' which were prepared, in which functions have been juxtaposed to reflect the operating links that need to occur between them. As with any complex room it is inevitable that compromises will need to be made – auditory privacy for

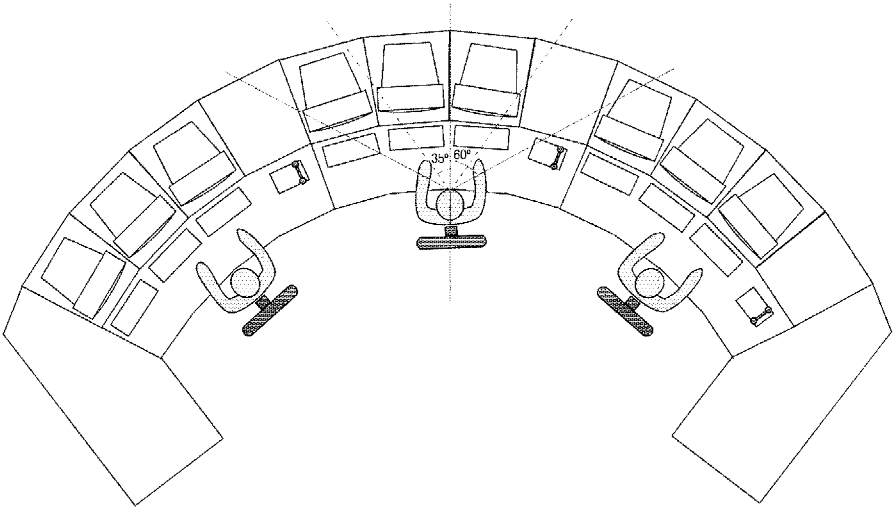


Figure 12.3. Work-station layout.

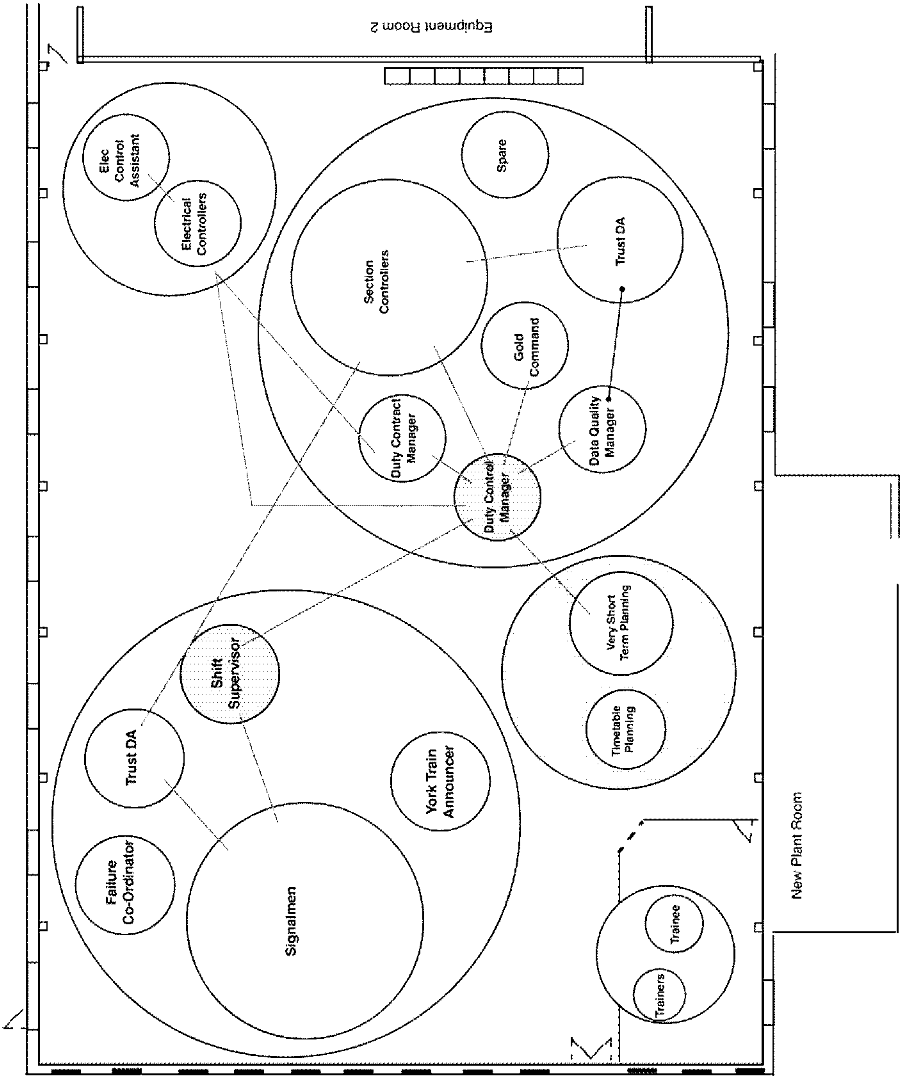


Figure 12.4. 'Bubble diagram'.

individual operators directly conflicts with a need for a supervisor to overhear what team members are doing at any time. A number of these bubble diagrams emerged – meeting in varying degrees the essential operating links identified.

The next stage of the process was to translate these 'bubble diagrams' into potential room layouts by replacing each bubble with the correct numbers and sizes of work-station, Figure 12.5. Overall work-station dimensions were based on those which emerged from the individual work-station design and mock-up trials.

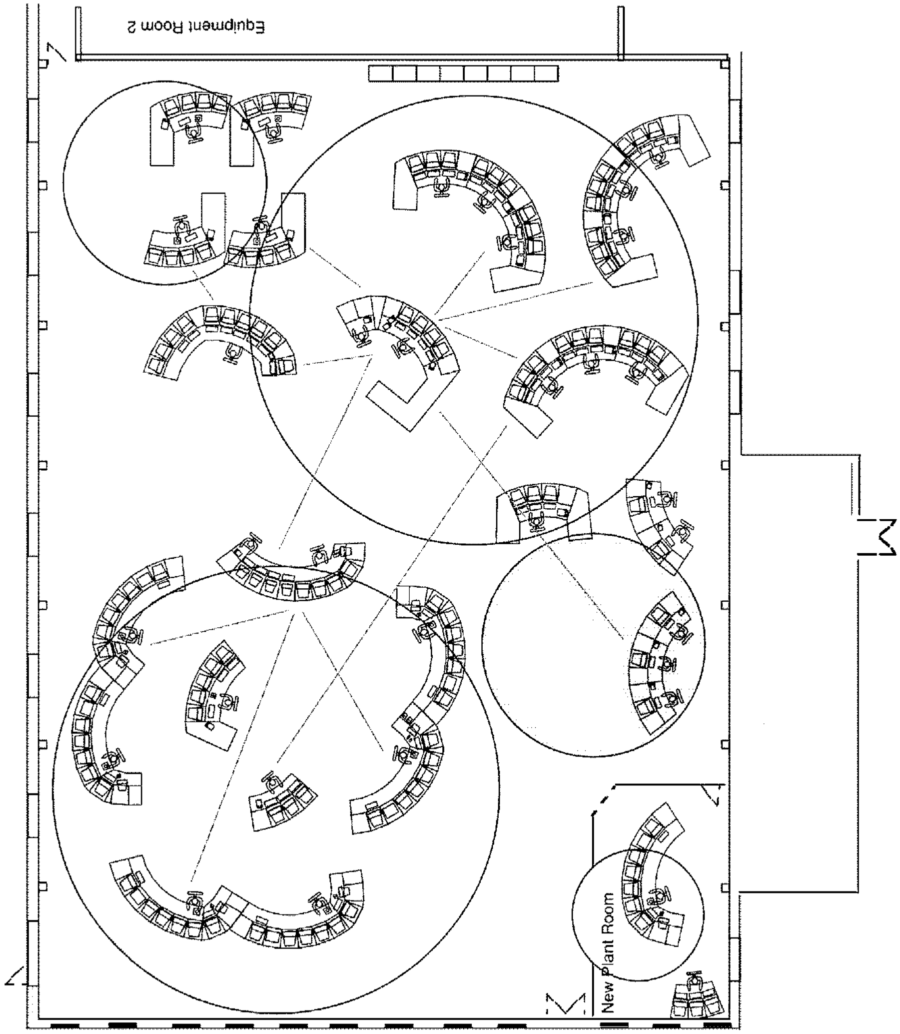


Figure 12.5. *Control room layout.*

The alternative room layout options were again subject to user trials – this time using a full-scale mock-up of the entire control room as shown in Figure 12.6.

The process adopted to determine the most suitable layout was similar to that used for individual work-stations. Users were presented with the control room layouts and were asked to run through simulations covering various operating conditions – normal, degraded and system failure. During these simulations the users were invited to comment on the performance of the layouts and how well they supported the operations being undertaken. Through a process of change and modification initial layouts were refined until all agreed that they represented the

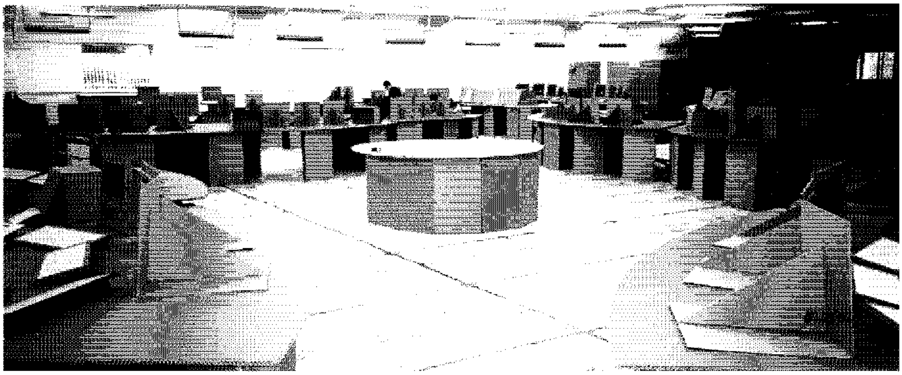


Figure 12.6. Room trials.

best that could be achieved within constraints of equipment provision and building shape and size. This agreed layout is the one now installed as shown in Figure 12.7. Computer-aided-design (CAD) modelling, as shown in Figure 12.8, provides highly realistic impressions of the way in which a room will look from various eye-points. It also provides the opportunity to examine, at an early design stage, what the visitors' first impression on entering the control space might be. However powerful CAD modelling is, it is not a substitute for full-scale mock-ups, since the subtler issues of interpersonal distances and the potential for team working are best explored at full scale. Both techniques were used in the design of the York Integrated Electronic Control Centre (IECC) allowing control staff and managers to gain confidence that the solutions selected would meet their operational needs.

Figure 12.7. Installed control room.

Figure 12.8. CAD modelling.

Simulating proposed systems, with real users, provides an effective and proven method of reducing the risks associated with the introduction of new technology and the design of interfaces. The aviation industry has been using simulation techniques for many years and would not plan to introduce new methods of working without having rigourously checked them under simulated operating conditions. The simulations used for the York IECC were relatively limited in scope, looking at the operating links between operators and the way in which alternative layouts matched the operating demands. Workloads were examined separately though direct observation and mathematical modelling. For future projects it is intended to adapt the techniques used in aviation and bring these two aspects together within a simulation environment. Ergonomists working with users will develop computer interface designs, test workloads and simulate team working under a range of operating conditions. The end result should hold no untoward surprises since high fidelity modelling will have minimised the potential for error.

12.8 Summary

The evidence 'on the ground' demonstrates that the specialist input needed in the planning and design of control rooms is often inadequate. As we place greater burdens on our control centres we can only expect that the pressures on control room performance will increase. Members of the public have a right to expect that control centres, particularly those where life-threatening events are handled on a regular basis, are designed responsibly.

The forthcoming standard will provide a necessary framework for those facing the challenges arising from a new generation of control centre.

Acknowledgements

The author is grateful to Railtrack's LNE Zone for permission to publish the work carried out on the upgrading of the York IECC Signalling Centre.

Chapter 13
Design of alarm systems

Matthew Bransby

13.1 Introduction

This chapter relates to the design of control room alarm systems. Alarm systems are found on many user interfaces to large systems, e.g. in control rooms of power stations or chemical plants, in control centres for railway, road, air traffic or military systems, in aircraft cockpits, on bridges of ships, *etc.*

Alarm systems are important because they are systems that provide stimulus – typically an audible warning – to make the operator aware of an operational problem. They should direct the operator's attention to an abnormal situation so that s/he can take preventative action. Modern processes and systems are often fitted with sophisticated protective systems to prevent hazard to people or major damage. However, action by operators is often very important to correct minor problems before they escalate into major disturbances and also to avoid unforeseen combinations of events where unprotected risks can arise.

Failings in alarm systems have contributed to several major accidents. For example:

- In the Three Mile Island accident (Rubinstein and Mason, 1979) the large number of alarms contributed to the slow response of the operators in diagnosing the problem.
- In the Milford Haven explosion in 1994 (UKHSE, 1997; Bransby, 1998) the operators were loaded with a sustained flood of alarms (one alarm every 2–3 seconds in the five-hour period leading up to the accident). This flood prevented them from fully appreciating the situation and caused critical alarms to be missed.
- In the Channel Tunnel fire (Brown, 1999) the slow response to fire alarms had an adverse effect on the management of the incident.

These major accidents provide obvious examples of alarm system failings. However, there is also strong evidence that large numbers of smaller and less obvious difficulties with alarm systems can have a very significant financial impact (Andow, 1998; Bransby and Jenkinson, 1998).

This chapter provides an introduction on how to develop alarm systems to reduce these problems. It draws very heavily on an HSE Research Report on alarms (Bransby and Jenkinson, 1998) and the EEMUA Alarm Systems Guide (EEMUA, 1999). The latter document provides much more detailed and comprehensive guidance. In order to emphasise practical application, many of the ideas in the chapter are presented through industrial examples.

13.2 Functional requirements for alarm systems

The primary function of an alarm system is to direct the operator's attention towards plant conditions requiring timely assessment or action.

This implies the following principles:

- **Defined response.** Every alarm should reflect an operational condition that the operator should do something about, i.e. every alarm should require an operator response.
- **Usability**. The alarm system interface should make new alarms obvious, should provide a clear indication of the problem and should guide the operator to what s/he should do about it.
- **Workload**. The alarm system should be effective in attracting attention, i.e. it should be designed not to demand too much of the operator's time/attention.

These are important principles, thus:

- The principles should be applied rigourously during the design of alarm systems and of individual alarms.
- Commissioning should demonstrate that the principles are satisfied.
- Audits should be made throughout operational life to ensure satisfactory performance is maintained.

13.3 Designing individual alarms

The bulk of the problems that occur with alarm systems originate from failures in the design of individual alarms. It is essential that the design process is rigourous. Every alarm should be justified and properly engineered. The key questions to answer about every alarm are:

- What is the relevance of the alarm to the operator?
- Is the alarm telling the operator what s/he needs to know?

However, these only provide a start to the design process. There are many other design issues that need to be resolved for each alarm. Some of these are as follows:

13.3.1 Purpose of alarm

- What is the purpose of the alarm? Of what hazard or process risk does it provide warning?
- What is the severity of risk in terms of injury to people, environmental impact, plant damage or financial loss?
- What other protection is there against the risk? Could this fail? Is the alarm safety related? If so, what does that imply?

13.3.2 Operator response

- What response should the operator make to the alarm? (If no response is required from the operator, should this be an alarm?)
- What alarm message does the operator need?
- Does the operator need to access other information to decide how to respond?
- How quickly must the operator respond?
- What written alarm response procedure is needed?
- What stress will the operator be under when this alarm comes up? What are the chances of him/her making the right response?

13.3.3 Prioritisation

- What are the consequences of the operator failing to respond to the alarm?
- Is the alarm time-critical?
- What priority should be given to the alarm?
- Should the priority be the same in all conditions?

13.3.4 Alarm setting

- What are the limits on normal fluctuations of the process variable?
- At what value will hazard or loss occur?
- Should the alarm setting change according to operating conditions?
- What noise or process fluctuations affect the process variable? Could this cause the alarm to annunciate repeatedly?

13.3.5 Suppression

- Will the alarm occur in large plant disturbances, when the plant is starting up, when the plant is shut down, *etc.*?
- Is the alarm operationally significant in all the conditions that it occurs in?
- Will other alarms occur in conjunction with this alarm?
- Could the process variable generating the alarm go 'bad' or out of range?

13.3.6 Management control

- What procedures are required to change the alarm? Who can change it?
- Should the operator be able to 'shelve', temporarily, this alarm?
- How will the alarm be tested and maintained?

A key point is that the process of designing alarms should be driven by the operator's needs, not by engineering features in the equipment. Modern equipment often provides considerable diagnostic information. It is important that this information is carefully assessed in terms of its value as an alarm. Problems frequently arise because alarm signals generated from equipment are simply transferred into operator alarm systems without any analysis or review.

Example 1 Engineering of alarms
A typical modern escalator used in an underground station is fitted with 100–150 diagnostic signals/alarms and will also be monitored through closed circuit television (CCTV) cameras. Automatic protection is provided to stop the escalator, e.g. if objects get caught in handrails or steps. There are also passenger emergency stop buttons and a key-locked local control panel. What alarms are required by the operator responsible for remote supervision of the escalator?

The operator needs alarms to warn if an escalator has been stopped by:

- the passenger emergency stop; or
- the automatic protection

The operator would also want to be able to see on his/her interface if an escalator has been stopped from the local control panel. However, this may not need to be an alarm since the action should only be taken by an assistant working under his/her direction.

If the escalator has been stopped by the automatic protection, the operator needs to know:

- if it is a fault that could be due to a passenger being trapped in part of the escalator;
- if the fault leaves the escalator in a safe state for use as a fixed stairway;

- if the fault is in the internal equipment so is very unlikely to have caused injury;
- whether the escalator can be restarted or whether maintenance is required.

If there is any chance of passenger injury, the operator should look at the escalator on the CCTV monitors and send the assistant to investigate. The operator should also send the assistant if (a) the escalator cannot be used as a fixed stairway and needs to be closed off, or (b) the escalator can be restarted.

If the fault is one needing maintenance attention, then the operator may need to raise a maintenance request. Many of the alarm signals generated by the escalator are indications of what has caused the escalator to trip. It is useful for the operator to be able to pass this information on to the maintainers so that they can bring appropriate spares with them. If technically feasible, it might be preferable to automatically direct this information to the maintainers.

There are a number of further alarm signals generated from the escalator that warn that the escalator is likely to fail in the future and preventative maintenance is required. Again this information is primarily relevant to the maintainers. However, it may have some value for the operator. For example, in some cases the operator can continue to run an escalator, but is prevented from restarting it if it is stopped.

The above analysis indicates that the alarms basically fall into a small number of categories of progressively reducing operational importance:

- emergency stop button pressed;
- tripped, trapped object in (handrail/stair);
- tripped, cannot be used as fixed stairway;
- tripped, restart may be possible;
- tripped, maintenance required (fault details);
- warning, restart unavailable, maintenance required (fault details).

This provides a basis for the design of the alarm messages and prioritisation.

This example illustrates that to design effective alarms it is often necessary to process logically alarm signals to make them more meaningful and to suppress low value information. This logical processing can become quite complex.

13.4 Design of alarm handling systems

Alarms can be displayed in fixed format on discrete alarm annunciators or in a variety of formats on programmable display devices, e.g. visual display units (VDUs). The common alarm graphics on programmable devices are:

- Alarm lists.
- Animated alarm indicators on plant mimics and plant overview graphics.
- Alarm banners or mini-alarm lists on other graphic displays.
- Fixed format fascia-type windows.

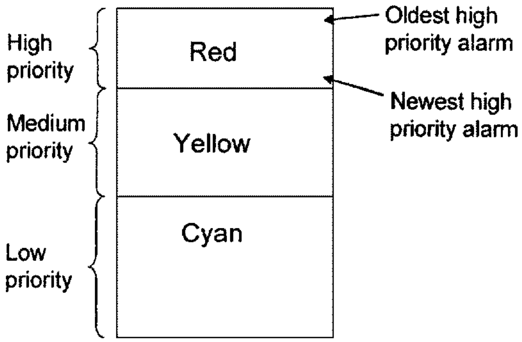
An alarm list is very often chosen as the primary alarm graphic, with the list kept on permanent view. However, recent experimental research at the Halden Institute for Energy Technology (Fordestrommen, 1999) suggests that operator preference is for a three-screen combination (i.e. an alarm list, a fascia-type display and a plant overview animated with critical alarms). These three screens are kept on permanent view – with other screens for selectable control and information graphics.

This chapter will limit itself to a discussion of alarm list displays. More general discussion of alarm display design may be found in (Dicken, 1999).

Numerous different variants of alarm list graphics have been developed by different manufacturers and across different industries. A range of designs seems to be usable and acceptable to operators. Unfortunately there are other designs that are very much more difficult to use. Since it is often commercially unrealistic to change the design of the alarm list handling on many process control systems, it is essential to check before purchase that it meets good practice. Some issues to consider are:

- Can the operator easily keep track of where s/he is in the list or does the displayed view move without any operator action?
- Is it easy to navigate to see new alarms and also to see older unacknowledged alarms?
- Do audible warnings increase stress during alarm floods or can they be silenced?
- Is it easy to select a view of all new high priority alarms in an alarm flood?
- Can nuisance alarms be easily removed from the list by shelving?
- Is it easy to access related controls and information displays from an alarm on the alarm list?

Example 2 Alarm list display



The list above illustrates how a particular alarm list display works. The list stored inside the process control system has been broken into three segments for the high, the medium and the low priority alarms. These are colour-coded Red, Yellow and Cyan respectively. When a new alarm occurs it is entered at the bottom of its section of the list and it pushes all the alarms below it, in the lower priority segment(s), downwards.

The new alarm flashes and sounds an audible warning when unacknowledged. It is acknowledged by clicking on it with the mouse, and this makes the alarm text stop flashing and stops the audible warning. When an acknowledged alarm becomes inactive, it is automatically removed from the list, which repacks.

If an alarm becomes inactive before it has been acknowledged, its colour is changed to white, but the audible warning continues. Clicking on it will silence the audible warning and cause the entry to be removed from the list.

From the above, it follows that the list stored in the process control system contains all the active alarms and all the unacknowledged inactive alarms. The part of the list shown on the operator's display can contain up to 20 alarms. The view is changed using scroll buttons that scroll the view up or down by 20 alarms.

Comments on the usability of this list display include the following.

- **Considerable navigation up and down the list may be needed to see new alarms**. This will be a problem if there are several pages of standing alarms. For example, if the operator was looking at the Yellow alarms and a new Red alarm came in s/he would have to page up to see it. Similarly paging down would be necessary if it was a new Cyan alarm. This paging up and down would not be necessary in a more conventional alarm list that is not priority segmented.
- **The list highlights important alarms in alarm floods**. If there are many incoming alarms the operator can just concentrate on the top of the list, which shows the high priority alarms, and can temporarily ignore the lower priority alarms. This is a strength of this design. However, if priority filters are provided, a conventional list can be just as good.
- **The view can change without any action from the operator**. If the operator is looking at a part of the list and a new higher priority alarm comes in or a standing higher priority alarm becomes inactive, then the list will shuffle up/down. This is undesirable as the list will be hard to read whilst it is moving. Also the operator can 'lose his/her place' if the list changes whilst s/he is looking away.
- **The audible warning can cause unnecessary stress**. The only way of silencing the audible warning is by acknowledging all alarms. This may take some time, since every alarm has to be individually acknowledged and there is the navigation up/down to do as well. It would be preferable to provide a silence button. Then the operator could get rid of the irritation of the noise and deal with the alarms in the way s/he wants. For example, some operators will leave alarms unacknowledged until they have investigated and dealt with the problem causing them to go off.
- **The operator may not be aware of alarms becoming inactive**. Acknowledged alarms that go inactive will disappear off the list without any action from the operator. Whilst it is generally not considered necessary for operators to have to acknowledge alarms that become inactive (because of the workload that it

imposes), there can be benefits in leaving inactive alarms in the list until the operator chooses to clear them off with a list repack.

- **Repeating alarms may flood the list.** In the description given it is not clear how an alarm that repeatedly annunciates will be dealt with. It is possible that it will cause multiple entries to be made in the displayed list. This can be very undesirable and significantly reduce the usability of the list.
- **The flashing alarm text will be hard to read.** Best practice is to flash a marker next to unacknowledged alarms rather than flash the text itself.

An absolutely crucial issue in the design of alarm lists is how repeated annunciations of the same alarm are dealt with. It is very important that the list should not get filled up with repeats of the same message because, in practice, it is virtually impossible to ensure that repeating alarms never occur.

Example 3 Repeating alarms

A level transducer measures the water level in a pool at the bottom of a power station cooling tower. This is shown in Figure 13.1. The tank containing the water is 1000 mm deep. The transducer measures over the range 0–900 mm. Measurements outside this range are set 'bad'. The water level is normally controlled at 700 mm. A high level alarm is set at 800 mm. (See Figure 13.1).

Wind will cause waves in the pool. What effect would these have on the alarms that are generated? What could be done about these alarms?

The waves can cause a repeating high alarm to be generated if the distance of the average water level from the high alarm level is less than the wave height (see Figure 13.2). This alarm can be eliminated by use of deadband. Note that the waves will break through the deadband if they get very large and cause repeating alarms. Breakthrough is almost occurring at point A in Figure 13.3. So deadband will not work in a hurricane!

An important point illustrated by this example is that, even in well-designed systems, it is very difficult to be certain that there are no unusual circumstances in which repeating alarms will not occur. It is essential that alarm list displays remain usable even when flooded with the same one or two alarms repeating over and over again.

The two basic functions that should always be provided in alarm lists for handling repeating alarms are:

- **Shelving** (i.e. a facility for the operator to temporarily block out nuisance alarms).
- **Single line annunciation** (i.e. designing the list so that a repeating alarm only takes up one line in the list display).

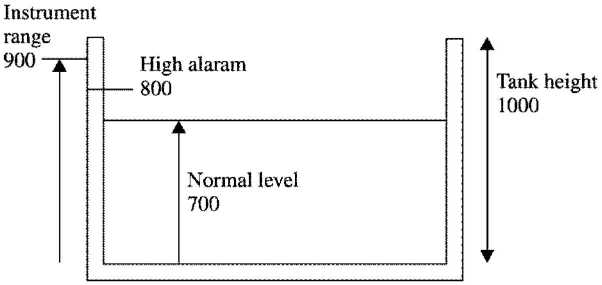


Figure 13.1. Example 3-I.

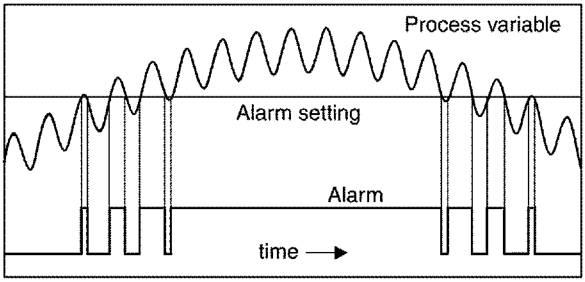


Figure 13.2. Example 3-II.

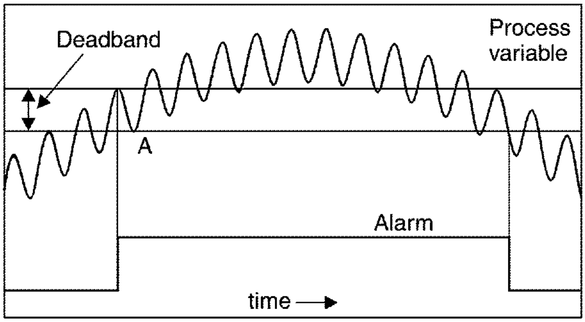


Figure 13.3. Example 3-III.

13.5 Measuring system performance

The examples above have highlighted some aspects of alarm system design. However, following good design practices is not enough; designs also have to be demonstrated to work in real operation.

A key aspect of alarm system performance is the alarm load imposed on the operator. The operator must be able to manage the alarms displayed to him/her as well as deal with all the other operational demands on his/her time.

A simple estimate of the operator workload, W, being demanded by the alarm system is given by the equation:

$$W = R^{*}T$$

where:

R is the average rate at which alarms are presented
T is the average time taken by the operator to respond to each alarm.

To illustrate this equation, suppose that the optimistic assumption is made that, on average, the operator will only take 30 seconds to read an alarm, diagnose its cause, decide how to respond and carry out that response. Then, if alarms are presented at an average rate of one per minute, the alarm system is consuming 50 per cent of the operator's time. This sort of load is unmanageable if it goes on continuously and the operator has many other tasks to do.

In practice many operators are unacceptably heavily loaded with alarms – both in normal steady operation and particularly so in major plant upsets such as plant trips. Operators often get hundreds of alarms in the few minutes after a major upset (Bransby and Jenkinson, 1998).

In order to reduce, systematically, alarm loads and also make alarms more useful it is very desirable to measure alarm system performance and use these measurements to drive an improvement programme. Performance measurements may be made using operator questionnaires, operator logging of problematic alarms, or analysis of logs of the alarms that have occurred (EEMUA, 1999).

Example 4 Analysis of alarm log
The following shows an extract of 30 entries from an alarm log obtained during the commissioning of a railway station. It has been edited to exclude the date and the tag number and to simplify the text messages.

```
06:52:45.03 Air Compressor Pressure HIGH
06:53:00.00 Air Compressor Pressure HIGH ACKNOWLEDGE
06:54:18.65 Stair 11 Door OPEN
06:54:21.55 Stair 11 Door CLOSED
06:54:49.55 Control Room Door OPEN
06:54:55.95 Control Room Door CLOSED
06:56:16.64 Control Room Door OPEN
06:56:23.03 Control Room Door CLOSED
06:56:34.83 Control Room Door OPEN
06:56:41.03 Control Room Door CLOSED
06:56:47.92 Air Compressor Pressure NORMAL
06:56:48.03 Control Room Door OPEN
```

```
06:56:55.13 Control Room Door CLOSED
06:56:57.00 Air Compressor Pressure NORMAL ACKNOWLEDGE
06:57:12.00 Control Room Door CLOSED ACKNOWLEDGE
06:57:18.00 Stair 11 Door CLOSED ACKNOWLEDGE
06:57:42.56 Effluent DISCHARGE
06:58:06.00 Effluent DISCHARGE ACKNOWLEDGE
06:58:12.56 Effluent NORMAL
06:58:21.00 Effluent NORMAL ACKNOWLEDGE
06:59:28.89 Air Compressor NOT RUNNING
06:59:43.00 Air Compressor NOT RUNNING ACKNOWLEDGE
07:00:27.60 Control Room Door OPEN
07:00:34.10 Control Room Door CLOSED
07:01:04.02 Control Room Door OPEN
07:01:10.72 Control Room Door CLOSED
07:02:02.36 Air Compressor RUNNING
07:02:12.00 Control Room Door CLOSED ACKNOWLEDGE
07:02:26.00 Air Compressor RUNNING ACKNOWLEDGE
07:04:45.00 Air Compressor Pressure HIGH
```

From this log can you work out:

- The average alarm rate.
- The most frequent alarm.
- The longest-standing alarm.
- Any regularly repeating alarms.

- **Average alarm rate**. The duration of the log was 12 min (06:52:45 to 07:04:45). There were 30 entries, but 9 were alarm acknowledgements.

  Some of the 21 remaining entries were alarms entering their active state (e.g. door OPEN, pressure HIGH, effluent DISCHARGE, compressor RUNNING), and some were alarms returning to normal state (e.g. door CLOSED, pressure NORMAL, effluent NORMAL, compressor NOT RUNNING). However, it appears from the log that the operator has to acknowledge both the alarm and the return to normal state. Thus there were 21 ''events'' requiring acknowledgement.

  On this basis the alarm rate is 21/12, i.e. 1.8 alarms per min. Research shows (Bransby and Jenkinson, 1998) that this is a high rate that is hard to deal with on a continuous basis. The average rate measured in a survey in the power and chemical industries was about 0.5 alarm/min.

  The data presented above was actually collected during the commissioning of the alarm system. The analysis revealed that the alarm load was likely to be excessive, so work was initiated to review alarms and eliminate those which were spurious or of low value.

  It would be have been preferable to eliminate potential nuisance alarms at an earlier stage in the design by ensuring that a rigourous justification and design process was followed. It would have been necessary to impose this on all equipment contractors and sub-contractors. Unfortunately, however, few purchasers appreciate such issues when they are specifying systems and placing contracts.

- **Most frequent alarm**. The most frequently occurring alarm is 'Control Room Door OPEN/CLOSED' for which there were six active/clear pairs in the logged period.

  Door alarms are needed for monitoring the security of the railway station. However, this particular alarm indicates that someone has come into the control room where the operator is working. The opening of this door is very obvious to the operator, so its display serves little purpose and the alarm should be removed.
- **Longest-standing alarm**. The longest-standing alarm for which there is a HIGH and NORMAL record is the 'Air Compressor Pressure' alarm. This is active between 6:52:45 and 6:56:47. However, the 'Air Compressor NOT RUNNING' alarm that becomes inactive at 6:59:28 clearly stands for longer. There could also be some alarms that remained standing throughout the logged period and thus would not appear on the log at all. It has to be concluded that it is not possible to calculate the longest-standing alarm (or the average number of standing alarms) from this section of the log.

  This problem could apply for even a very long period of log. To overcome the difficulty it is desirable if the logging system is designed to include a regular printout of all standing alarms, e.g. 'a midnight snapshot'.
- **Regularly repeating alarms**. Although it is not very clear from this log, the air compressor alarms are regularly repeating. The compressor operates under automatic control and automatically comes on when the pressure is low and goes off when it is high. This results in a regular cycle in air compressor pressure. Unfortunately, the high-pressure alarm has been set at a level where it is activated by these normal fluctuations. Analysis of the alarm log suggests that the compressor running alarm should be eliminated as it does not require an operator response and that the setting of the high-pressure alarm should be raised.

  Note that the effluent alarms are further examples of indications of normal equipment operation that do not require any operator response. They should be reviewed.

Note that in analysing alarm logs it tends to be easy to collect and analyse lots of data from normal steady operation. However, it is important to also analyse plant upsets in detail. These are the times when the plant tends to be at most risk and when good performance is most beneficial.

Analysis of alarm logs will identify problem alarms. These will then need to be reviewed to see how they should be redesigned to make them more useful. The review process needs to involve all key parties (e.g. operations, engineering and safety representatives) and should follow appropriate modification procedures.

Unfortunately, in practice, alarm system problems are often only revealed after a plant has been designed, constructed and brought into operation. In this case, rather than driving an improvement programme from analysis of logs, it may be preferable to work systematically through all plant sub-systems reviewing the installed alarms and considering their purpose, their priority, their interrelationship, their frequency of occurrence, *etc*.. Organising a review in this way ensures that a logical and coherent approach is followed across all sub-systems.

Even with a thorough review of every alarm it can be difficult to eliminate alarm floods during major plant upsets. In these circumstances many systems can be simultaneously driven through severe transients and simultaneously generate alarms. It is important that the alarm system interface is designed so that the operator can still manage the plant effectively during an alarm flood. Approaches to consider are:

- Minimise the overload by logical processing, elimination of spurious alarms, *etc*.
- Provide overview displays of critical plant safety parameters.
- Display safety critical alarms on individual mimics independent of the process control system so that they cannot become hidden in an alarm flood.
- Provide training and clearly defined procedures for handling plant upsets.
- Ensure that alarms are effectively prioritised so that it is acceptable to ignore lower priority alarms during alarm floods.

13.6 Alarm prioritisation

Alarms should be prioritised according to two factors:

- The **expected severity of consequences** (safety, environmental and economic) that the operator could prevent by responding appropriately to the alarm.
- The **time available** to the operator to carry out the required response.

Written rules should be developed for the prioritisation of alarms and these should be applied consistently to all alarms. It is commonly accepted that using **three priority bands** within any type of display provides effective discrimination without being over-complex.

Example 5 Prioritisation of alarms
Figure 13.4 shows two alarms that indicate failure of redundant power supplies to a critical item of equipment. How should these be prioritised?

- **Issues that influence prioritisation**
  - What response should the operator make to the failure of one power supply?
  - Should the alarm be directed to the maintenance engineer rather than the operator?
  - If the power supplies are fully redundant, what is the consequence of the operator not responding to the occurrence of one alarm?
  - What will be the consequence if both power supplies fail?

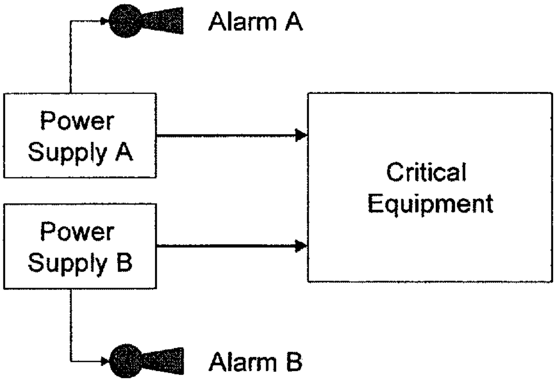


Figure 13.4. *Two alarms that indicate failure of redundant power supplies to a critical item of equipment.*

  - Will the failure of both power supplies result in other alarms?
  - If there are no other alarms, can the occurrence of a second alarm have the same priority as the occurrence of one alarm?

- **Implications of single power supply failure**

  - Consequence: there is no effect on the functioning of the critical equipment. However, there is reduced system integrity and increased risk of the critical equipment failing for the period whilst alarm is active.
  - Operator response: initiate maintenance action (e.g. raise work request).
  - Urgency: the work request would probably be categorised as normal (e.g. repair within a few hours) rather than immediate. Some operator delay in raising the request should not matter.
  - Priority: low – although that does not mean the alarm can be ignored.

- **Implications of double power supply failure**

  - Consequence: failure of the critical equipment possibly causing an immediate financial or safety impact. Other consequential alarms may be generated.
  - Operator response: respond to loss of critical function and subsequently initiate maintenance action.
  - Urgency: high.
  - Priority: high.

This analysis indicates that the single power supply failure alarms should be low priority. However there should be logic so that if both power supply failure alarms occur, these should be suppressed and a single high priority alarm should be generated.

13.7 Conclusions

The examples in this chapter have only provided an introduction to some aspects of alarm system design. There are many important issues left out such as the design of field sensors, risk assessment, safety related alarms, logical processing of alarms, procurement of systems. The EEMUA Alarm Systems Guide (EEMUA, 1999) provides comprehensive advice covering these and many other issues.

The key message to reiterate is that alarm systems are a critical part of many control room operator interfaces. Good design can improve profitability. Bad design can result in accidents, injuries and deaths. It is an area where there are real benefits from following best practice.

References

1 ANDOW, P.: 'Abnormal situation management: a major US programme to improve management of abnormal situations'. Paper presented at IEE Colloquium on *Best Practices in Alarm Management*, London, March, 1998

2 BRANSBY, M. L.: 'Explosive lessons', *Computing and Control Engineering Journal*, 1998, **9**, pp. 57–60

3 BRANSBY, M. L., and JENKINSON, J.: 'The management of alarm systems'. HSE Research Report no CRR 166/1998

4 BROWN, S.: 'The role of alarm systems within safety-related systems – the links with IEC 61508 and IEC 61511'. IBC Seminar on *Alarm Systems*. Paper presented at IBC UK Conferences Ltd, London, March 18–19, 1999

5 DICKEN, C. R.: 'Soft control desks and alarm display', *Computing & Control Engineering Journal*, 1999, **10**, pp. 11–16

6 EEMUA: 'Alarm systems: a guide to design management and procurement'. Publication No 19 1 (The Engineering Equipment and Materials Users Association, London, 1999)

7 FORDESTROMMEN, N. T.: 'Lessons learned in the experimental alarm system CASH'. IBC Seminar on *Alarm Systems* (IBC UK Conference Ltd, London, March 18–19, 1999)

8 RUBINSTEIN, E., and MASON, J. F.: 'An analysis of Three Mile Island. The accident that shouldn't have happened. The technical blow-by-blow', *IEEE Spectrum*, November, 1979, pp. 33–42

9 UK Health and Safety Executive (UKHSE): 'Explosions and fires at Texaco refinery, Milford Haven, July 24, 1994' (HSE Books, 1997)

Further reading

10 BRANSBY, M. L.: 'Alarm systems and the COMAH regulations'. Paper presented at IEE Colloquium *Control of Major Accidents and Hazards Directive: implications for electrical and control engineers*, London, November 4, 1998

11 BRANSBY, M. L., and JENKINSON, J.: 'Alarming performance', *Computing & Control Engineering Journal*, 1998, **9**, pp. 61–7

12 BROADHEAD, N.: 'Optimisation of the Sizewell B alarm system', IEE Colloquium on *Best Practices in Alarm Management*, London, March 1998

13 Channel Tunnel Safety Authority: 'Inquiry into the fire on the heavy goods vehicle shuttle 7539 on 18 November 1996', 1997

14 IBC Seminar on *Alarm systems*, IBC UK Conferences Ltd, London, March 18–19, 1999

15 IEE Colloquium on 'Stemming the Alarm Flood', *IEE Digest*, London, June 17, 1997, **97/13**, (6)

16 IEE Colloquium on 'Best Practices in Alarm Management', *IEE Digest*, London, March 25, 1998, **98/279**

17 JENKINSON, J.: 'Alarm reduction in nuclear power plants: results of an international survey'. IEE Colloquium *Stemming the Alarm Flood*, London, June 17, 1997

18 O'DONNELL, M., and CAMPBELL BROWN, D.: 'Too much of a good thing? Alarm management experience in BP Oil. Part 2: implementation of alarm management at Grangemouth Refinery'. IEE Colloquium on *Stemming the Alarm Flood*, London, June 1997

19 STANTON, N. (Ed.): 'Human factors in alarm design' (Taylor and Francis, London, 1994)

Chapter 14

Decision support in process control plants

Charlotte Skourup and Arthur Aune

14.1 Introduction

As process control plants become more complex, the demands on the operators increase. Operators need information about the process, such as the production goal, the history and ongoing maintenance tasks, as well as experience in operating a plant. It is common practice in most process plant control rooms to provide decision support systems for operators. These decision support systems help and support the operators, e.g. in providing action decision help, state estimation and diagnosis. Existing decision support systems normally generate support and advice based on different types of process models, whereas support systems for transferring individual experience between operators have not been incorporated into the traditional systems.

In today's process control plants, information is continuously generated and stored on different media with much of it never being viewed again. In addition, operators acquire individual knowledge from daily work experience and training sessions that often may be difficult to share with the rest of the operating staff. Such experience is valuable as decision support for other operators.

This chapter gives an introduction to decision support systems for use in process plants. Factors for designing useful systems for decision support, such as situations in which the operators need support and the different types of decision support, are discussed. The latter part of this chapter provides a presentation of a case-based reasoning approach for decision support that has been partly implemented with successful results. Case-based reasoning is a technique to handle incomplete domain knowledge and complex problems. The proposed decision support system also integrates a three-dimensional (3D) visualisation to improve the operator's ability to diagnose and make decisions. Typically, new multimedia and visualisation techniques are used to improve existing decision support systems instead of designing new systems to utilise already existing knowledge such as individual experience and to present more intelligent information.

14.2 Human information processing

Process control operators work in a complex environment with a large variety of tasks. Frequently, they have to make decisions and solve problems at different levels. It is essential to understand how people perceive information input and process information to understand the operators' work tasks and hence, their need for decision support. Rasmussen's three-level model of human cognitive behaviour that is well-known and accepted in many environments describes how information processing of operators takes place at three different levels (Rasmussen, 1983):

- **Skill-based behaviour** is an unconscious process where the action to be performed is a direct result of the input. People often find it difficult to explain a skill-based action such as riding a bicycle.
- **Rule-based behaviour** takes place when the operator recognises the sensory input, i.e. the human has gained experience from previous situations. The input is matched against stored patterns like 'IF-THEN' conditional rules.
- **Knowledge-based behaviour** is the most complicated cognitive process to perform. Information processing at this level is goal-oriented. The operator has to identify observed and unknown input. Based on the input and the overall goal, the operator decides on the task and makes a plan.

An action performed by an operator is related to skill-based, rule-based or knowledge-based behaviour, or a combination of several types of behaviour. Other chapters in the first part of this book describe the human cognitive processes in more details.

14.3 Decision support for operators

The purpose of a decision support system is to assist the operator without replacing him or her, such that there is an improvement in the overall quality and the effectiveness of their work (Aamodt and Nygård, 1995). The operator and the computer-based decision support system have different strengths which can complement each other and hence, improve the overall problem solving process. A challenge is to decide when to provide decision support, the type of decision support, and how much decision support the operator needs.

The design of a decision support system strongly depends on the degree of automation between the operator and the control system. At the lowest level of automation, the operator controls the process manually with no support from computers. At the highest automation level, a fully automatic control system controls a process without interruptions from the operator. In between, the degree of automation varies from open-loop to close-loop control. At each level of automation, a decision support system has different roles in supporting both the operator and the control system. The operator may however not necessarily see a significant difference between the control systems at the middle level of automation. In open-loop control, the control system provides the operator with a basic level of support, e.g. when the control system takes raw data, puts it into a context and presents it as information to the operator. Examples of information presented to the operator are the state of process variables, trends and generated reports. The presentation itself is a type of decision support. Moving towards close-loop control, the control system performs a higher degree of automation under normal circumstances. The operator may however intervene and/or take over the control of the process if s/he judges the process as not being stable or optimal, or if s/he wants to change the goal of the process. In addition, the control system often includes a knowledge-based system (expert system) to provide decision support in the form of problem identification and diagnosing. In abnormal situations, such a system may suggest the cause of the problem and actions to perform.

A decision support system is often integrated in the control system as control systems generally interpret incoming information and serve as an expert system advising the operator on what to do next (Sheridan, 1992).

14.3.1 Situations for supporting the operator

Decision support systems in process control may support the operator in a wide spectre of operational tasks. The focus of this chapter is on decision support during plant operation including the start-up and shutdown phase. Examples of operators' tasks are:

- monitoring (goal-oriented);
- parameterisation;
- diagnosis;
- event and alarm handling;
- analysis of critical incidents and accidents.

Even though all these tasks relate to plant operation, the operator requires different types of decision support related to the specific task and the situation. While monitoring the process and the control system, the operator often has to inspect several hundreds of process variables and the interrelationship between these variables. The operator needs an overview of the situation to estimate how the real process satisfies the planned goal such as quality, quantity and efficiency. In contrast, the operator's primary objective during event and alarm handling is not the planned goal, but the diagnosis of the situation. The operator still needs an overview of the situation in addition to information about the cause and how to solve the problem.

For each of the listed operational tasks, the operator performs several tasks in the interaction with the control system. Figure 14.1 illustrates Sheridan's categorisation of operator supervision tasks (Sheridan, 1988).

Note that the human monitoring must not be confused with the operational monitoring. With regard to Rasmussen's three-level model of human information processing, the operator normally monitors the process and interprets the output based on learnt skills at the skill-based level of behaviour. At the rule-based level, the operator teaches the control system new procedures and intervenes in the process to improve the automatic control, e.g. by changing control parameters. Finally, the operator plans a new operational task and learns from his/her experience at the knowledge-based level of behaviour. Decision support systems must present information that is adapted to the operator's level of behaviour.

14.3.2 Types of decision support and roles of decision support systems

Before the design process of decision support systems starts, the designer must consider issues concerning what type of support the operator needs and how to generate this support. The outcome must then be matched against the typical roles of decision support systems, which are to:

- identify a problem;
- estimate a process state;
- diagnose a situation;
- provide action help;
- support alarm handling.

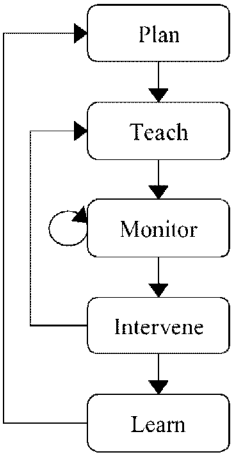


Figure 14.1. Human supervisory functions and their relationships (Sheridan, 1988).

Decision support systems vary in the type of support. Whereas some systems produce a complete plan for the operator including actions to be performed, other systems provide the operator with additional information such that the operator's awareness about a specific situation increases and hence, the operator is able to perform a plan on his/her own. In the first example, the system takes over the operator's tasks and thus, may create a distance between the operator and the process. Such systems have a high degree of automation and may be useful in complex situations such as state estimation in a petrochemical plant with more than 10,000 process variables. The latter type of support systems provide more intelligent support often in combination with a high quality presentation like diagnosis in a power plant (Alty *et al.*, 1995). These systems improve the interaction with operators in that they present information which may widen the operator's situation awareness (Endsley, 1995). As a result, the operator may detect new symptoms, or may be reminded of a previous situation somewhat similar to the new problem situation. This type of system copies the human information-processing process. In practice, however, decision support systems often combine different types of support in order to provide optimal functioning for operators in process plants (Hancke *et al.*, 1994).

As already stated, decision support covers a broad field of systems. A decision support system generally embodies a knowledge base and intelligence for extracting the right content from the knowledge base related to a specific situation. The knowledge base combined with intelligence makes up a knowledge-based system that can solve problems, make decisions, plan or reason about a situation in a problem domain. Often, a knowledge-based system holds a knowledge structure such that it can learn from information input and, in some cases, also update the existing knowledge structure. Such a system does not necessarily copy the human information processing process, but may, for example, generate rules concerning the environment (rule-based system), incorporate prototype cases that cover previous problems (case-based reasoning system) or may learn complex interrelationships between a large number of variables like a black box (neural network) (Mille *et al.*, 1995). Different types of decision support systems correspond to the various requirements for decision support. Therefore, it is important to learn about the operator's tasks, the situations for providing support and the different types of decision support systems.

14.4 Case-based reasoning as decision support

The rest of this chapter contains a presentation of a framework for a decision support system for operators in typical process plants such as power plants (nuclear power plants are exceptions) and chemical plants. The operators work both inside the control room and outside in the plant. Much research carried out to date has been targeted at providing decision support for the operator during critical situations where the consequences could damage the process equipment and/or put human life at risk. This decision support system gives support for operators in unexpected situations defined as:

> A problem situation in which the operator does not have a ready response immediately. Furthermore, at least one observation of the system is unexpected or outside normal behaviour, but the situation is not necessarily critical to the process or the safety. (Skourup, 1999: 162)

Such situations are troublesome and are frequently not solved. The decision support system is based on the case-based reasoning paradigm. The purpose is to reuse already existing information and knowledge in process plants by including an experience repository available to all the operating staff members (Skourup and Alty, 1998; Skourup, 1999). Especially, individual experience acquired during plant operation and problem solving is valuable knowledge that often is difficult to formalise and thus to transfer, for example, to colleagues. The experience repository makes up a knowledge base that contains cases encapsulating individual experience as a significant contribution to future problem solving. The system retrieves cases from the experience repository that matches a new problem situation in some way.

14.4.1 A case-based reasoning approach for fault diagnosis

The decision support system uses the techniques within case-based reasoning to manage individual experience. The case-based reasoning paradigm is providing a method for problem solving. A reasoner retrieves previous experiences similar to a new problem and reuses this experience to solve the new problem. In process control, operators already use the case-based reasoning paradigm, but often without realising it. They retrieve similar experience and adapt the solution to a new situation. In the suggested decision support system, the operator defines a new problem situation by a number of symptoms. A symptom is an observation that contributes to the complete description of the specific situation. The symptoms are stored in a new case in the experience repository. The operator also adds positive and/or negative weights to each symptom reflecting the importance of it regarding the situation context. Later on, when the problem situation is solved, the case must be completed in that the operator, or a domain expert, adds the cause, actions that have been performed and the outcome of implementing these actions. Hence, the content of a case includes both general information as well as individual knowledge such as the weighting of symptoms and personal interpretations of different situations. The case-based reasoning paradigm includes four main steps:

- retrieval;
- reuse;
- revise;
- retain.

However, the steps are performed differently in case-based reasoning systems depending on the purpose of each system. In this decision support system, the system retrieves a set of previous cases that are somewhat similar to a new problem situation. Then, based on a 3D visualisation, the human operator has to chose one, or more, previous case(s) as the best match(es) for the new situation. The operator performs the process of reusing information and knowledge from previous cases. In addition to direct use of the content of a case, the retrieved cases are also used to widen the operator's perception of the particular situation and to suggest new, or different, ways of solving the problem. In the revise step, the operator, or the domain expert, evaluates the suggested solution. Finally, the operator completes the case description of the new situation and stores the case in the experience repository.

14.4.2 The retrieval procedure

In general, the result of a retrieval procedure, e.g. in a traditional textual information retrieval system is presented to the user in an ordered list. As the human problem solver trusts the computer-based decision support system and its ability to perform its tasks, it is a paradox that the first ranked case is not necessarily the best match to the new situation (Bainbridge, 1983). However, the problem solver will, in most cases, chose the first ranked case as the best match without considering the additional cases. Alternatively, humans continue to search for better information based on the result of the automatic search (Hendley *et al.*, 1995). Often, such a manual search is unstructured and the result is unsuccessful in that the human gets lost or may forget the intention of the search. The aim of the proposed 3D visualisation of previous, partially matching cases is to make different aspects of similarity visible to the operator and to visualise the relationships between the retrieved cases. Thus, the operator should have a better basis for deciding which of the retrieved cases is optimal to use as support for the new unexpected situation.

14.5 The importance of the user interface

The user interface performs a key role in supporting the interaction between the operators and the knowledge base in the decision support system. During the last decade, one direction of research on designing decision support systems has been to incorporate different kinds of multimedia such as graphics and 3D visualisations (Hancke *et al.*, 1994; Alty *et al.*, 1995; Walker, 1995). However, the content and shape of user interfaces are often based on traditional rules and concepts for control room designs even though, for example, 3D visualisation opens the way for new and untraditional solutions (Sturrock and Kirwan, 1998). The decision support system presented in this chapter includes a 3D visualisation that aims to visualise previous experience in relation to a new problem situation. The visualisation presents knowledgeable information and is an integrated part of the decision support system. The purpose of the 3D visualisation is to make the relationships between the outcome of the retrieval procedure more visible to the human operator. Thus, the operator's judgement of the best matching case concerning the situation context has added in dimensions. Further, the operator becomes an integrated part of the decision support system. The main purpose of the visualisation is to make the complex relationships among the cases visible to the operator in order to improve the operator's ability to judge the relevance of the retrieved cases.

14.5.1 Visual decision support

Humans are used to perceiving 3D information from everyday operation, and it can be hard to learn how to recognise and interpret representations in two dimensions. During the last few years, the use of 3D modelling and virtual reality has increased considerably as a result of the dramatic advances in the capacity of information technology. The process industry has not been an exception. However, it is not the 3D visualisation itself, but the way it is used that can improve some human tasks, or perform a better support in certain situations. Thus, visualisation aims in bridging the gap between the abstract data world and the real world such as operation in process plants (Walker, 1995). Operators do not need more data, they require more information, which means that data is put into a context such as in an advanced information presentation. Hence, the suggested 3D visualisation is abstract and cannot be modelled in reality. However, humans are very good at recognising patterns, even abstract ones, and at learning just from seeing and observing input (Mones-Hattal and Mandes, 1995).

14.5.2 Visual decision support for case-based reasoning approach

In the case-based reasoning approach introduced above, the operator interacts with the knowledge-based system to perform the matching and adaptation procedures that are parts of the retrieval and reuse steps. The purpose of the 3D visualisation is to present the retrieved cases for the operator so that s/he is able to make a better judgement with regard to selecting which information input will help attain their goal of performing the task (Endsley, 1995). Thus, as the visualisation presents more intelligent information, the operator may then be able to solve the situation without any further support or one of the previous cases may represent a similar situation directly. The dimensions of the three axes together should reflect the relationships between different viewpoints.

In this decision support system, a case is complex and rarely includes only three features. Therefore, each of the three dimensions in the visualisation should represent a complex measurement that expresses the relationships between the new case and a retrieved case. The problems of visualising these relationships is similar to the visualisation of information in traditional information retrieval systems. In case-based reasoning, cases have the same role as documents whereas features are similar to terms. Thus, translated from information retrieval systems based on (pure) text documents, a 3D visualisation should present the relationships existing among cases, among features, and between cases and features (Krohn, 1996).

14.5.3 Metaphors used in visualisation

The proposed 3D visualisation makes use of metaphors such that the visualisation becomes easier for humans to perceive and understand. First, in everyday life, humans are used to viewing and judging visual input with regard to some sort of landscape or ground (Walker, 1995). Hence, this visualisation is based on a landscape plane in the xy-plane. Such a visualisation enhances the 3D effect, even when the landscape is projected in perspective onto the two-dimensional (2D) graphic screen. Figure 14.2 illustrates the landscape plane and the xyz-axes and their respective origins.

Second, objects representing retrieved cases are located at this xy-plane. The new problem situation is not illustrated directly in the 3D visualisation since it is the relationships between experienced situations and the new situation that are of interest. However, the new unexpected situation is always indirectly located nearest to the user that is in front of the landscape.

The objects should visualise simple patterns which humans are able to recognise and compare. Hence, we use signs as metaphors to represent cases. Signs are well known to all people. Furthermore, a sign has qualities that can be extensively used for visualising the complexities of the cases. Each board represents a symptom describing the situation. As the total number of symptoms is known in advance, their locations including both the angle between the boards and the distance from the top are predefined. The symptoms within the same group are located at the same distance from the top and next to each other. A sign representing a specific case includes boards for each registered symptom. These boards are then placed according to the predefined angle and distance for the specific symptom. Figure

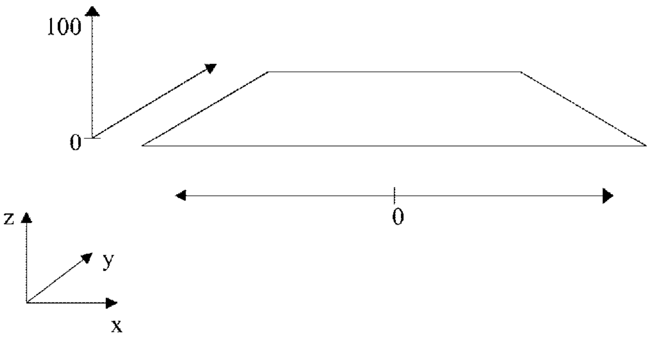


Figure 14.2. The landscape plane and its dimensions.

14.3 illustrates an example of a sign representing a case that consists of six symptoms.

When the operator clicks on one of the boards, the name of the symptom appears. The pattern constituted by the boards should give the operator an overview of the content of each case such that the operator can perform a rough comparison of different cases. However, the content of previous, somewhat similar, cases may be valuable in the problem solving process. Therefore, the operator can double-click on the sign and detailed information concerning the case appears.

Furthermore, a sign is coloured with regard to its physical location in the process plant. Such a physical location is related to the cause of a problem situation. Examples of physical locations are high-pressure preheater and turbines. Next to the landscape visualisation is found a list including different locations in the plant and their colour codings. Hence, the operator can tell at a glance where in the process plant that a previous case has taken place.

14.5.4 Dimensions

The distance from the user to an object should reflect the similarity between the new unexpected situation and a retrieved case. To create a feeling of closeness between the new situation and the retrieved cases, the previous cases having most in common with the new case are located closest to the problem solver. As the visualisation only illustrates the retrieved cases in the landscape plane, the ideal previous case is placed in front of the user. As stated earlier, the 3D visualisation presents the relationship among problem situations, among features of these situations and between a situation and its features along the x-axis, the y-axis and the z-axis, respectively.

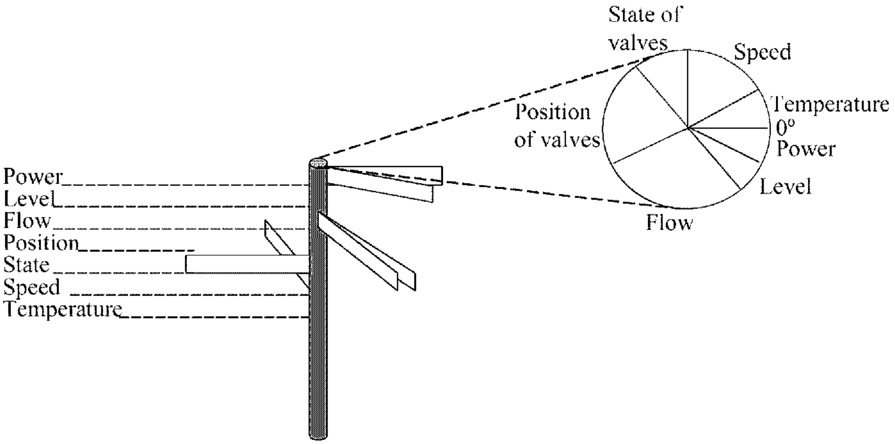


Figure 14.3. Example of a sign metaphor and its symptoms.

14.5.4.1 Relationships among cases

The x-direction represents the relationship among cases, i.e. the number of common features between two cases. If the intersection between the features of a new case, *P*, and a retrieved case, *Q*, includes all features of both cases, then *Q* is identical to *P* in the x-plane. Note that the values of the features are uninteresting for this dimension and hence, these are not necessarily similar. The left-hand side of the x-axis is the number of features in the retrieved case, *Q*, that does not exist in the new case, *P*, i.e. the number of elements in the set *A* where *A* is:

$$A = \{f \mid f \in Q \wedge f \notin P\} \tag{14.1}$$

Where:

f is a feature

If all features of *Q* are a part of *P*, then the right-hand side of the x-axis is the number of features in *P* that does not exist in *Q*, i.e. the number of elements in the set *B* where *B* is:

$$B = \{f \mid f \in P \wedge f \notin Q\} \tag{14.2}$$

As the operator can update the content of the new problem situation at any time, a previous case located at the left-hand side of $x = 0$ can become a better match for the new situation in the x-direction if one of the features in *Q* also becomes a feature of the new case. Similarly, a previous case located right of $x = 0$ is, in theory, also able to move closer to $x = 0$ if one of the features existing in *P*, but not in *Q*, is deleted from *P*. This situation is more an exception than a rule. Normally, the location of a retrieved case moves to the right as the operator extends and completes the description of the new problem situation.

14.5.4.2 Relationships among features

This dimension focuses on details of the individual features that the new situation and the retrieved case have in common. This measurement reflects the normalised similarity between these features as:

$$Sim_f(P,Q) = Sim(P,Q)/n \tag{14.3}$$

Where:

$Sim_f(P, Q)$ is the normalised degree of similarity between all common features in P and Q

Sim(P, Q) is the degree of similarity between all common features in P and Q

n is the number of common features

The measurement compares the values of the features taking into consideration the different types such as numerical types, enumerated types and textual strings. If all common features between two cases are equal, the retrieved case is located at $y = 0$, i.e. in front of the landscape model and hence, closest to the problem solver. As the values of the features become more distinct, the retrieved case is located far away from the problem solver. Hence, the connection between similarity and closeness is reflected in this dimension as well.

14.5.4.3 Relationships between cases and features

The z-axis represents the complex relationship that exists between the retrieved case and its features compared to the new unexpected situation. However, this dimension reflects the closeness function between two cases. The closeness function takes into consideration the weights of each feature with regard to the situation context. The operator decides these weights as the situation goes on. The closeness function is given as:

$$SIM(P,Q) = \alpha \, . \, Sim(P,Q) + \beta \, . \, 100 - Dsim(P,Q) + \gamma \, . \, Psim(P,Q) \qquad (14.4)$$

Where:

$\alpha = 50$

$\beta = 25$

$\gamma = 25$

Dsim(P, Q) is the difference between all common features in P and Q where Sim(P, Q) $<$ *50*

Psim(P, Q) is the number of features where Sim(P, Q) $\geq$ *50 divided by the number of features in case P*

A previous case that is identical to a new situation has $SIM(P, Q) = 100$ and hence, the object is full high in the z-direction, that is $z = 100$. Features that exist in *P*, but not in *Q*, are ignored in the closeness function.

The higher the score is between two cases, the better is the match. This relationship is visualised as the height of the sign representing the previous case. The relationship between the similarity and the high corresponds with the human perception of high objects; these seem to be more important compared to smaller objects.

It is not trivial which of the retrieved cases is most similar to the new situation. The landscape visualisation visualises the relationships between the different cases in three dimensions. It is then the human operator's responsibility to decide which one, or more, of the cases can be partially adapted to the new problem situation. However, a previous case is also valuable for the new situation if it triggers parts of the operator's knowledge such that the operator's situation awareness is extended, or the operator is reminded of other experiences that may be helpful (Endsley, 1995; Mones-Hattal and Mandes, 1995).

14.6 Case study

This section gives an example of the 3D visualisation where the visualised situations are taken from a coal-fired power plant. Data used for the case study is

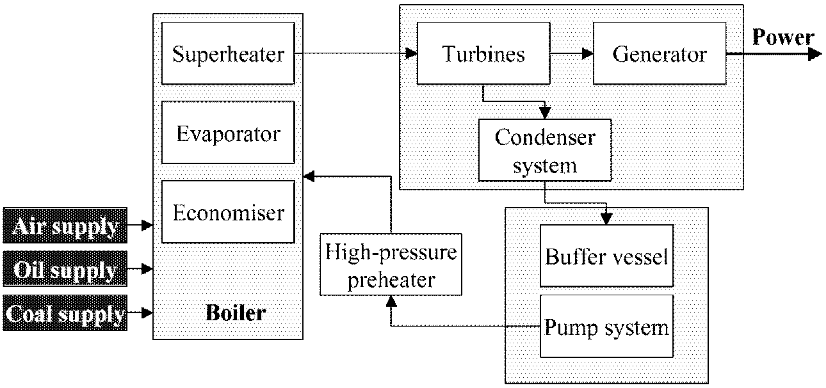


Figure 14.4. Overview of coal-fired power plant.

logged from a power plant simulator at Clausthal University, Germany. Figure 14.4 illustrates an overview of the plant where each of the boxes represents the various physical locations that a cause can be related to.

In this example, the operator discovers a new problem situation that he/she is unable to solve immediately. He/she makes a search in the experience repository to get support on which problem he/she is facing and how to solve the problem situation. Table 14.1 shows the symptoms that the operator has observed.

The search results in five matches that are accepted as being of interest for this unexpected situation. Figure 14.5 illustrates the 3D visualisation of this search.

The operator can navigate in the solution space and get detailed information about similar unexpected situations. The content of a case provides decision support at two levels:

- Direct support in form of directly transferable solutions.
- Indirect support as new information input may direct the operator's focus on other aspects of the situation.

Table 14.1. Set of symptoms for increasing power generation.

Symptom	*Value*
Power	750 MW
Power	Increasing
Level in high-pressure preheater 60	Increasing
Level in high-pressure preheater 70	Strongly increasing
Flow – feed water pump 1	Strongly increasing

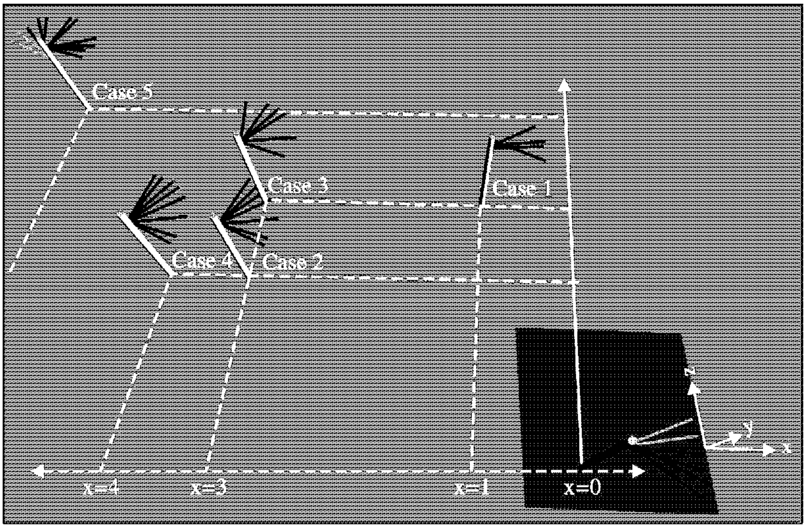


Figure 14.5. 3D visualisation of retrieved cases.

14.7 Conclusions

This chapter has introduced essential factors of decision support systems that are important to consider when designing such systems for operators in process plants. Basically, a good insight into the process and its behaviour, the operator's tasks and behaviour, and the interaction between the operator and the process are significant for a successful decision support system.

Decision support covers a wide group of systems that function differently. This chapter has reviewed basic principles for decision support systems. In addition, an unconventional decision support system based on the case-based reasoning paradigm has been presented. This system focuses on the reuse of individual operator experience as support for colleagues in unexpected situations. The decision support system also includes an advanced 3D visualisation. The visualisation integrates the operator in the decision-making process by making complex relationships between cases visible to the operator. The purpose of this decision support system is to present experience that can be used directly to make decisions in the problem solving process and to widen the operator's situation awareness.

References

1 AAMODT, A., and NYGÅRD, M.: 'Different roles and mutual dependencies of data, information, and knowledge – an AI perspective on their integration', *Data and Knowledge Engineering*, 1995, **16**, pp. 191–222

2 ALTY, J. L., BERGAN, J., and SCHEPENS, A.: 'The design of the PROMISE multimedia system and its use in a chemical plant', *in* EARNSHAW, R. A. (Ed.): 'Multimedia systems and applications' (Academic Press, London, 1995)

3 BAINBRIDGE, L.: 'Ironies of automation', *Automatica*, 1983, **19**, pp. 775–9

4 ENDSLEY, M. R.: 'Towards a theory of situation awareness in dynamic systems', *Human Factors*, 1995, **37**, pp. 32–64

5 HANCKE, T., HERBST, L., ARNOLDT, C., and HINZ, W.: 'New tendencies in man-machine systems for process control'. Proceedings of the ASME Joint International Power Generation Conference, Phoenix, AZ, USA, 1994

6 HENDLEY, R. J., DREW, N. S., WOODS, A. M., and BEALE, R.: 'Narcissus: visualising information'. Proceedings of the Information Visualization Conference, 1995, pp. 90–6

7 KROHN, U.: 'Visualisation for retrieval of scientific and technical information'. Unpublished Ph.D. dissertation, Technical University of Clausthal, Papierflieger: Clausthal-Zellerfeld, Germany, 1996

8 MILLE, A., CHARTRES, J. M., NIEL, E., FUCHS, B., and CHIRON, B.: 'Intelligent work-station for immediate decision helping in process supervision: a new paradigm?', *Computers Chemical Engineering*, 1995, **19**, pp. S815–20

9 MONES-HATTAL, B., and MANDES, E.: 'Enhancing visual thinking and learning with computer graphics and virtual environment design', *Computers and Graphics*, 1995, **19**, pp. 889–94

10 RASMUSSEN, J.: 'Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models', *IEEE Transactions on Systems, Man, and Cybernetics*, 1983, **13**, pp. 257–66

11 SHERIDAN, T. B.: 'Task allocation and supervisory control', *in* HELANDER, M. (Ed.): 'Handbook of human-computer interaction' (North-Holland, Amsterdam, 1988)

12 SHERIDAN, T. B.: 'Telerobotics, automation, and human supervisory control' (MIT Press, Cambridge, MA, 1992)

13 SKOURUP, C.: 'Operation support in process plants: a framework for utilising an experience repository'. Unpublished Ph.D. thesis, Norwegian University of Science and Technology, Report no. 99-8-W, 1999

14 SKOURUP, C., and ALTY, J. L.: 'A case-based reasoning approach to fault diagnosis in control rooms'. Proceedings of the 7th IFAC/IFIP/IFORS/IEA Symposium on *Analysis, design and evaluation of man-machine systems*, Kyoto, Japan, 1998, pp. 221–6

15 STURROCK, F., and KIRWAN, B.: 'Interface display designs based on operator knowledge requirement', *in* HANSON, M. (Ed.): 'Contemporary ergonomics' (Taylor & Francis, London, 1998) pp. 280–4

16 WALKER, G.: 'Challenges in information visualisation', *British Telecommunication Engineering*, 1995, **14**, pp. 17–25

Chapter 15

Train controllers, interface design and mental workload

Wendy Macdonald

15.1 Introduction

Train controller workload levels are of interest from various viewpoints (Lenior, 1993; Neerincx and Griffioen, 1996). Workloads that are too high or too low are likely to cause a deterioration in the quality of controllers' situation awareness, strategic planning and decision-making performance. Deterioration in controllers' performance quality may in turn affect operational efficiency and safety, to an extent dependent on the 'fail-safe' nature of the control system. Workload levels also affect controllers' personal well-being, indicated at an individual level by measures of increased stress and perhaps lower job satisfaction, and at an organisational level by increased absenteeism, staff turnover and associated costs.

Workload levels need to be managed in a variety of ways. Most basically, design of the control system itself, particularly the controller-system interfaces, should take proper account of the nature and level of task demands that system design imposes on the controllers operating it. If the system is significantly modified in some way, for example by automating some of the controller's tasks, possible effects on controller workload need to be assessed. Most immediately, such assessment provides information for designers to use in optimising system interfaces. Just as importantly, it enables managers to estimate future staffing requirements, since workload levels are an important consideration in determining how much territory or traffic is allocated to each controller. Other issues that should be resolved in the light of information on controller workload levels include controller job design, particularly design of shift rosters, rest break regimes, possible limitations on shift duration and limitations on total hours to be worked in a week.

All of the above issues were raised during the project from which data are reported in this chapter. Focus here is on the analysis of train controllers' work and workload to determine the effects of introducing a more automated control system. The project was conducted in Melbourne, Australia, at 'CENTROL' – the control centre for all trains in the state of Victoria, except trains on the Melbourne suburban passenger network. CENTROL traffic includes some long-distance passenger trains but the majority are freight trains. By international standards, traffic levels are light to moderate.

One of the main control systems used at CENTROL to maintain safe separations of trains was a paper-based system termed 'Train Orders' (TO). This system relied entirely on transmission of information between controllers and field staff (usually train drivers), by means of radio or telephone conversations. A more automated control system termed 'Alternative Safe Working (ASW)' was in the process of being implemented in some areas. When fully implemented, the ASW system was expected to cover a large part of the state, with each controller work-station handling a larger area and more traffic than with TO. This was expected to require fewer controllers, based on the assumption that controller workload per train controlled would be much lower with the more automated ASW system than with TO. However, the extent of the presumed decrease in workload was uncertain, and controller workload was seen as a factor limiting the amount of territory which should be allocated to each work-station.

Accordingly, the present study compared TO and ASW systems in terms of their relative workload levels. This was done by a qualitative analysis of task demands and related aspects of interface design in terms of their probable effects on controller workload and situation awareness, and by determining quantitative differences between the two systems in controllers' mental workload. Before presenting results from these investigations, some underlying conceptual issues are outlined.

15.2 Mental workload and related factors

The term 'workload' is commonly used to refer simply to the amount of work which has to be performed, but the more technical meaning of the term within the field of ergonomics/human factors engineering is somewhat different. According to a recent text on this topic:

> The term *workload* has intuitive meaning for most people; everyone has experienced periods of high or low workload in their daily life in response to different situations ... However, ... operational definitions proposed by psychologists and engineers continue to disagree about its source(s), mechanism(s), consequences(s), and measurement. Furthermore, although workload and performance are clearly related, it has proven to be a much more complex relationship than originally thought. (Huey and Wickens: 1993)

Despite disagreement about definitions, there is general consensus among researchers and practitioners in the field that workload level is a product of the *interaction between operator and work task characteristics* (Hart and Staveland, 1988; Huey and Wickens, 1993; Tsang and Wilson, 1997). In the present study, the focus was on the influence of *work task* characteristics on workload levels. Individual controller characteristics were not of direct interest, since all controllers were assumed to be appropriately selected and trained, and the same controllers were employed on both systems.

A key concept relevant to the mental workload of train controllers is that of 'situation awareness', which forms the basis for controller decision-making and performance. According to Endsley's widely accepted definition:

> 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 and Rodgers, 1994: 6)

Situation awareness is dependent on the controller's 'mental model' of system operation. An accurate, detailed and comprehensive mental model of system functioning incorporates knowledge of functional relationships between different task elements and events, their varying probabilities and associated contingencies. Controllers' situation awareness is continually updated from a range of information sources.

Increased automation may reduce the total number of discrete tasks to be performed by the controller, or reduce the total time required, but it cannot necessarily be assumed that mental workload will decrease as a consequence (Sarter *et al.*, 1997). With low levels of automation, controllers have a more 'hands on' role and are likely to obtain most or all of the information needed to maintain their situation awareness as a by-product of their control activities. With increased automation, however, they may need to proactively seek out some types of information in order to maintain their situation awareness. Thus, a reduction in workload associated with a reduction in required control actions in a more automated system might be offset to some degree by an increase in workload associated with the need to be more proactive in maintaining situation awareness. That is, with increased automation the *nature* of workload might change significantly – perhaps more than its overall *level*. Design of the controller-system interface should facilitate the maintenance of controllers' situation awareness with minimal increase in their workload. To this end, it may sometimes be desirable for controllers to perform tasks that *could* be done automatically but whose performance serves to maintain a comprehensive awareness of system status.

Tasks differ in terms of the level of conscious attention they demand – a variable that is not necessarily obvious to an uninformed observer. Some tasks appear complex to an observer but may actually be performed in a fairly automatised fashion, demanding relatively little attention; other tasks might entail fewer observable responses but more complex decisions and concentrated attention – imposing a heavier workload. The 'Skills-Rules-Knowledge' (SRK) performance model classifies behaviour according to whether it is skill-based, rule-based, or knowledge-based – with respectively low, moderate and high levels of attentional demand. Rasmussen (1987) and co-workers developed this model during the late

1970s and 1980s during pioneering research on the activities of supervisory controllers in large chemical and power plants. It has been widely applied in evaluating human–machine-system interactions in complex environments such as large control rooms, especially with respect to the design of information display and control systems (Woods *et al.*, 1987; Wilson and Rajan, 1995; Sarter *et al.*, 1997).

At a knowledge-based level there is active decision-making with full attention required, while at a rule-based level people recognise familiar situations and apply appropriate rules, requiring some attention. At a fully skill-based level, automatised performance routines are 'run off' in response to learned cues, with little or no attention required. The potential for performance of a given task to become rule- or skill-based depends on the consistency of its component stimulus-response relationships; the more invariant these relationships, the more automatised performance can become. For train controllers, even with highly routine activities there will always be some knowledge-based control, since they need to monitor overall system performance and assess whether goals are being met. However, control systems might differ in the extent to which they demand knowledge-based rather than more automatised rule- or skill-based performance.

Another factor that can influence workload is the sensory quality of information presented via system interfaces. Displayed information should be well above operators' sensory thresholds and have good signal to noise ratios, in order to minimise the attentional resources required to perceive it. Interface design also influences controller workload by the particular combinations of information modalities and codes that are used for particular tasks. In accord with Wickens' (1987) multiple resource theory, mental workload will be lower when information inputs are distributed over different, non-conflicting stimulus modalities (e.g. requiring both looking and listening), and when responses are via the most compatible response modalities (e.g. oral response to auditory input and manual response to visual input). Within each modality, the form in which information is coded – categorised as verbal or visuo-spatial – also has an effect: verbally coded information is most compatible with auditory input and oral output modes, and spatially coded information is most compatible with visual input and manual output modes. These interface design factors particularly influence controllers' capacity to perform different sub-tasks concurrently, thus influencing their efficiency as well as workload.

15.3 Identification and analysis of specific workload sources

In addition to their ongoing tasks of maintaining situation awareness and strategic planning, controllers perform a range of specific control tasks. Prior to the present study, staff of the Public Transport Corporation had documented details of all such tasks, together with the average times required to perform each of them. Taking this information as a starting point, TO and ASW control systems were analysed and compared for each of the 10 main tasks related to the primary controller functions.

The demands of each task as they relate to controller workload and situation awareness were identified for each task separately.

The four most centrally important of these tasks were ‘issue train order or authority’, ‘establish authority queue’, ‘fulfil train order or authority’, and ‘field location reports’. Another six tasks were less central but critically affected system functioning: grant/refuse track permission, accept return of track permission, issue absolute occupation, return absolute occupation, enter advice of gang under protection, and record completion of protection. These 10 tasks were analysed in detail. A variety of subsidiary tasks, such as inputting data on train loads to either train graphs or a computer database, were excluded from detailed analysis since these had less direct impact on system performance. Analyses of the four most important tasks are presented below.

15.3.1 Task 1: Issue train order (TO) or authority (ASW)

Performance of this key task gives a train driver permission to proceed to the next section of track.

Performance Times: TO = 180s; ASW = 10s.

- **Cue(s) for task initiation**

  TO

  Cue (1) – Own situation awareness (particularly of train movements, outstanding and future orders/permissions, taking account of potential conflicts both current and future, including consideration of all track-related activities).

  Cue (2) – [If too busy for Cue (1)]:
  Request via radio/phone, typically following fulfilment of TO.

  Auditory signal from: *either* phone *or* locomotive radio (less intrusive signal than phone)

  ASW

  Cue: Return of Authority

  With ASW the preferred strategy is simply to wait for the return of authority and issue the next one at that point. This difference from TO may contribute to a greater sense with ASW of being ‘controlled by the system’ (as some controllers commented) rather than being ‘the controller’. A relatively low sense of control can increase stress levels, thus increasing perceived workload.

- **Information modality and quality**

  TO

  Planning: Visual (paper graph); quality OK
  Writing: Visual (order book); quality OK

Transmitting: Auditory + visual (talk to driver; read aloud); visual quality OK; auditory quality usually OK but can be a problem if radio transmission problems
Feedback: Auditory (driver reads back); quality can be a problem if radio transmission problems

ASW
Planning: Visual (electronic graph); quality OK
Keying: Visual (screen); quality OK
Transmitting: Visual (screen) + some auditory via alarms; quality OK
Feedback: Visual (screen) + some auditory via alarms; quality OK

- **Required performance components**

TO

If Cue (1) – Controller-initiated:

Plan in relation to situation awareness – strategic decision-making:
Plot trip line on graph
Prepare train order: Write out details in block capitals
Transmit to driver:
Make call to driver
Dictate train order to driver, spelling out location names in full, each number separately, *etc.*
(Writing and transmitting of order often done concurrently)
Feedback from/to driver: Listen and confirm as driver reads back; issue a readback time

If Cue (2) – Driver-initiated

Decide priority of cue and whether to respond (strategic decision-making)

When workload allows, most controllers prefer to initiate this task (Cue 1). Waiting for a request (Cue 2) subjects them to an intrusive auditory cue, interrupting other activities and thus increasing perceptual and decision-making demands, or potentially increasing stress because of the need (if busy) to keep the caller waiting. However, the controller may 'answer' the phone or radio by lifting the receiver and placing it on the desk, or pushing a radio button to stop the ringing/buzzing and enable the caller to hear ongoing activities, until the controller can attend to the call.

Answering the radio/telephone:

Converse with driver, decide priority and whether to deal with immediately (strategic decision-making: decisions to defer action until later increase subsequent memory load and possibly also stress; on the other hand, a decision to act immediately means that planning and preparation of the order are performed under greater time pressure than might be the case if deferred.)

If deal with now:

Plan in relation to situation awareness – strategic decision-making: Plot trip line on graph *or* make notes on graph and defer plotting until later – memory load.
Prepare train order: Write out details in block capitals (greater mental demand when being done concurrently with driver conversation)
Then as above for Cue (1)
ASW
Plan (in relation to situation awareness):
Prepare authority:
select train
select authority
Propose transmission
Monitor screen to see accepted

- **Response modality**
  TO
  Oral, manual (paper and pencil)
  ASW
  Manual (keyboard/mouse); possibly also oral

- **Comments**
  The strategic planning and decision-making which is the basis for both TO and ASW operations are mentally demanding, requiring attention to a complex set of factors. In this respect, both systems are similar in workload determinants. The actual transmitting of both TO and authorities and obtaining feedback is routine, 'rule-based' performance, although requiring significant attention at the time but with lower mental demands than planning. With ASW, signal transmission is very much faster (via the work-station rather than voice).

15.3.2 Task 2: Establish authority queue (ASW only)

Establishing an authority queue entails the preparation and queuing of consecutive authorities for execution, as appropriate, by the system; it is performed only in ASW.

Performance Time = 30s.

- **Cues for task initiation**
  Cued by Controller's own situation awareness and strategic planning of train movements and track permissions/occupations.

- **Information modality and quality**
  As for Task 1 – ASW

- **Required performance components**
  Plan in relation to situation awareness – strategic decision-making
    After authority issued and train still selected:
  Prepare authority queue
    select authority queue
    select extended authority proposal
    select last location of queue
  Propose
  Check graph updates

- **Response modality**
  Manual (mouse/keyboard)

- **Comments**
  Choosing to establish an authority queue is a strategy, available only in ASW, which decreases the controller's overall workload since authorities can be prepared when convenient to the controller rather than when drivers relinquish their previous authorities. However, use of this facility decreases the frequency of automatic updates to the controller's situation awareness that occur if an authority is issued at the time the driver requires it. Also, as train traffic increases, the possibilities for establishing queues decrease because of the smaller separations between trains and greater complexity of their interactions; thus, use of this strategy is less likely to be possible when workloads are highest.

15.3.3 Task 3: Fulfil train order (TO) or relinquish authority (ASW)

This task must be performed at the end of one track section, prior to the train driver being issued with an order or authority for the next track section.

Performance Times: TO = 20s; ASW = 10s.

- **Cue(s) for task initiation**
  TO
    Auditory signal from
      usually (a) locomotive radio
      or (b) Telecom phone
  ASW
    Visual + auditory screen alarm
      Within ASW, the acknowledgement/relinquishment of an authority automatically updates the electronic graph (except when radio transmission problems).

- **Information modality and quality**
  TO
    Auditory (radio or phone signals plus speech)
      Stimulus quality of speech is usually good: well above threshold level, clear, attention-getting – but intrusive because auditory.
      If radio transmission is poor, stimulus quality may deteriorate significantly.
  ASW
    Visual + auditory (alarm)
      Good quality visual cue plus auditory alarm (1 beep per 5 seconds) – both clear, well above threshold level

- **Required performance components**
  TO
    Decide priority of cue and when to respond (see corresponding section of Task 1 analysis, above)
      If do:
    Answer radio/phone
    Decide priority of message and whether to deal with immediately
      If not: need to make call later – memory load.
      If do:
    Converse with driver
    Record information from driver on TO form – update of situation awareness
  ASW
    decide priority of alarm and when to respond
      If not, increase in time pressure because alarms remain active until response made
      If do:
    Relinquish authority:
      select alarm (brings up authority screen)
      select release section
    Observe that the graph has updated – update of situation awareness

- **Response modality**
  TO
    Oral, manual (paper and pencil) – TO form + graph
  ASW
    Manual (mouse/keyboard) – plus oral (radio, speech) if problems with signal return.

- **Comments**
  With both systems, performance should be at the routine, rule-based level provided that radio transmission is operating effectively. It is usually initiated by the driver, with ASW communications being solely via the work-station unless there appear to be significant radio transmission problems. In deciding

whether or not to defer issue of a Fulfilment, an additional strategic consideration for the controller is that deferment until the next issue of order or authority will probably result in some economy of overall performance time. With ASW, there is additional workload if (a) the driver forgets to initiate the task, resulting in no radio signal being sent, or (b) there are radio transmission problems. The information available to the controller as to whether (a) or (b) is the cause of an absent signal was inadequate, increasing workload considerably.

15.3.4 Task 4: Report location – field

This task is performed to enable the controller to maintain awareness of the location of all trains. Signal transmission is automatic for ASW.

Performance Times: TO = 30s; ASW = 10s.

- **Cue(s) for task initiation**
  TO
  Usually Cue (1): Auditory signal from *either* phone *or* locomotive radio
  Sometimes, Cue (2): *if location reports not initiated by field staff*:
  Cued by own situation awareness – based on which there is a perceived need to update information. In this case, call initiated by controller, via above means.
  ASW
  Within ASW, the acknowledgement/relinquishment of an authority automatically updates the electronic graph (assuming that radio signal transmission is effective).

- **Information modality and quality**
  TO
  Auditory (radio or phone signals plus speech)
  Good quality cue signal (clear, well above threshold level, attention-getting) – but intrusive because auditory
  When radio transmission poor, speech signal may be of poor quality.
  ASW
  Visual (screen)
  Good quality cue signal (clear, well above threshold level, but less conspicuous because visual information is less attention-getting. This level of conspicuity appears to be appropriate for the task.
  except when there is a problem with radio transmission, in which case:
  Auditory (radio or phone signals plus speech) must be used in addition to visual, screen-based information
  In this case, quality is the same as normal TO conditions.

- **Required performance components**
  TO
    - If Cue (1): As for Task 1
    - Answer telephone/radio – to eliminate auditory signal
    - Decide priority of cue and whether to speak to the caller immediately or defer for a short time until less busy (strategic decision-making)
    - Converse
    - Decide priority of message and whether to deal with immediately (strategic decision-making)
      - *If do*:
      - Transcribe information received onto graph (paper and pencil) – may use ruler to update graph *usually* at time of call *or* later (strategic decision-making)
    - *If deal with later*:
        - Remember (memory load) to
        - Make call later using telephone/radio
    - If Cue (2):
    - Make call using telephone/radio
    - Transcribe information received onto graph as described above.

  ASW
    - Monitor screen display of train positions – ongoing controller activity of central importance in maintaining situation awareness
    - (Additional activities if radio transmission problems)

- **Response modality**
  TO
    - Oral, manual (paper and pencil)

  ASW
    - None, unless problems with radio transmission; if problems, manual (keyboard/mouse); possibly also oral

- **Comments**
    - TO and ASW are similar in responsibility but with ASW the system displays information cueing the controller to take action when needed, whereas controllers using the TO system are dependent on their own situation awareness and memory. In both cases, updating of the train graph (either paper- or screen-based) should be followed by the controller reviewing implications for future running – knowledge-based strategic decision-making with significant mental demands.

15.3.5 Radio transmission problems and ASW alarms

Radio transmission problems occurred intermittently during the period of data collection. In both TO and ASW systems, poor transmission increased the difficulties experienced by controllers in conversing with drivers or other field staff. The observed effects on task performance ranged from the relatively frequent need to repeat a few words to occasional temporary abandonment of the task. Since TO operation relies primarily on voice communication, difficulties in conversing with field staff had a greater impact on TO than on ASW system functioning.

The problems caused within ASW were of a different nature. The human auditory perceptual system has a greater tolerance for degradation in the quality of input information than does the ASW work-station. Since ASW operation relied primarily on the transmission of radio signals to and from the controller work-station rather than on conversations with field staff – particularly in issuing and relinquishing authorities and receiving location reports – degradation in the quality of radio signal transmission caused significant disruption to ASW operations. Such disruption was observed on several occasions, particularly in the context of uncertainty about whether or not an authority had been returned.

To resolve this uncertainty, controllers did not simply repeat normal task activities but performed additional actions:

Monitor informal auditory cues associated with radio signal transmission – ongoing during periods of uncertainty.

Interrogate the authority 'window' on the VDU display as necessary to determine whether the driver has 'relinquished' the authority (pressed a button on the locomotive control unit). *If authorities NOT returned at appropriate times, continue monitoring for a short period. If still not returned, then:*

Make call to the driver.

Note: contact with the driver would normally be via locomotive radio, but if the problem is with radio transmission then depending on the nature of the radio problem, radioing the driver might either be impossible, or just worsen transmission problems for other users, which presents the controller with an additional problem.

Converse (ask him/her to 'send again').

Overall, the observed effects of radio problems were much greater for ASW than for TO operations. This was because with TO the effect was simply to increase the time taken for a partial or full repeat of normal task responses, whereas with ASW there was both an increase in time taken – generally greater in ASW than in TO relative to the time normally taken for the task – and an increase in ASW mental demands related to uncertainty about whether signal transmission failure had actually occurred and consequent doubt about the best response strategy. This contrasted with the effects of transmission problems in TO, where there was difficulty in hearing the necessary information which necessitated some repetition, but no uncertainty about the nature of the problem nor the appropriate course of action.

With ASW, even moderately busy periods were marked by a significant increase in the number of auditory alarms demanding some action from the controller. In both systems, controllers were free to prioritise actions according to their own preferred control strategies. However, controllers reported that they had less sense of such control in ASW because of the nature of the alarms which were to some extent intrusive and demanding of attention. On the other hand, the existence of the alarms reduced mental demand because it decreased the controller's need to rely on memory, i.e. the alarms tended to *increase* workload because of the increased time pressure but to *decrease* workload because they removed the memory demand related to remembering what actions were required. These effects were significant both when radio transmission failed, and during high workload conditions that were due to high traffic levels.

15.4 Different dimensions of controller workload

The above analyses were of value in identifying the specific sources of task demands in the two different systems. Their most appropriate application is to the design or modification of aspects of the control system hardware and software in order to avoid excessive workload while maintaining or enhancing situation awareness. The observed differences between the two control systems in the nature of task demands suggested that controllers' workload might vary not only in overall level, but in its nature – the balance between levels of the different *dimensions* of workload.

The NASA Task Load Index (TLX) (Hart and Staveland, 1988; Tsang and Wilson, 1997) was used as the starting point for further investigation of differences between the two systems of controllers' overall mental workload. The six standard TLX workload dimensions are: physical demand, mental demand, temporal demand, effort, own performance, frustration. However, the present application was not typical of those for which the TLX was developed, and the above six dimensions were not considered to be optimal for the present purpose. The 'Own Performance' scale was felt to be particularly inappropriate since results from some previous research using the TLX in occupational environments (O'Bryan *et al.*, 1991) found very little variation – probably because when the task of interest is being performed as part of someone's normal job, they are unlikely to believe (or at least, unlikely to *acknowledge*) that they are doing less than a reasonably good job. In the case of train controller performance, this is certainly likely to be the case. To establish how best to modify the standard TLX dimensions for use with train controllers, a semi-structured questionnaire was developed and administered informally to several of them.

From these interviews it emerged that the much less severe consequence of controller error in ASW compared with TO was perceived as a significant factor making ASW 'easier'. This was consistent with past research evidence that 'consequences of error' can be a determinant of perceived workload (Herbert, 1974), and is also consistent with a recent finding that 'working carefully to avoid errors' was a key element in the workload of people performing repetitive tasks in manufacturing industry (Macdonald, 2000). Further, the interviews highlighted the role of frustration as an important workload dimension, particularly in relation to the inadequacies of radio transmission within ASW. Frustration was seen as qualitatively different from and more important than 'stress', although 'stress' is subsumed within the definition of 'frustration' on the current standard TLX scale, which was retained. A key factor underlying frustration in this context appeared to be the perceived *avoidability* of the problem causing it – in this case, it was perceived that management action could remedy the radio transmission problem if managers chose to do so.

Based on these findings, together with results of task analyses, the 'Own Performance' scale was omitted, being replaced by importance of avoiding errors; 'mental demand' was subdivided into 'mental demand' and 'perceptual demand'. The resultant set of workload dimensions was defined in terms appropriate for the specific characteristics of train controller workload, as follows:

- **Perceptual demands** of the work
  task information is *able to be noticed* easily when needed; *able to be seen or heard* easily: information written on the train graph or elsewhere; details of screen displays.
- **Mental demands** of the work
  demands of *thinking* and *planning; decision-making; switching attention* between different aspects of the work while *maintaining task priorities; remembering* to do things; *recalling* information when you need it.
- **Importance of avoiding errors**
  awareness of the *consequences of errors.*
- **Physical demands** of the work
  physical discomfort or tiredness relating to *posture* (head/neck, arms, body); *reaching and twisting*; effects of *seating* and *work-station layout*; adequacy of *opportunities to walk around.*
- **Time pressure**
  work rate demands – *time available to deal with things.*
- **How much effort** you have to make
  your level of *concentration*; how much *attention* you need to give.
- **How much frustration** you feel
  includes *stress, annoyance, irritation, etc.*

The above workload dimensions were incorporated within a short questionnaire to obtain controller rankings of their relative importance within the two systems. Mean rankings from 6 of the total 12 controllers who were familiar with both systems are shown in Table 15.1, separately for TO and ASW, each separated into 'when things are very busy' and 'when things are fairly quiet'. Low values represent a high ranking, indicating that the dimension is perceived as one of the more important contributors to overall workload.

Looking at the ranks summed over all four conditions, it can be seen that 'mental demand' was seen by controllers as the most important overall contributor to workload, particularly when things are busy when it was ranked '1' for both TO and ASW systems. Next most important dimensions were 'perceptual demands', followed by 'importance of avoiding errors'.

Within TO, the relatively high ranking of 'time pressure' for the busy condition suggested that, consistent with time-line analyses of workload conducted during an

Table 15.1. Average rankings of the relative importance of workload dimensions as contributors to overall workload for TO and ASW systems, separately for busy and quiet conditions. Values shown are mean ranks of the seven dimensions: 1 = most important, 7 = least important.

	Workload Dimensions – Ranked In Order Of Importance						
System Condition	*Perceptual demands*	*Mental demands*	*Avoiding errors*	*Physical demands*	*Time pressure*	*Effort*	*Frustration*
TO: BUSY	2	1	4	7	2.5	5	6
QUIET	2	3	1	5	6.5	4	7
ASW: BUSY	4	1	2	7	6	5	3
QUIET	1	2	3	5.5	7	4	6
SUM OF ALL CONDITIONS	9	7	10	24.5	22	18	22

earlier phase of this project, high workload levels were close to the maximum acceptable. Related to this, the lower ranking of 'importance of avoiding errors' with busy conditions relative to quiet further suggests that high levels of time pressure might negatively influence performance quality.

Within ASW, the relatively greater importance of 'frustration' during busy periods probably reflects the accumulation of alarms waiting to be dealt with, possibly exacerbated by radio transmission problems (as observed during the course of the project). The experience of 'frustration' appears, from interviews, to be related to the perception that radio problems could be solved, thus lowering mental demands to a more acceptable level, if the organisation chose to spend the necessary money. The perceived organisational failure in this regard apparently resulted in significant frustration when it directly increased workload during busy periods.

Finally, the relatively higher ranking of 'perceptual demands' during quiet conditions may simply be due to the relatively low levels of other dimensions during quiet periods: there is not much to do other than monitor the screen displays.

15.5 Overall differences between TO and ASW in workload dimensions

With ASW, 'time pressure' was one of the least important contributors to workload, which is consistent with controllers not having reached the maximum level of workload with which they could cope. For TO, 'time pressure' was equal second in importance when conditions were busy. Frustration was one of the least important contributors to TO workload, but with ASW, 'frustration' was ranked third in importance when conditions were very busy, probably due to problems with radio transmission.

It seems that there were two key reasons for these differences. First, the high workload levels experienced with TO during busy periods changed the relative importance of some workload dimensions. Second, the incidence of radio transmission problems and their impact on the frequency of auditory alarms that could not immediately be dealt with led to greater 'frustration' with ASW than with TO.

15.6 Quantifying overall workload

As well as ranking the relative importance of the different workload dimensions as reported above, controllers rated the magnitude of overall workload, using a standard psychophysical rating technique – magnitude estimation, anchored by a criterion condition (Alteras-Webb and Dekker, 1994). The criterion condition, against which other conditions were rated, was 'ASW workload level on a very quiet day'. Overall workload for this criterion condition was assigned an arbitrary value of 10 units.

Given this as a criterion or calibration condition, mean ratings of overall workload levels for the other conditions are shown below.

ASW – Quiet: 10 (arbitrary units, specified as a criterion)
– Busy: 61
TO – Quiet: 15
– Busy: 80

Clearly there was a large effect of being busy on perceived levels of overall workload within both TO and ASW systems. This effect was much greater than the perceived difference in workload between ASW and TO systems. The difference between systems was *relatively* greater for quiet conditions (10 versus 15) than for busy conditions (61 versus 80).

The reliability of data from the questionnaire is limited by there being only 6 respondents from among the total of 12 controllers who at that time had experienced both systems. Another limiting factor is the relatively high incidence of problems with radio transmission during the period immediately preceding and during data collection. As discussed above this would be expected to increase workload relatively more with ASW than with TO operations, and may well have inflated the estimated magnitude of overall ASW workload relative to TO workload.

15.7 Future monitoring of ASW workload levels

Overall, it was clear that workload levels with ASW were considerably lower than with TO, and that the traffic levels being handled by the ASW system at the time of the study had not yet challenged operators' maximum performance capacity – at least if radio transmission problems were excluded from consideration. The extent to which ASW workload could be further increased – by increasing the area of track per controller – without exceeding acceptable workload levels was not specifically predictable from the present data. It was therefore recommended that workload levels should be routinely monitored in future to ensure that task demands do not become excessive, and to enable the most effective deployment of controller resources over the areas controlled.

It was recommended that relationships should be determined between:

- Main **task parameters**: numbers of trains and other track users; their separations and types of movements; numbers and types of problems in train operations (e.g. train breakdowns); work-station and control system-related problems (e.g. radio transmission); other task parameters.
- A set of **controller performance measures**: performance frequency of each of the primary controller tasks; possible performance indicators of approaching overload (see below).
- More direct measures of **workload experienced**: controller ratings of workload levels, using either a unitary scale, or the TLX-based dimension scales (which would have greater diagnostic value), at pre-determined intervals throughout specific shifts for which the above task and performance data are known.

Analyses of the above data should then be used to identify the task and performance measures which are the most reliable predictors of workload, controlling for other factors which might influence these relationships such as indicators of probable controller fatigue (e.g. hours worked recently, night shifts) and controller skill (e.g. extent and recency of experience in ASW control system). From this could be derived a workload index comprising a composite of readily available task parameters and performance measures which would be both predictive of workload levels and convenient to monitor routinely. On this basis, future changes in controller workload could be managed in a way that maximised safety and efficiency of system operation while also protecting the well-being of controllers.

15.7.1 Possible indicators of imminent 'overload'

Unlike machine performance, human performance tends to degrade gracefully as it approaches overload. Humans are very adaptable and tend to adapt to changing workload levels by changing their performance strategies in ways that protect their performance of the tasks they perceive to be most important. Accordingly, the most useful performance indicators of imminent overload are likely to be those reflecting such changes in strategy – including changes in normal patterns of sub-task scheduling. Examples of performance indicators which could be investigated for this purpose include: duration of delays in attending to phone/radio calls; frequency of phone/radio calls answered and those where the caller is told that they will be called back; durations of delays in calling back; frequency of asking caller to wait while another task is completed; frequency of recording caller or other information as a temporary note to be further dealt with later; frequency of deferring other actions, or other, less critical tasks; duration of delays in responding to work-station alarms; numbers of additional alarms present at any one time. However, such changes might sometimes be adopted for reasons other than approaching overload, which means that their reliability as workload indices must be established in terms of their empirical relationships with task parameters and with more direct measures of overload, as suggested above.

15.8 Implications for interface design

While the main focus of the present study was on controller workload, it was evident from the analysis of task elements and associated aspects of the control system interface that some components of the interface played a particularly critical role in determining workload.

Notable among these were the ASW auditory alarms, the most commonly occurring of which indicated that a task was waiting to be performed. For example, an auditory alarm indicated when a train driver at the transition point from one 'authority' section of track to the next had sent a 'relinquish authority' signal from the train cabin to the controller work-station. This signal automatically updated the electronic graph depicting train locations and activated an auditory alarm to indicate to the controller that the driver required an authority to be issued for the next section of track. In earlier prototypes of the interface such alarms had been much louder than at the time of the study. They had been modified in response to controllers' objections that such intrusive alarms made them feel controlled by the system rather than in control of it. Nevertheless, these alarms still appeared as a source of additional workload. In particular, their continuing high salience, albeit at a lower intensity than originally, made it more difficult for controllers to follow their preferred strategies, since these often entailed deferment of actions that would silence the alarms. This tended to increase controllers' frustration levels, particularly when multiple alarms were present.

Clearly, the alarms serve to increase system efficiency by cueing controllers to perform necessary tasks; further, they reduce controllers' workload related to system monitoring, planning and related memory load. A challenge for designers is to produce interfaces and alarm systems that effectively cue controllers to perform required actions, preventing them from being overlooked, but that are not so intrusive as to cause frustration when the controller's response to them is necessarily delayed by the performance of other, higher priority tasks.

Acknowledgements

This chapter is presented with the permission of the Public Transport Corporation of Victoria and VicTrack, Australia. Thanks are particularly due to John Herring, Caroline Dowell and Peter Stagg, who were responsible for documenting details of TO and ASW tasks prior to the present study.

References

1 ALTERAS-WEBB, S., and DEKKER, D. K.: 'Measuring perceived task difficulty using magnitude estimation: a demonstration and replication'. Proceedings of the Human Factors & Ergonomics Society 38th Annual Meeting, Santa Monica, CA, 1994, pp. 335–9

2 ENDSLEY, M. R., and RODGERS, M. D.: 'Situation awareness information requirements analysis for en route air traffic control'. Proceedings of the Human Factors & Ergonomics Society 38th Annual Meeting, 1994, pp. 71–5

3 HART, S. G., and STAVELAND, L. E.: 'Development of NASA-TLX (task load index): results of experimental and theoretical research', *in* HANCOCK, P. A., and MESHKATI, N. (Eds.): 'Human mental workload' (North-Holland, New York, 1988)

4 HERBERT, A.: 'Factors in the perception of work difficulty'. Report No. 53, Institute of Applied Psychology, 1974 (Director & Editor: Gunnar Borg, University of Stockholm, Sweden)

5 HUEY, B. M., and WICKENS, C. D. (Eds.): 'Workload transition: implications for individual and team performance' (Committee of Human Factors Commission on behavioural and Social Sciences and Education, National Research Council, National Academy Press, Washington, DC, 1993)

6 LENIOR, T. M. J.: 'Analyses of cognitive processes in train traffic control', *Ergonomics*, 1993, **36**, (11), pp. 1361–8

7 MACDONALD, W. A.: 'Workload, fatigue and job satisfaction in repetitive blue-collar work', *in* DE WAARD, *et al.* (Eds.): 'Human system interaction: education, research and application in the 21st Centruy' (Shaker Publishing BV, Maastricht, 2000)

8 NEERINCX, M. A., and GRIFFIOEN, E.: 'Cognitive task analysis: harmonizing tasks to human capacities', *Ergonomics*, 1996, **39**, (4), pp. 543–61

9 O'BRYAN, S. J., MACDONALD, W. A., and EVANS, O. M.: 'A comparison of some workload analysis techniques'. Proceedings of the 27th Annual Conference of the Ergonomics Society of Australia, Ergonomics Society of Australia, Canberra, 1991, pp. 139–47

10 RASMUSSEN, J.: 'Cognitive control and human error mechanisms', *in* RASMUSSEN, J., DUNCAN, K., and LEPLAT, J., (Eds.): 'New technology and human error' (John Wiley & Sons, New York, 1987) pp. 53–61

11 SARTER, N. B., WOODS, D. D., and BILLINGS, C. E.: 'Automation surprises', *in* SALVENDY, G. (Ed.): 'Handbook of human factors and ergonomics' (John Wiley & Sons, New York, 1997)

12 TSANG, P., and WILSON, G. F.: 'Mental workload', *in* SALVENDY, G. (Ed.): 'Handbook of human factors and ergonomics' (John Wiley & Sons, New York, 1997)

13 WICKENS, C. D.: 'Attention', *in* HANCOCK, P. A., and MESHKATA, N. (Eds.): 'Human mental workload' (North-Holland, New York, 1987)

14 WILSON, J. R., and RAJAN, J. A.: 'Human–machine interfaces for systems control', *in* WILSON, J. R., and CORLETT, E. N. (Eds.): 'Evaluation of human work: a practical ergonomics methodology', (Taylor & Francis, London, 1995, 2nd edn.)

15 WOODS, D. D., O'BRIEN, J. F., and HANES, L. F.: 'Human factors challenges in process control: the case of nuclear power plants', *in* SALVENDY, G. (Ed.): 'Handbook of human factors' (John Wiley & Sons, New York, 1987)

Chapter 16

Power generation: The advanced control desk

Andrew Lichnowski and Chris Dicken

16.1 Introduction

The privatisation of the utility sector has brought enormous changes in the way these markets operate and consumers buy their services. Today, over 10 years since the end of nationalisation and monopoly supply, the electricity industry comprises a large number of private companies competing fiercely to sell electricity directly to industrial, commercial and domestic consumers. The generators that provide the power, the energy and the system ancillary services equally compete on price to produce electricity competitively into the market. This has brought with it new opportunities in the way electricity power plants need to be managed and controlled.

Since privatisation, the delivery of power, energy and other ancillary services has been subject to a number of complex contractual obligations and monitoring procedures and these are changing continually. The pressure of competitive pricing places a greater emphasis on plant availability, generation efficiency, plant life and maintenance costs. Local and global environmental concerns are resulting in increased regulation and tightening emission limits, requiring improved combustion control. All these factors have resulted in a major shift in what is expected from power station operations staff. In particular, the role of the operator has moved from equipment controller to plant manager, balancing competing requirements to achieve the optimal production. To assist the operator's team in achieving better performance there has been a need to introduce new control and information systems that guide the commercial requirements, and improve the technical controllability and flexibility of the power plant.

It has proved an additional challenge to address these issues on the older coal-fired power stations, with their disparate control systems and uncertain commercial lives. The competitive market has increasingly forced the coal-fired units into a more marginal position with a consequential change in their operational behaviour, away from the base load operation for which they were originally designed. This has created new opportunities, which has resulted in a significant number of UK generation units being upgraded with new plant management and control systems. An example is the Advanced Plant Management System (APMS) that has now been completed on 23 UK power plants. Figure 16.1. shows a photograph of one of these new control desks.

Predominantly this chapter will describe the authors' experience of how their company has responded in this arena. Initially, it describes a number of plant management and control functions that were previously performed manually and have been automated. This has involved detailed analysis and re-design of the operator tasks. It has also required the design of many new operator interfaces and automatic control functions. The latter part of the chapter will describe some of the more detailed technical and human factors issues involved in producing a full 'soft desk' operator interface.

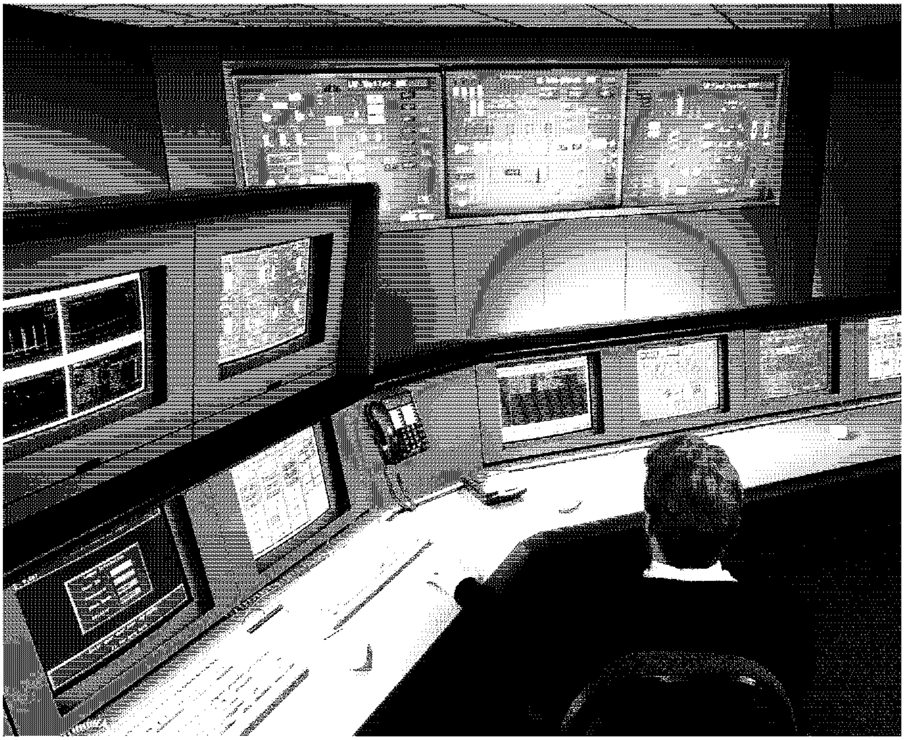


Figure 16.1. Power station control room APMS system.

16.2 The integrated operator interface

For the control room operator, optimisation of power plant control is considerably assisted if the controlling elements are integrated together. One of the challenges faced is how to introduce integration and full 'soft desk' control on older legacy process control systems. The approach requires an upgrade that enables a flexible approach to process control refurbishment without necessarily requiring wholesale replacement of existing systems. It also needs to be adaptable to suit the business case of each power plant.

One solution adopted was the development of the APMS based on a Supervisory Control And Data Acquisition (SCADA) package with an open systems approach permitting integration of third party packages. It also included the integration of new and legacy control systems and implementation of Added Value Applications, all managed by a uniform operator 'soft desk'.

This 'soft desk' was a key component in the refurbishments. Previously the standard desk consisted of a back-panel of indicators and a control desk of hard control and auto-manual stations, with a set of advisory visual display units (VDUs). These VDUs had been added progressively and were often connected to different software systems. The vision for the new, totally 'soft desk' was that it would consist of a VDU screen/mouse driven system with a consistent 'look and feel'. It would integrate all plant information and support all control action and commercial decision-making, giving access to a multitude of information not previously available and thereby facilitating the developing enhanced operator role. As extra automation was installed and control strategies enhanced, the amount of individual plant item operation would decrease. However, the operator would still be crucial to the overall efficient and safe operation of the plant and would need to be able to carry out investigations and take action as necessary.

The electricity production cycle is currently undergoing a dramatic change with the introduction of the New Electricity Trading Arrangements (NETA). As the existing systems are upgraded and altered, to meet the new requirements, the strength of the 'soft desk' approach is confirmed, as it allows the integration of new functionality without wholesale re-design of the interface.

16.3 From business to plant and back

The key interfaces in the business and plant control cycle are shown in Figure 16.2. This diagram focuses on those features important for the power plant control room. These can be broadly segregated into:

- The plant management system.
- Electronic dispatch and logging.
- Integrated load monitoring.
- Energy metering.
- Plant condition monitoring.
- Information technology.

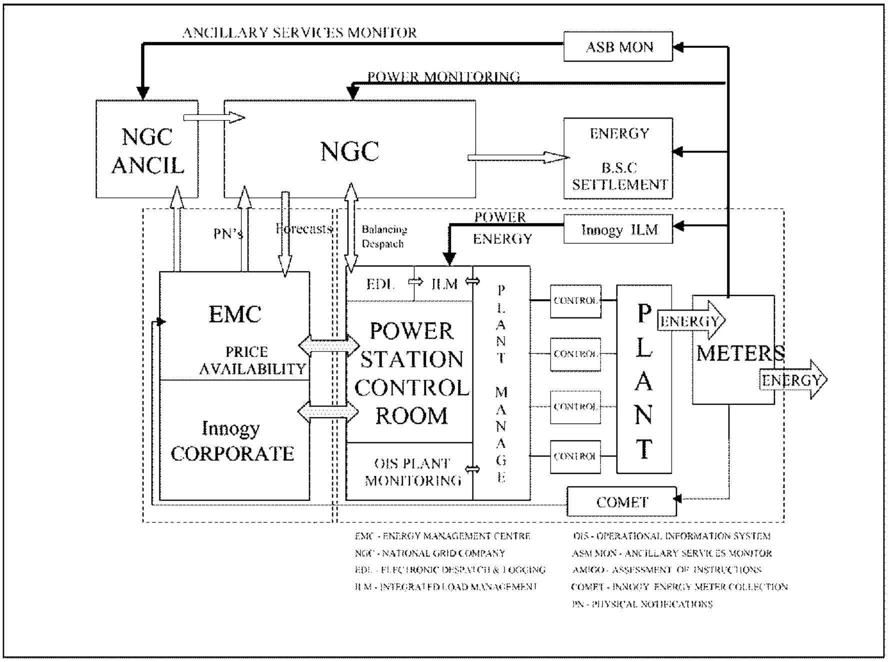


Figure 16.2. Powerstation control room – key interfaces

The control room has to interface with the demands of the customer (the company's energy management centre, EMC) and with the operational requirements of the system operator (the National Grid plc, NGC). To do this the operator needs to manage the plant in a way to deliver the services at the least cost, whilst maintaining compliance with its environmental and regulatory constraints.

The electricity production cycle starts early on the previous day, with each station control room sending plant status data, estimated generation capability, start-up costs and loading rate information to the company's energy management centre system via the Electronic Dispatch and Logging (EDL) system.

Within the NETA framework, the emphasis of scheduling is 'self dispatch' i.e. each generator decides when to run his/her plant, dependent on the existing contracts for supply, rather than as previously submitting data to the system operator (SO) who then scheduled plant throughout the whole supply system. The EMC, taking into account the contracts for supply, calculates the amount of generation needed and runs a scheduling program to determine which units will be required to run to meet this demand. Each individual station's schedule is passed back to the station and the complete schedule is passed to the SO for use in determining what extra generation may need to be requested to run to 'balance' the system. At this stage this data is just indicative. Further processing and if necessary re-scheduling is carried out as time proceeds to the 'schedule' day and indeed on the day itself as time proceeds to the 'gate' (3.5 hours ahead of real time) at which point the schedule must become firm.

The individual power plants prepare for production based on the above schedule information, with the EDL system (described below) providing the main operator interface.

The key systems listed above then provide the operators with their information and control requirements during the whole production process. The loop is closed by the metering systems that monitor the power flows across the grid boundaries and which, as well as providing real-time data for control systems, also provide the inputs into the systems for financial settlement back to the business. Hence, the operator information and control systems closely tie the operational production of electricity to the business requirements of the company.

16.4 The key system interfaces

As can be seen from the above, the whole business cycle process is reliant on a number of well-integrated systems. Some of these are control systems whilst others are very much geared towards the information needs of the operator. In all cases it is imperative to give feedback to the operator concerning their plant state. The sections below describe those systems that require most operator interaction and show how they have transformed the delivery of their information requirements.

16.4.1 Electronic Dispatch and Logging (EDL)

In the control room the EDL system provides the vital direct link between the company's energy management centre, the system operator and the production process. It replaces the previous manual system that relied on telephone and paper fax dialogue. The important plant parameters, needed by the energy management centre for scheduling plant at least cost, are entered in the EDL system and transmitted to the energy management centre system. The EDL system gives direct access to a number of pre-determined plant states, as well as allowing the operator to 'tweak' the parameters to take account of 'on the day' conditions. All current data can easily be viewed and updated if necessary.

After the EMC scheduling process has been completed, the plant generation profile is sent electronically from the energy management centre via the information technology (IT) infrastructure directly to the EDL system. In addition, 'balancing' dispatch instructions are sent directly from the system operator. The EDL system processes all this information to produce unit dispatch instructions that are displayed on the control room desk. The display of instructions also generates an audible signal. 'Balancing' dispatch instruction messages are automatically checked for accuracy and consistency with the previously declared parameters. The operator accepts or rejects the instruction. The unit dispatch information is logged for future comparison and analysis and sent to the Integrated Load Monitoring

(ILM) system for input to the process control system. The integration of these activities ensures an unambiguous record of the dispatch instruction and provides accuracy for the plant control strategy. It also allows the operator to view the predicted generation of their unit and thereby have early warning of load change requirements.

At any one moment the generation unit plant may be in a variety of operating states and offering a mixture of services. The following sections describe some of the key products and techniques that have been introduced to assist the operator in achieving the performance required of the dispatch instructions.

16.4.2 Start-up Management System (SMS)

This involves bringing the plant from a shutdown condition, up to synchronising the electrical generator to the National Grid system. Depending on the temperature of the boiler, the start-up times vary from under an hour to several hours. However, the unit must always synchronise within a +/− five-minute window of the instruction time, otherwise access to the Grid may be delayed with financial consequences. Starting up a power plant requires a sequence of operations to be performed to bring auxiliary plant into service and to warm the boiler along a profile that matches thermal constraints of construction materials. Optimising the process requires a careful balance between minimising the start-up energy costs and limiting plant damage.

Traditionally, the start-up of a coal-fired unit has been performed manually using the built up expertise of well-skilled operators. However, analysis of the start-up methodology reveals significant variations in the techniques employed. To improve the operation, a Start-up Management System (SMS) has been developed using a G2 expert-based system interfaced to the plant management system.

The overall objectives of SMS are to:

- standardise start-up procedures to achieve consistency between units/operators;
- provide a specific start-up plan and schedule for the prevailing plant conditions;
- enable start-up procedures to be optimised to achieve the fastest, most reliable start-ups while controlling plant damage;
- provide procedure-specific alarms, information and advice to reduce the likelihood of uneconomic plant damage and delays.

During the start-up, SMS displays to the operator the relevant part of the start-up plan and indicates the recommended next activity. The rate of progress through the start-up is compared to the plan and indicated to the operator. The system monitors activities to ensure that they are only carried out when the correct conditions are met, and provides warnings if they are not. Monitors can be incorporated into the system to check that key plant parameters stay within prescribed limits during specific periods of the start-up.

Automation of activities is achieved by the use of 'plant-managers', which reproduce the type of monitoring, decision-making and control actions carried out manually by an operator. For example, a firing manager operates the boiler firing based on the optimum combination of mills and oil burners for the prevailing plant conditions while a drains manager operates the boiler drains to control the rates of rise of temperature of the critical boiler headers and also manages the boiler pressure.

The SMS scheduler orchestrates the start-up in light of the instructed synchronisation time and the prevailing plant conditions. It co-ordinates the firing and drains managers, and reschedules the synchronisation time should the plant performance be less than optimal.

In practice, the application of the SMS has considerably reduced the variability of the start-up process, giving the operators greater predictability and consistency in achieving synchronisation times. The operator still has important strategic decisions to make and has to manage unplanned incidents. However, much better support in terms of understanding the options and implications is now provided.

16.4.3 Integrated Load Monitoring (ILM)

Compliance with dispatch instructions through the timely and accurate delivery of energy and power is an essential primary task for the control room operator. Achieving it requires a concise interpretation of the instruction and an understanding of the compliance monitoring criteria. Previous to the NETA rules, the system operator monitored the generation performance of power plant using a set of rules known as Assessment and Monitoring of Instructions by the Grid Operator (AMIGO), the principle of which was based upon a penalty points system. Under these rules, generators were required to achieve the following targets:

- Under steady state generation the mean error and standard deviation of actual achieved generation versus instructed generation was calculated over a 30-minute period. To avoid the award of a penalty point the mean error had to be less than 2 per cent and the standard deviation less than 2.5 per cent.
- During load changes a tolerance of 5 per cent of Generator Rated Capacity (GRC) was allowed. At the end of a load change the unit had to be within the tolerance from the instructed load value to prevent the award of a penalty point. In addition, this also applied at a time 10 minutes after the end of a load change.

A generating unit with more than 12 penalty points in a day was deemed to have had unsatisfactory performance. Four days of unsatisfactory performance in a two-week period resulted in the generator being given a warning. If repeated over the next two sets of 14 days, the generator could be forced into a test situation and possible monetary penalties.

The above rules are difficult to interpret, especially in a real-time situation. To assist the control room operator an ILM system was developed to display the target and to monitor the performance of the plant in accordance with the rules. A highly visible graphical/trend display was developed which has proved invaluable in relating this information to the operator in an easily understandable way. This system is now in the process of being modified to reflect the new trading conditions. The system acquires the current dispatch instruction from the EDL and at any given moment calculates exactly what a generating unit should be producing and profiles the load into the future. It also acquires real-time data from the plant itself (via the metering system), constantly updating the current situation with regard to actual power generation. The new trading arrangements place great emphasis on balancing energy production over half-hour periods. The ILM system can monitor this situation and give advice and guidance to both the operator and the process control system on how the target generation should be met.

16.4.4 Operational Information System (OIS)

The long-term performance of the plant is monitored using a historical database capable of recording data at up to one-second intervals with online access for one year's data and indefinite storage on back-up tape. The OIS is networked on the IT infrastructure and gathers data via a gateway from the plant process control systems. Anyone on the power station can look at plant operation, analyse data, assess efficiency, *etc*. Access can also be granted to other remote specialists looking at specific plant issues, e.g. plant life management, turbine run-up, boiler performance and environmental investigations. Some OIS displays are used as control room operational support tools, for example for the detailed analysis of plant efficiency.

16.5 The transition from hard desk to 'soft desk'

The above examples illustrate the importance of integrating all process control data together and being able to process the raw data to give meaningful information to the operator. Early development was performed on disparate systems and resulted in hybrid interfaces, with the hard desk still being used for basic control action. This traditional power station control interface consisted of a horseshoe-shaped control desk with large, freestanding, back-panels some distance away. These were invariably hard-wired and could only give the operator basic raw data, i.e. none of it was processed to give information and certainly not to display 'knowledge'.

With the rapid changes in technology and the requirements highlighted above, the move away from these traditional 'hard desks' to totally computer-based systems with VDU screens and 'mouse' input (soft desks) became essential. The previously mentioned change in role for the operator, to plant co-ordinator, evolving not only in response to the new business requirements, but also due to the greater amount and more sophisticated automation, represented a major cultural change for operating staff. In making this transition from the well-established traditional hard desk it is important to ensure that the benefits that can be obtained are not negated by implementing a system which is alien to the user, i.e. the operator.

The following sections describe the issues confronted and solutions adopted within the power industry. They are however equally applicable, in a generic sense, to other industries making a transition to 'soft desk' operation.

16.5.1 Philosophy and physical layout

Designing a new operator interface is not a trivial task, whether on a green field new build site or on a refurbishment project, where the 'ghost' of the existing system is always around. Both the issues concerned with the transition and those concerned with the 'soft desk' itself need to be considered. In ensuring that a 'soft desk' would provide the required interface, a number of basic tenets were considered as crucial:

- **Ease of transition.** A standard back-panel/hard control desk has much to commend it in terms of basic information display and ease of operation. The meters, dials and push buttons, in their fixed positions, are readily accessible, give excellent spatial awareness, provide overview and detailed data and allow for fast operation. What is not so easily provided is the wealth of extra data provided by the computing and control systems, the access to that data processed into information and knowledge, the advanced control systems interface and the flexibility to easily change and upgrade. In changing an operator interface to a new desk, providing equivalent means of retaining some of the key points of the hard desk whilst exploiting the advantages of the new system is a desirable objective. This helps to ease the operator learning curve.
- **Maintenance of operator awareness.** Whilst increased automation provides a means for increasing efficiency and enhancing other facets of operations, the operator still remains the critical item in the whole picture. It is necessary, therefore to ensure that operator awareness of the state of the plant is fully maintained.
- **Ease of operation.** A new desk means a move from push buttons for initiating actions, to the use of computer mice. This, along with the need to bring information to view as required, means that the system must be easy and quick to use, with a minimum number of display changes/mouse clicks in order to provide the operator with the access that is required.
- **Screen 'real-estate'.** With all access to information being through the VDUs and the amount of information that can be displayed on a single format being limited, it is important to ensure that enough screens are available to display simultaneously the required data in all plant circumstances.
- **Alarm provision.** Increased automation provides a relatively calm operating scenario when the plant is in a steady state. However, given the fact that the operator has a more restricted view on plant data and the importance of alarms in times of upset, the display of alarm information must be given a high status.

- **Operator involvement**. The move to a 'soft desk' is a major cultural change for operating staff. It cannot work without the full involvement of the operators and without a high regard being given to their input. They have a key role in determining what information and controls should be provided on schematics and in the definition of alarm data.

16.5.1.1 Physical desk design

The physical desk layout and the provision of the control and display devices form the basis on which to build a good operator interface. In moving to a 'soft desk' it is essential to ensure that sufficient display area is available to cope with the amount of data that may need to be viewed by the operator and also that the information can be seen by the appropriate personnel. Unfortunately, there is no easy answer to the question of how many display screens should be provided for any installation. The ideal method is to perform a detailed task analysis of the operator's job, but this is not always practicable. A number of factors need to be taken into account, including the amount of automation installed, the size and complexity of the plant, the intended level of manning of each process, and the required reliability and redundancy of the system.

In the many refurbishments, between 6 and 10 standard (21") VDUs were provided and large (50" diagonal) screens driven by a projection system were also used. The large screens give an ideal display area for overview schematics which can be left permanently on display and which are easily visible simultaneously by a number of people. These could be other operators (during an emergency), the shift controller or even control room visitors. The standard-sized screens are used for detailed displays and control actions. The provision of the large screens has given the operator an equivalent to the hard desk back-panel. Although different in feel, it provides the same characteristics of:

- fixed spatial information;
- an easily understandable overview of the plant;
- generalised alarm information akin to fascia alarm annunciation panels.

Schematic selection and control action are provided by 'point and click' using computer mice. The software allows for a single mouse to be allocated to a number of screens, including the large screens, and to move transparently across the boundaries.

16.5.2 Navigational considerations

Within a 'soft desk' system, the operator has to view the data in a different way to a hard desk, as it is not all permanently on view and there is also much more information available. This requires that the system is extremely flexible to allow viewing of different data on different screens and that the selection of information should not only be quick and easy, but available in a number of ways. A basic philosophy of the system has been that any schematic should be available on any display, giving maximum flexibility to cope with any given set of circumstances and the ability to change the normal situation. Basic navigation through the display schematics is provided by the following.

- **An index schematic**. Each schematic has access to this index schematic by a single button click. The operator can then choose the new schematic to be displayed by reference to labelled buttons. The index schematic is divided into main plant areas and then a logical sequence of schematics below each area. The selection of the correct schematic is then made quick and easy, and is only two mouse clicks away from the starting position.
- **Display hierarchy**. The consideration of how schematics relate to each other has been an important aspect in the total system design. From any given schematic, various options are usually available. These include more detailed schematics (moving down the hierarchy), schematics concerned with closely related plant systems (moving across the hierarchy) and individual plant item schematics (often control panels). These additional schematics are accessed by 'hotspots'. Clicking on a hotspot will bring up the new schematic. Hotspots can be implemented either as labelled 'buttons' on the screen or associated with an equipment icon. The display hierarchy is illustrated in Figure 16.3.
- **System index**. By use of a menu facility, a scrollable alphabetical list of all schematics is available. The required schematic can then be chosen.
- **Short cut keys**. Function keys on a standard QWERTY keyboard can be defined to bring up instantly a pre-selected schematic. This method of selection can be adapted by having a 'soft' set of function keys along the bottom of each display screen, giving the same functionality, but without the need for the hard keyboard.

The techniques described above are not exhaustive, but do provide a basic set that have been proved in practice and should be considered when designing any 'soft desk' interface.

16.5.3 Schematic design

The design of the schematics within the system is closely allied to the navigational aspects of the system. Within a full 'soft desk' implementation it is important to ensure that schematics are developed which:

- cover the whole range of operator actions required;
- are well designed within themselves;
- allow for fast navigation through the multitude of available schematics.

To provide the above, a number of different types of schematics can be considered:

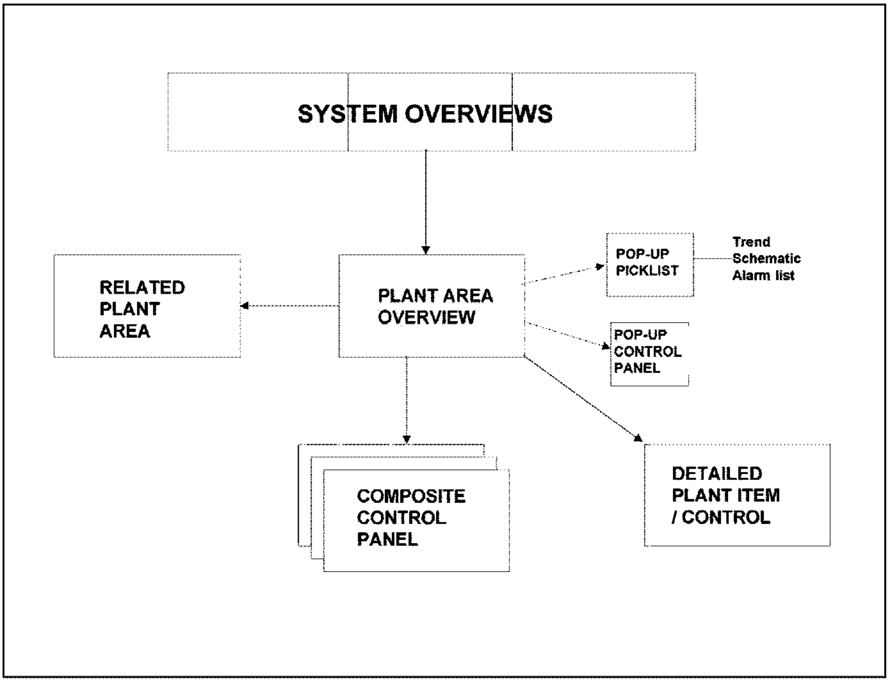


Figure 16.3. Schematic hierarchy.

- **Overview schematics**. The plant overview schematics are used to give an instant picture of the operational status of the complete plant and are mainly concerned with efficiency, availability, safety and environmental compliance. They should use graphics much more than numeric information. They should give alarm overview information and are intended to be used in a 'pattern recognition' mode, being permanently on view. Overview schematics are an essential design feature of 'soft desk's.
- **Plant area/detail schematics**. These schematics take the view down to the next level, allowing the operator to both have an overall picture of the status of the particular plant area plus the ability to investigate, in detail, the state of the plant components. Access to control of individual items can be achieved from these schematics by the means of small pop-up control panels. Many of the plant schematics follow the physical structure of the plant and its associated pipework, giving a better picture of the status of interacting systems.
- **Control schematics**. These are used to bring together all the modulating and sequence controls for a single system (e.g. each coal mill). They display values of key parameters, and provide mouse driven 'push button' access to control the plant items.
- **Trend schematics**. These can either be pre-configured and instantly accessible or bespoke schematics can be built as required by the operator.

Many of the schematics are plant-orientated, often following the grouping of information on the hard desk. However, task orientated schematics have also been developed. These are seen as an important method of reducing the number of displays that are required to be on view and of easing the operator's task. The production of such schematics is dependent on the existence of an integrated full plant database.

16.5.3.1 Schematic style

It is always important on any control system that individual schematics should be well designed. This is particularly so when they provide the only view that the operator has on the plant. Giving the operators a key role in the basic schematic design team is invaluable – they, after all, are the people who know what information they need to see. However, it is imperative to arrange training sessions to ensure that they understand how to approach schematic design in a formal and organised manner, as well as learning basic ergonomic concepts of schematic design. The team also requires the back up of an ergonomist and professional engineers, to provide non-operational help and encouragement.

Many basic ergonomic principles on information display, e.g. when to portray analogue or digital information, to relate size to the importance of the information being displayed and how to avoid cluttering schematics whilst not wasting space have been incorporated into the schematics. One very important facet, that has been addressed as a 'soft desk' issue, has been the need for careful design of plant object icons. These icons need to show a wealth of information. For instance, what plant item type is being represented, what is the current position of the item (e.g. open/closed), what is its operational status (e.g. auto/manual), is the item in alarm, has it been isolated and are there any operational messages/notes associated with this item? It can be seen therefore that the dynamics of such icons can become quite complex.

16.5.3.2 Detailed information and control actions

One of the advantages of a 'soft desk' is the potential for allowing access to much more information. Plant area and plant detail schematics include hotspots configured for individual plant items. These hotspots then all give access, via the right mouse button, to a pick-list of information or control actions associated with that plant item. The options can include:

- Viewing an associated schematic.
- Viewing a trend of the point's history.
- An alarm list with a pre-defined filter (e.g. plant area) applied.
- A message/note box, which may contain information on the plant item's behaviour from the previous shift.
- Viewing a basic point information window – giving detail such as the state of the item, alarm limits, current values, database point, identity, *etc.*

Using the left mouse button will bring up a mini-window, which is typically a control panel for that plant item (e.g. valve, pump, *etc.*) and includes detailed information (e.g. open/close). This will allow the operator to perform control actions on the item. For major plant items this control panel can be a sequence control panel and will also give information as the sequence proceeds.

16.5.4 Alarm integration

Although the overview screens provide a replacement for the hard desk back-panels they do not fully replicate the functionality. The operator is therefore more reliant upon the alarm system, which is an integral part of the system. Alarm indication is provided both by annunciation on overview, plant schematics and a permanently visible alarm list display.

The alarm indications on the overview schematics are used in a similar manner to alarm fascias, with the spatial recognition being important here. Alarm indication at this higher level is provided by means of grouping alarms within the system in plant areas. More detailed investigation can then be undertaken with a plant overview or more detailed schematic, where individual alarm indication is displayed.

The alarm display list is a chronologically ordered view onto the alarms as they occur. Alarms can be given one of three different priorities to help the operator distinguish between urgency of alarms. Other powerful features have been included in the alarm manager such as alarm shelving, plant category filtering, operating mode filtering and chattering alarm suppression. In addition, the mouse can be used to select an alarm and give facilities to bring up an associated schematic, trend, detailed information on the point, textual help information or access to operating procedures.

16.6 Conclusions

This chapter has described a range of control room operational support systems, interfaces to commercial systems and advanced control techniques to assist the operator in optimising the control of power plant. It has shown how these systems serve to enhance the operational environment and how they provide flexible platforms that can be adapted more easily to suit new operating regimes and modified to take account of new trading and regulatory frameworks.

In addition, many issues have been considered that need to be taken into account when moving from a hard control desk to a soft screen-based control desk and the way solutions to these issues have been successfully implemented. Soft desk control systems operation in power plant is now a fact of life and will continue to progress and evolve as the technology and business needs drive it. Moreover, the techniques used here are equally applicable to many other industries.

Chapter 17
Human-centred design for railway applications

Lynne Collis and Felix Schmid

17.1 Introduction

A human-centred approach to the design of work places in the railway industry is essential if railways are to retain their lead over other modes of transport in terms of safety, timeliness of the transport product, level of passenger care and energy efficiency. In this chapter, a number of case studies are presented that illustrate some of the principles of human-centred systems design, as applied to railways. The case studies relate to the handling and management of information in the context of complex systems composed of many critical sub-systems. They are used to describe situations where human-centred approaches could be applied. The case studies presented are based on real situations, but should not be viewed as criticisms of particular approaches.

Railway undertakings and manufacturers of railway equipment pursue a variety of strategies for the presentation and handling of safety critical and other information. Some of these approaches are the result of many years of experience while others were developed to 'fix' particular problems.

17.2 Functional requirements

A high level Integration Definition for Function Modelling (IDEFØ) (Yeomans and Schmid, 1998) type representation of the system railway is shown in Figure 17.1.

This input-output diagram represents the way in which the system responds to a transport demand while respecting the constraints imposed by the timetable, regulations and 'rules of the route' and using the resources of staff, power, track and rolling stock.

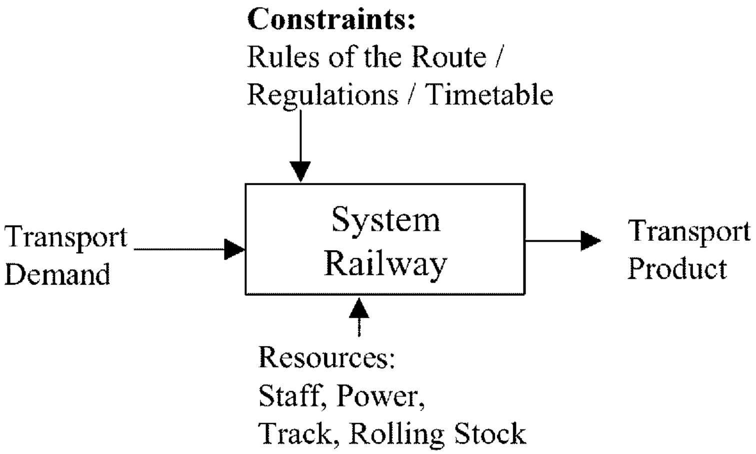


Figure 17.1. IDEFØ diagram of the system railway.

Unlike the road mode of transport, railways require close co-operation between the personnel responsible for the management and maintenance of the infrastructure, staff operating the trains and the people in charge of the train control function (dispatchers and signallers). Similar levels of co-operation must be achieved in the air transport industry but the three-dimensional nature of air space and the lack of a fixed guideway offer pilots more freedom and more opportunities to avoid conflicts.

On the railways, these responsibilities and relationships are defined by safety regulations and standards and by rules of the route. The rules of the route are agreed between the organisation responsible for the railway's fixed elements, such as rails, power and signalling equipment (the infrastructure controller) and the train operating companies. Note that this is the structure currently existing in the UK (Railtrack, 2000). This framework governs the balance between train operations and maintenance by defining conditions under which such measures as temporary speed restrictions and temporary line closures may be imposed.

Once a physical route has been set, with points acting as variable geometry infrastructure elements, a train can only avoid an obstruction or a train travelling in the opposite direction by braking before reaching the point of potential impact. Because of the limited coefficient of adhesion of the steel wheel to steel rail interface, there are substantial braking distances involved (e.g. about 3500m at 250km/h). Track quality must be to a high specification for the same reason since train drivers cannot respond in a timely manner to faults in the running line.

In most occupations, slips and errors resulting from a cognitive failure (Reason, 1990) similar to that of passing a Signal Passed At Danger (SPAD), will result in a reprimand, a warning or a fine. However, throughout the world, laws and regulations governing train drivers' and signalling staff's responsibilities allow the imposition of prison sentences for neglect or dereliction of duty, combined with sanctions such as dismissal and loss of pension rights.

It is thus essential that the tools provided for drivers and other railway staff are of the highest standard possible, within the limitations of the situation. Particular care must be taken in the design of the human–machine interfaces in railway control systems because any slips or errors may lead to inevitable loss of life or financial penalties. Poor traffic management, for example, will lead to lost train paths, reduced capacity and therefore the imposition of penalties.

Stress caused by worry over the risk of fines being imposed on an organisation may lead to errors of judgement or wrongful behaviour by individuals. Since humans are involved in most aspects of train control, any solution to a problem must take into account their capabilities.

It is possible to design a human–machine system which appears to satisfy the requirements of the specification but which leads to totally inappropriate unconscious 'training' of personnel. Such a combination of a regularly occurring situation and an unexpected human reaction is reported to have led to a recent major accident. Other similar occurrences present risks of further accidents. These situations are described in the first case study (see also Schmid and Collis, 1999).

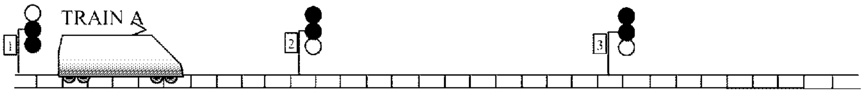
17.3 Case study 1: Training wrong behaviours

Many railway operators, including those in the UK and those using British signalling practice, expect drivers to know in detail the routes over which they travel. They have to learn the physical features of the route and, in particular, the speed restrictions applicable. Usually these relate to poor quality track or to curves where the speed for safe passage is lower than on straight track. On most railway systems world-wide such locations will be protected by technical means. On the British system, there is currently no Automatic Train Protection (ATP) system and the Automatic Warning System (AWS) serves simply to alert the driver to expect a signal at danger. However, there is a system, the Train Protection and Warning System (TPWS) currently being developed (Ford, 1999). This is a lower cost version of ATP.

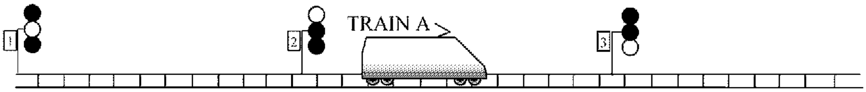
British practice in signalling system design normally ensures that signals are positioned in such a way that drivers can respond in a timely fashion while travelling at the line speed determined by the track alignment and the rolling stock capability. In some cases though, speed restrictions are introduced to overcome the problems caused by poor signal spacing, e.g. where there is insufficient physical distance to give adequate warning to drivers at the higher speed. The physical experience of the driver thus does not match the requirement for the rule.

Such a situation occurred in a particular location on Britain's West Coast route. The signal for which this speed restriction had been introduced would normally be green, allowing the train to proceed. One day, a driver failed to reduce the speed and was therefore unable to stop at the signal which showed a danger aspect he had not expected. Unconscious training had taken place and had resulted in a situation which could only have been prevented by more sophisticated technical means.

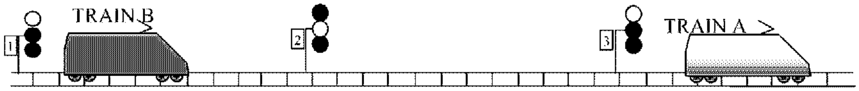
A similar situation frequently arises on the routes approaching the rail termini of major cities, such as London and is shown in Figure 17.2.



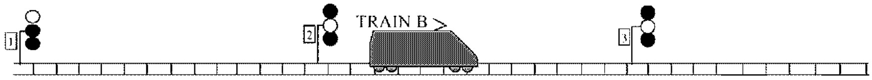
1. Signal 1 turns to danger (red) as Train A passes, to protect it from the rear



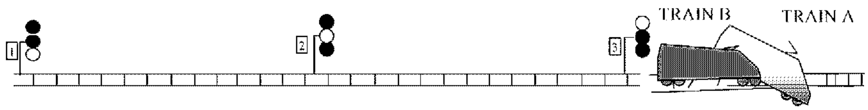
2. Train A advances. Signal 2 turns to red as Train A passes and signal 1 turns to yellow to warn Train B to brake, so it can stop at signal 2.



3. Train A advances. Signal 3 turns to red as Train A passes, signal 2 turns to yellow. Signal 1 turns to red as Train B passes it.



4. Trains A and B advance. Signal 3 turns to yellow as Train A clears the section, normally just before Train B arrives at that signal.



5. Train A has stopped unexpectedly (e.g. due to a break down, passenger action, obstacle on the line). Train B advances but does not brake in time, as signal 3 normally clears to yellow just before Train B approaches. Train B starts braking too late, passes signal 3, which is at danger to protect train A, and collides with Train A.

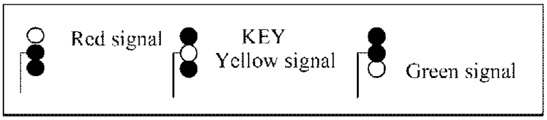


Figure 17.2. Sequence of colour light signals illustrating the potential for unconscious training.

A red aspect is used to provide protection from rear-end collision for either a moving or a stationary train. A single yellow aspect is used to indicate to the driver that the next signal will be red, that is, at danger. Since the routes approaching London stations are used to capacity, drivers habitually see yellow aspect signals which are not followed by red signals, as the preceding train has continued on its route and the signal has become less restrictive. Accidents have occurred when drivers have believed that the yellow aspect would not be followed by red, because at the same time on other days, the preceding train had indeed moved on and the next signal also presented a single yellow aspect.

Infrequently, the preceding train may suffer technical problems or be held up due to an extended station dwell time, due to passenger action or in response to an obstacle on the line. The driver of the following train may then fail to start braking in time and, therefore, may be unable to stop at the red signal protecting the rear of the train. An example of this is illustrated in Figure 17.2, which shows a normal sequence of colour light signals for a busy railway in diagrams 1 to 4 and an accident in diagram 5. Again, unconscious training has occurred, resulting in an accident which could only be prevented by technical means for preventing the train from passing the signal at danger.

17.4 Case study 2: Low frequency high impact events

Humans tend to react well to situations which are part of their normal experience and for which they have been trained. Situations which occur rarely and which have potentially disastrous consequences thus require special solutions, appropriate to the risks encountered.

17.4.1 Background

A major railway company, operating both passenger and freight services over a topographically difficult route, was faced with the urgent need to introduce a new corridor service able to carry road trailers with a corner height of up to 4m, on intermodal ‘piggyback’ trains. The double track railway featured many tunnels and other physical obstacles constraining the kinematic envelope. It was therefore decided to adapt the structure loading gauge of both tracks for the new traffic, as far as possible. On the grounds of cost and technical feasibility, modifications in the most difficult areas were to be restricted to the running line where an enhanced loading gauge could be achieved with relative ease. Since the line was signalled for bi-directional working, this appeared to be a good solution.

However, from the outset it was realised that special precautions would be necessary to prevent oversize intermodal trains travelling on a stretch of track which had not been adapted for them. Such an incident could result in a train being toppled over into the path of a train on the other running line, with potentially disastrous consequences.

17.4.2 Technical considerations

The route is equipped throughout with colour light signalling and relay based interlockings in standard fail-safe technology. The signalling system is of the 'distant and main' type with the colours red, yellow and green. Light combinations allow the display of a limited range of speed information. Automatic route setting to set up train paths was being extended over the whole line, using a train-describer based on the transmission of train numbers between control centres. Train type and service pattern are encoded in the train number.

It was thus easily possible to add a criterion that a train was outside the normal loading gauge. However, the transmission channel used to signal a train number to the next control centre is not safe, i.e. it must be expected that sometimes a wrong number may be keyed into the system or a correct number may be changed in transmission. It could thus not be guaranteed to route a piggyback train at all times onto the running line with the enhanced loading gauge. The risk assessment for accidents on this line was very unfavourable, with the potential for a large number of fatalities, albeit with a relatively small probability. This was due to the mountainous nature of the terrain and the highly intensive train service.

It would have been possible to develop a new train describer system offering inherent error checking and including a method for the fail-safe entry of the train type. However, within the time available this was not feasible. Since the railway authorities overseeing the project were happy to accept a two-channel approach to safety systems design, it was decided to opt for an arrangement where technical means were supplemented by human intervention.

17.4.3 Two-channel safety approach

The railway company had experience of two-channel safety systems where neither channel is guaranteed to fail to ensure safety but where the probability of both channels failing simultaneously was calculated to be very small. The Automatic Train Stop (ATS) system which had been used by the company for many years is of this type. In this system, the first channel represents the human, i.e. the driver, who observes the signal. The second channel is implemented as the inductive transmission of the signal status to the vehicle, using track mounted coils and magnets.

It is dogma that driver and technical systems, perceived as independent, will not fail at the same time, even though this has happened in the past. The company's experts therefore decided to use a similar approach to guarantee safe train operation on the piggyback routes. The initial approach is described in the next section. After further consideration, a modified solution was adopted, which is outlined in section 17.4.7.

17.4.4 Initial approach: piggyback interdiction signal

Route information, available within the relay based interlockings, allowed the fail-safe display of a new signal aspect indicating that points were set in a way which led from a piggyback approved track to a running line with standard loading gauge. A two channel system could therefore be implemented where the automatic train routing system, based on the train describer information, would program the paths for out-of-gauge trains. The driver would be instructed to stop if the route was not suitable for the piggyback train.

A new signal aspect was created for both advanced and main signal locations to indicate that a train's path was about to encounter the restricted loading gauge. Since the special signal aspects were only to be observed by drivers of piggyback trains, a new colour was chosen for the signal. A technical research programme led to the choice of the colour violet since the only available primary colour, blue, suffered from poor visibility. The new piggyback interdiction signal was therefore installed and connected to the points circuits of the interlockings controlling the critical crossovers. Due to the modular nature of the geographic interlockings, conversion proved to be straightforward and was completed together with the loading gauge adjustments. A typical layout is illustrated in Figure 17.3.

17.4.5 Driver's eye view of the initial approach

Drivers on the railway were expected to drive passenger trains, ordinary freight trains and piggyback trains as part of their normal work, often changing between train types in a single shift. In the course of their normal duties they would therefore encounter the violet signal in three different situations (summarised in Table 17.1):

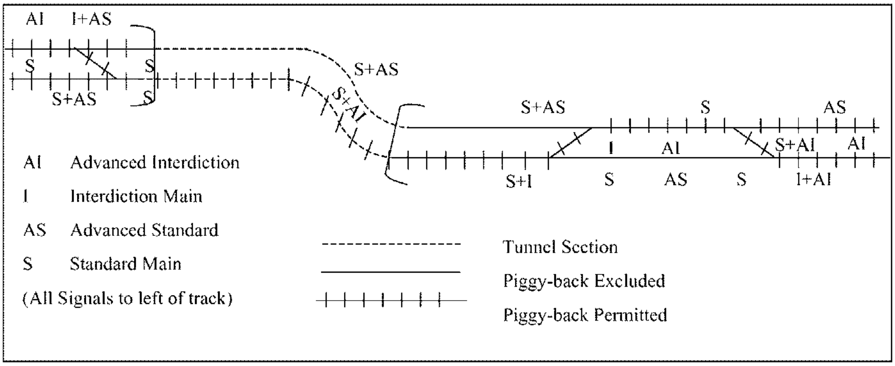


Figure 17.3. Typical layout of piggyback route.

1. When driving a passenger or standard freight train the driver would be presented with the violet signal several times per journey, purely depending on the path set by the automatic routing system. S/he was to ignore the signal since it was of no significance.
2. When driving a piggyback train with out-of-gauge vehicles, the route would normally be set such that the driver would never see a violet signal. Indeed, this was the situation the driver encountered before the introduction of piggyback trains and the installation of violet signals. No action was therefore required, beyond observation of the standard signal aspects and the associated speed restrictions.
3. A restricted route would be set for a piggyback train and, in this situation, only the driver of the out-of-gauge train would be expected to respect the instruction given by the violet aspect at the advanced signal. The driver response would be to stop the train at the main signal. In this situation, the meaning of the violet signal was thus that of the most restrictive aspect, red, even though it would always be paired with a green aspect (for standard gauge trains), normally signifying no restriction.

There are a number of reasons or combinations of reasons, all very rare, for a wrong route to be set:

- data transmission error (technical error) when reporting a train number electronically from one interlocking or control centre to the next;
- data entry error (human error) at the point where the piggyback train enters the train describer area;
- incorrect route setting in automatic control mode due to an undetected system fault (technical error);
- route setting error when system is switched to local (manual) control due to misunderstandings or wrongful interpretation of information about a train (human error).

Both human and technical errors will not occur frequently since the former attract severe sanctions while the latter only avoid detection if a malfunction results in an alternative ‘reasonable’ state of the system, unlikely in a ‘fail-safe’ system.

Table 17.1. Summary of stop signals seen by train driver.

Signal	*Route*	*Train type*	*Response*
Red	Any	Standard	Stop
Yellow	Any	Standard or Piggyback	Prepare to stop at next signal
Green	Any	Standard	Proceed
Violet + Green	Any	Standard	Proceed
Red	Permitted	Piggyback	Stop
Red	Interdicted	Piggyback	Stop
Violet + Green	Interdicted	Piggyback	Stop at main signal

Under normal circumstances, train drivers would thus either not be presented with any instances of the violet aspect or they would be expected to ignore this potentially very restrictive aspect. It was estimated that a driver could be faced with less than three instances of a violet signal with a restrictive meaning per year, while having to ignore a non-restrictive violet aspect up to 17 times per three hour period.

17.4.6 Critical review of the approaches taken

The exception handling procedures inherent in the violet signal approach are clearly technology-centred. A solution was chosen on the basis of technical feasibility and cost rather than with the objective of creating the best result possible within tight constraints.

The driver was allocated ultimate responsibility without being given control over the situation. In fact, there was a clear departure from the principles of the two-channel safety approach as described for the ATS system above. The key difference is that the ATS system consists of two independent channels where the failure of either channel has no effect on the other. In the case of the violet signal, the human channel only comes into effect as a backup for the technical system, i.e. if the automatic route setting malfunctions. Only then is driver action required. The provision of a nominally safer system therefore brought about an environment which, very infrequently, exposes drivers to situations where a mistake has an extreme impact (low frequency/high impact). Quite rightly, the unions and government safety bodies objected strongly to such a stress inducing change in members' duties.

17.4.7 Final system chosen

The development of new methods for exchanging information between track and train, e.g. the European Train Control System (ETCS) balise (Hedberg, 1995), allowed a relatively straightforward technical solution for controlling the operation of out-of-gauge trains. A piggyback train is only allowed to leave the marshalling yard once an exchange of information between a balise and a locomotive's transponder has confirmed the appropriate set-up in the on-board computer system. Balises at critical locations are then able to stop the train before a hazardous situation can arise. The final system chosen is thus a non-failsafe balise supplementing the driver and the technical solution of automatic route setting with violet signals.

Minimal cost was one of the most significant reasons for the original plan to use an interdiction type method of signalling. Adoption of the violet signal as a permission to proceed aspect for trains with outsize loads would have resulted in higher costs because of the need to equip many more locations with additional failsafe signals. It would have allowed, however, the signalling of 'path clear for piggyback', i.e. a message not in contradiction with the green aspect. Drivers with piggyback trains would only have been allowed to proceed with both green and violet aspects.

17.5 Case study 3: Controlling speed on TGV Nord

TVM (Transmission Voie-Machine) 430 is an advanced ATP system which ensures that the driver maintains the speed of the train below the braking envelope (Pincock, 1998). This signalling system provides proceed, speed restriction and stop orders by providing indications in the train driver's cab, rather than by controlling lineside signals. Lineside markers, which are numbered, reflective metal signs on posts, show the point by which the order (such as to stop) must be obeyed. This system also incorporates continuous ATP, which monitors and controls the train's speed if the driver fails to brake in time, so that it stays below a prescribed speed envelope. This envelope is usually 5 to 20 km/h over the speed restriction displayed, within a given range.

This process is illustrated in Figure 17.4, where the braking curve and cab display of a train operating on a TVM 430 signalling system is shown. The diagram has been simplified, reducing the number of speed bands shown. TVM 430 is currently used on high speed railways with a maximum speed of 300 km/hour, notably on French railways (Société Nationale de Chemins de fer Françaises–SNCF). Flashing indications show that, at the next marker, a more restrictive (slower) aspect will be displayed.

The train protection systems of the TVM family have some of the characteristics of an Automatic Train Operation (ATO) system since a train driver could, theoretically, relinquish the control of the train's speed to the on-board computer system, designed to fail to safety. S/he could simply maintain timetabled departures

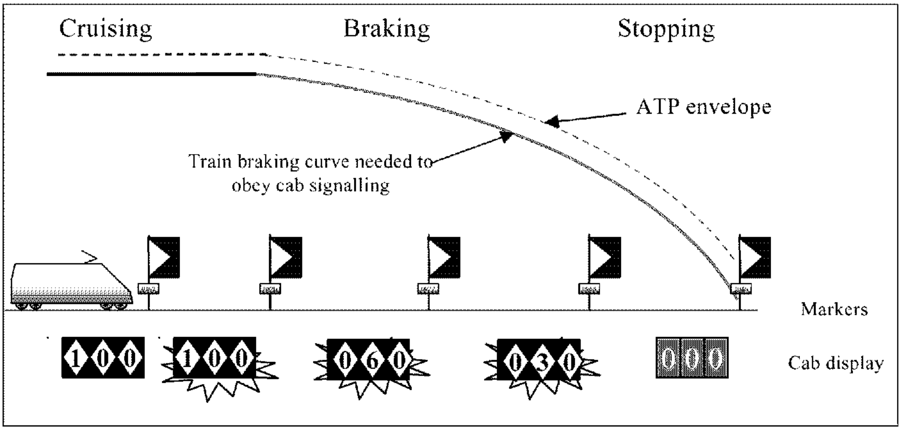


Figure 17.4. TVM 430 braking curve and cab display.

from stations and handle exceptional situations, effectively acting as machine minders with an obligation to observe the track, but no more. However, since TVM 430 is essentially a block-based speed control system, driven by conventional (relay) interlockings, such an approach would result in poor energy efficiency, frequent changes from acceleration to coasting and deceleration and even in 'bunching' of trains.

SNCF chose to make it the driver's responsibility to comply with the timetable while ensuring minimal energy use and best possible comfort for travellers. Even on the TGV Nord line with its many long straight stretches, the driver is expected to use his or her route knowledge. Based on the working timetable s/he adjusts the power supplied to the motors for good timetable and energy performance, while observing the top speed mandated by the TVM 430. The task is suited to the capability of humans since it is a multivariable control task with a limited number of parameters.

A key benefit of this human-centred approach can be found in the relaxed way in which drivers deal with the problem of operating the TGV trains on the conventionally signalled approach to Paris. Being challenged to take meaningful decisions throughout the whole of a journey, drivers cope easily with a changed situation even though this can be very demanding when the final part of the journey is affected by congestion.

17.6 Case study 4: Dispatching and train graphs

The changing nature of the interface between information systems and human skill is clearly demonstrated in the development of the type and scope of information provided to the operator in the signal box, and the impact of this on the skills required.

17.6.1 History

The layout of the manual lever frame box was defined by the space required for the levers and the mechanical interlocking frame. A geographical layout of the track could be provided on a mimic display panel, as illustrated by the photograph of Dalton Junction signal box in Figure 17.5. This shows, above the levers, a diagram of the lines being controlled and which covers an extremely small area in comparison with that controlled by later electro-mechanical and relay-based signal boxes.

Initially, such displays provided only static information about the track layout, as a reference for the signallers. Without track circuits or dynamic information displays, they had to use the timetable, visual checks and a manually completed train register for recording the position and effecting the regulation of trains. The fallibility of such information contributed to several accidents, the worst being the

Figure 17.5. Dalton Junction Box, c.1902 (Photograph John Hinson, reproduced with permission).

Quintinshill crash of 1915, where a train was forgotten during a shift hand-over, and 227 people died (Schneider and Mase, 1968).

Whilst interlocking of signalling has been developed to prevent the routing of trains onto occupied lines, a clear overview is essential for signallers authorising trains to pass signals at danger after technical failures, and for effective incident management.

17.6.2 Static information and timetables

Early train timetables were simply lists of departure and arrival times. However, these soon proved inadequate when managing complex situations. Many railway administrations therefore started to use train graphs (time-distance diagrams) to support the work of signallers and train staff.

In the era of paper timetables this was a very commonplace approach since it allowed the dispatcher to update the graph to show the actual as opposed to theoretical running of trains. This then allowed the detection of conflicts, in terms of connections and track occupation.

17.6.3 Computer control with paper supervision

The next stage of automation, the introduction of computer-based train describers, led to a situation where all the information was presented in numerical form. This increased the demands on the intellectual performance of the operator because of the need to visualise and then analyse complex patterns (a two-step process). As a result, stress levels increased while the inappropriate use of the cognitive skills of the human being reduced motivation levels.

Automatic route-setting has further differentiated the functions of regulation and route-setting through the use of computerised timetable functions, where timetable software interfaces directly with route setting computers. In the signal box of a large French terminal station, for example, the operators call up an alphanumeric display of the timetabled route, and either validate it themselves or modify it in agreement with the supervisor. The signaller checks the setting of the route and train position on the mimic display panel, while the supervisor uses a paper time-distance graph. This is printed with the theoretical timetable, which s/he annotates with lines showing the deviation from that timetable.

The graph method of displaying the progress and track occupation of trains makes good use of the supervisor's spatial reasoning skills and permits identification of the conflicts to be avoided. However, both operator and supervisor are spending more time than necessary on repetitive manual tasks in terms of validation keystrokes and line drawing, rather than on safety and regulation functions. An example of such a graph is given in Figure 17.6.

Recognising the causes for reduced performances, system designers started to introduce train graphs in computer-supported systems (Hille, 1997). The graphs are now presented as dynamic images on computer screens where trains running on time are shown in a neutral or type specific colour while trains running outside their normal paths are shown in a contrasting colour. A representation, as shown in Figure 17.7, affords the operator completely new avenues for analysing a situation and for finding a near optimal solution. This is a very good example of true human–machine co-operation.

Research being undertaken into interactive time-distance graphs may result in enhanced tools to further support the operators' management of the railway system.

17.7 Case study 5: The impact of the systems model on operational communications

Human-centred engineering must not only address the operator–system interface but also the basic relationships between information and system, so that an accurate

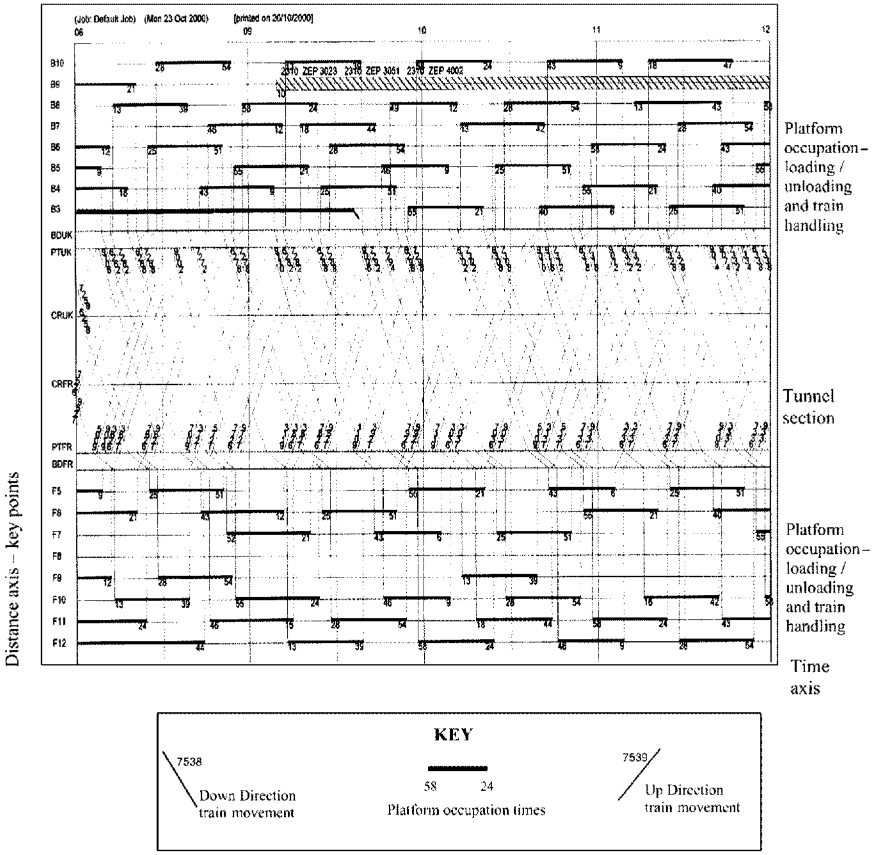


Figure 17.6. Train graph for a shuttle service.

system model is presented to the operator. This is particularly important for incident and emergency management.

Since it is recognised that the number of alarms displayed has an impact on incident handling, there exists a delicate balance between the display of too many alarms and over-simplification of the model. See, for example, the report into the Channel Tunnel fire of November 18, 1996 (CTSA, 1997).

17.7.1 Definition of system databases

A Supervisory Control and Data Acquisition (SCADA) system had been designed for a cross-border railway to control all electrical, mechanical, control and

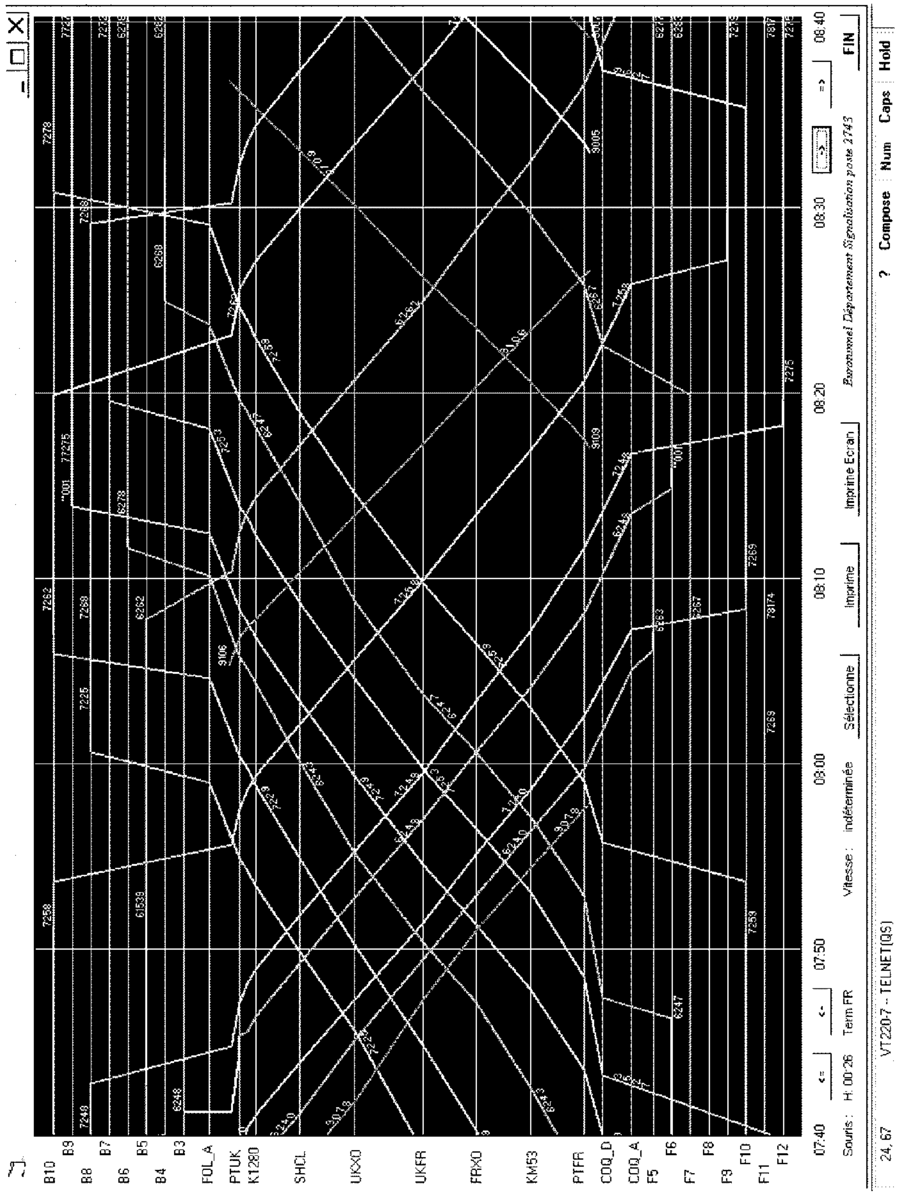


Figure 17.7. *On-screen time-distance graph (courtesy of Eurotunnel).*

communications equipment. This SCADA system was to be operated from one control centre, with generalist operators managing equipment either side of the border. A database was used to translate status information signals received from sensors and actuators into meaningful indications of process and to define the text-strings which represented these changes to the operator (Collis, 1998).

17.7.2 Implementation of systems integration

In order to define the parameters for the SCADA system, information on the inputs available from equipment was collected initially from the individual suppliers. Electrical switchgear, railway traction supplies, tunnel pumping and ventilation equipment, and control and communications systems had all been provided by different suppliers.

The system was complicated by its location on a national border, which required operators and technicians, or groups of technicians from different disciplines, to work together, despite language differences. This diversity in language and culture presented problems in designing an integrated system. Where design engineers sought to work together to resolve differences, the natural relationship was between the engineer working on the hardware for a certain system and the control systems engineer defining that area for the SCADA database, e.g. pumping. Both would often be of the same nationality. However, the operating and maintenance philosophies were different between disciplines and between countries. Each country formally defined, for example, the isolation process for railway traction supplies, but the level of intervention differed between groups of staff.

The operators using the SCADA system and those controlling the railway signalling system clearly needed to be able to communicate effectively with each other on common terms, particularly concerning overhead traction supplies and tunnel equipment. The complex relationships between the systems and between the operators and maintenance technicians made an overall approach essential, to permit all concerned to reach a complete understanding when communicating about any part of the process.

All the same, consistency of the database was initially only partially attained. Where consistency between events on the ground and those defined on the display screen was achieved, the bilingual operator-system interface did not necessarily include an accurate transposition of operating philosophies or a coherent translation into the second language. Examples of this became apparent when alarms and equipment status from a pumping station were displayed together:

- 'not pressurised' had been translated as 'not operational';
- 'no electrical faults' had been translated as 'fully operational';
- the generic status 'available' (i.e. no major alarms preventing system use) had been translated as 'operational'.

It was thus possible for the pumping station to be shown as not operational, fully operational and operational at the same time. Confusion could therefore have resulted for the operator and maintenance technician trying to communicate on the nature of the fault and the consequences for the system.

17.7.3 Solution: Objectives

Once the problem had been identified, several objectives were defined for a modification programme. These included:

- consistency between command and alarm handling;
- consistent terms for severity of alarms;
- consistency between the two languages used;
- accurate representation on the screen of what was happening on the ground;
- clear and consistent presentation of groups of indications which could be displayed together;
- accurate reflection of operating and maintenance processes, to enable operators and maintenance technicians to reach a complete understanding, particularly on matters relating to safety.

17.7.4 Solution: Methodology

Resolution of the integration problems required teams to work together in groups which reflected the relationships required within the database. A detailed study of the database was made, identifying all inputs and events, the equipment to which they related, and the associated text-string in each language. An extensive series of meetings was then organised, between system design engineers, technicians and operating staff, as shown in Table 17.2. These meetings established for each of the indications a text-string in each language that would achieve the set objectives. Computer algorithms and automatic sequences which governed the displayed status for each group of equipment were taken into account, to ensure that the chosen text-strings would be meaningful for the range of parameters.

Table 17.2. Staff relationships in the system.

Team combinations	*System elements*
Operators, system design engineers and maintenance technicians	Events on the ground displayed accurately on the screen
Staff (as above) of each nationality, and bilingual staff	Bilingual operator-system interface permitting staff to communicate effectively
Design engineers from different disciplines	Systems integration to ensure compatible processes where disciplines intersect
Engineers, operators and technicians	A system comprehensible to all groups requiring access

Proposed changes were subject to rigourous evaluation and review, to ensure that they would not introduce false indications into the system. Errors at this level could have a profound effect on the safety of operations. Once the changes had been incorporated into the SCADA system, a programme of detailed factory acceptance testing and commissioning was carried out before operations were commenced.

17.8 Conclusions

Extensive and detailed customer requirements capture is necessary at an early stage of any control systems project. This ensures that the system provides an accurate representation of the processes in the context in which it will be used by operators, technicians and all other parties involved in the project.

There is no right or wrong solution to the human-centred design of complex systems. However, there are a number of approaches which can support the design process. These may result in systems that are not only more user friendly but also more robust and effective than systems which have been designed solely on the basis of hardware and software considerations.

Care must be taken to present operators with information and indications which reinforce both training and good practice, avoiding the potential contradictions illustrated by Case Studies 1 and 2. By taking into account the skills of the operator, and the balance between peak and routine workloads, the designer may build on the key strengths of the operator and improve the functionality of the system as a whole. This is clearly demonstrated by case studies 3 and 4. Case study 5 demonstrates that, in any development, it is essential that a strong working team be created whose members have both the responsibility and authority for the complete process.

Co-operation between users, maintainers and designers of a system is of great benefit if the requirements for clear operational communication are to be achieved and if a control system is to be designed which can be operated without potentially unsafe regions of ambiguity or misunderstanding. The authors have attempted to show that communication between the people involved in designing, building, operating and maintaining a system is very important. The links between subsystems must be clearly defined and managed if successful co-operative working is to be achieved. A human-centred approach makes these factors central to the design philosophy adopted. It ensures that the system is appropriate for the human context in which it is to be used.

Acknowledgements

The authors would like to thank Eurotunnel and Dr T. Hancke for starting them on the human factors trail. They are grateful to John Hinson for permission to reproduce the photograph and Eurotunnel for permission to use the time-distance graphs.

References

1 COLLIS, L. M.: 'Making sense of remote condition monitoring for operations and incident management'. Proceedings of the IEE Seminar on *Remote condition monitoring for railways*, IEE Publications, Birmingham, November, 1998

2 CTSA: 'Channel Tunnel Safety Authority inquiry report into the Channel Tunnel fire of 18th November 1996' (The Stationery Office, London, 1997)

3 FORD, R.: 'Railtrack's train protection and warning system', *Modern Railways*, May 1999, **56**, (608), p. 337

4 HEDBERG, J.: 'Railways choose Eurobalise technology', *International Railway Journal*, September, 1995, pp. 27–8

5 HILLE, P.: 'Konzepte und strategien für KOBS', *Signal und Draht*, 1997, **89**, (9), pp. 19–22

6 PINCOCK, A. J. L.: 'Safe passage to Europe – Eurostar's in-cab signalling systems', *Journal of Light Rail and Rapid Transit*. Proceedings of the Institution of Mechanical Engineers, **212** Part F, London, 1998

7 RAILTRACK: 'Railtrack rules of the plan and route 2000–2001' (Railtrack plc., London, 2000)

8 REASON, J.: 'Human error' (Cambridge University Press, Cambridge, 1990)

9 SCHMID, F., and COLLIS, L. M.: 'Human-centred design principles'. Proceedings of IEE Conference, People in Control, Bath, England, June 21–23, 1999

10 SCHNEIDER, A., and MASE, A.: 'Railway accidents of Great Britain and Europe', translated by DELLOW, E. L. (David and Charles, Newton Abbott, UK)

11 YEOMANS, P., and SCHMID, F.: 'An introduction to IDEFØ' (IDEFine Ltd., Sheffield, and ARRC, University of Sheffield, UK, 1998)

Chapter 18

Integrated platform management system design for future naval warships

Iain MacLeod and Derek Smeall

18.1 Introduction

The Type 23 Frigate as shown in Figure 18.1 provides the mainstay of the primary anti-submarine warfare force of the British Royal Navy (RN). Measuring 133m in length and displacing 3500 tonnes, the Type 23 Frigate was designed in the 1980s, during a time when a potent threat from Eastern Bloc submarines still existed. However, times have changed and recent history has witnessed the role of the Type 23 Frigate moving away from hunting enemy submarines to more constabulary duties, such as enforcing economic blockades and assisting in disaster relief.

As the last Type 23 Frigates enter service, attention has already turned to their replacements. In the new world order, attitudes and strategic aims have changed dramatically. The next generation of Royal Navy frigates will be required to provide enhanced capabilities for land attack, integrated intelligence and interoperability between allies but with minimum risk to life.

In progressing from the Type 23 Frigate to its successor, the Future Surface Combatant (FSC), the most radical design changes will be in the vessel's propulsion and control systems. The drive to gain higher fuel efficiency has led engineers to consider moving from using a lot of traditional equipment to utilising new technology such as Integrated Full Electric Propulsion (IFEP). This produces the problem of how to integrate and govern this new technology while simultaneously driving down both development and through-life costs. A major step towards achieving this aim is to control the whole ship with an Integrated Platform Management System (IPMS). However, on a warship this single network will require supreme survivability through extreme system flexibility. The issue of

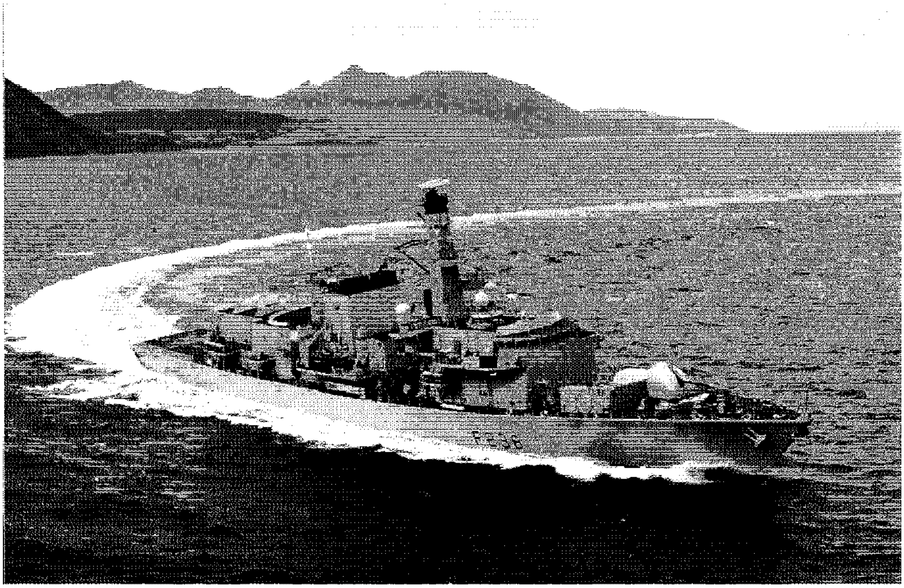


Figure 18.1. HMS Montrose – Type 23 frigate (Reproduced with permission from MoD/ BAE SYSTEMS Marine).

survivability goes beyond the use of back-up cable routes and in the authors' views requires a total rethink of human-control interface philosophy.

18.2 Demands of the twenty-first century

The latter part of the twentieth century witnessed dramatic changes in world order and subsequently in the role of the RN. The RN's traditional role as the protector of the Atlantic lifeline involving large convoy protection and battle groups has all but gone. Since the collapse of the Eastern Bloc, the capability to operate in open sea, i.e. 'blue water' operation, has been superseded. Future deployments will most commonly take the form of enforcing 'no go' zones in territorial waters, selective strikes against military infrastructure and support of land forces in coastal areas. With the decision of the UK Ministry of Defence (MoD) to use larger and potentially more capable aircraft carriers, a carrier group protection role may also be required.

Due to advances in communication technology, commanders will be under severe pressure to act impeccably under comprehensive media scrutiny. Therefore, in a world where one misguided weapon can launch a barrage of international condemnation, the RN will be expected to deter aggression, maintain area control and discharge munitions with surgical precision. Following this philosophy, design emphasis will move towards the introduction of enhanced defensive and area superiority mechanisms. It will be considered imperative that a vessel has the ability to enter a combat zone, enforce superiority by defensive means and utilise its offensive capabilities only as a last resort.

So, to summarise, the FSC will aspire to:

- exert superiority within its theatre of operation to deter enemy attack;
- co-ordinate and support joint force operations by utilising advanced integrated command, control and communication systems;
- react against aggression rapidly, prudently and with supreme efficiency.

18.3 The vessel

There are two primary contending designs for this new style of naval vessel. One is the next generation of low signature, fast, manoeuvrable monohull – illustrated in Figure 18.2. The other is the more radical three-hulled trimaran vessel. One common aspect to the design of both these vessels is the utilisation of an IFEP arrangement. Other platform criteria are left open to innovation and it is these areas we plan to discuss in this chapter. Recent feedback has indicated that naval vessels spend less than 20 per cent of their active lives in a combat state, the remainder of time being spent on normal cruising or in harbour. Thus, it is likely that the frigate of the future will operate, for the majority of its life, little differently to many other commercial marine vessels.

A significant benefit of the advances in processing capability is the potential reduction of manpower required to operate modern naval vessels efficiently. A crew target of 100 has been mooted for the FSC which is almost half the number presently required to crew a Type 23 Frigate. The manpower cost for a naval seaman in service, taking into account salary, training, accommodation, food, health care, *etc.*, runs to tens, and in some cases hundreds of thousands of pounds per year. Therefore, although new technology is expensive, if its installation results in a reduction in the crew requirement of even a small number of ratings, it could

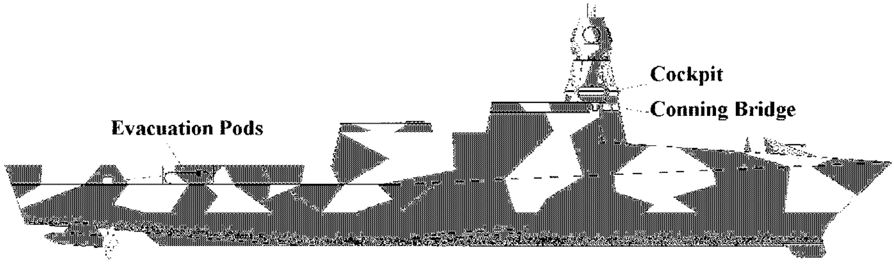


Figure 18.2. *The Future Surface Combatant, monohull option* (BAE SYSTEMS Marine Ltd.).

literally be saving millions of pounds per year. The downside of this scenario is that the remaining crewmembers on the FSC will have to be even more expert in a variety of disciplines. In addition to being multi-disciplinary, personnel will have to work in hazardous conditions and be able to cover the duties of fellow crewmembers if necessary. Thus, twin targets have arisen:

(a) Equipment commonality – not only between different classes of naval vessel but throughout the services.
(b) Human factors commonality – standardisation of control system symbolism and methodology.

The critical point in achieving these targets is the human–machine interface. Maximising human–machine interface efficiency to minimise human workload and hence possibility of error, will improve the efficiency of the combat machine, augmenting the vessels tactical performance and survivability. The key to maintaining effective command of operations is the integration of internal and external real time communication. The use of real time data links and multimedia display technologies can ensure that the sum of all assets is greater than all individual contributions. The IPMS will encompass all C^4I (command, control, communications, computers and intelligence) on board. Leaping forward to concepts for the next generation of vessels brings with it an element of risk. The rapidly expanding field of ship system modelling is helping to minimise design risk factors by allowing every aspect of new vessel design to be tested to a high degree long before construction commences – from hull shape to compartment ergonomics, and from electrical demand trends to weapon systems integration.

18.4 New philosophy

The pace of technological advances within many areas of civilian commercial equipment development has rapidly increased and left military technology standing on the hard shoulder. There is, therefore, an urgent requirement for the new generation of vessel to harness the abundance of available technology much more rapidly. However, budgetary restraints are another major driving force in the procurement of military hardware. The MoD Procurement Executive emphasises its own targets of reducing through-life ownership costs, improving project risk management, and the use of commercial off the shelf (COTS) equipment. There are potentially many areas where – utilising human factors engineering – this philosophy can be brought to bear in the design of the FSC. As stated previously, a naval vessel's life support system is the IPMS and this facility can benefit widely from technological advances. The top level of this control infrastructure is the bridge where a proficient human–machine interface is cardinal.

18.4.1 Ship's bridge philosophy

The totally enclosed bridge design on recent naval vessels has been criticised for its restricted field of view. Figure 18.3 shows a modified two-level bridge design that will provide all round vision for enhanced ship control and surveillance. Introduction of modern optical and display technology will strengthen crew capability to this end. In addition, the ability to reconfigure platform management options for the distinct requirements of tactical and peacetime operations is essential.

It is proposed that all navigation sensors, communications and ship control equipment are integrated. Thus, a screen may have a cruising function (e.g. machinery surveillance) and a combat function (e.g. weapons surveillance). An array of multi-function flat screens will provide the bridge crew with a fully re-configurable set of data displays. This will facilitate a reduction in essential bridge crew to two: an officer of the watch (OOW) and an observer.

18.4.2 Integrated platform management system

This will be the heart of the ship's integration. Unlike present naval control systems which tend to operate separate combat management, ship data management, machinery control, *etc.*, the FSC IPMS will use a central backbone architecture connected to all management systems via a common real time capable Local Area

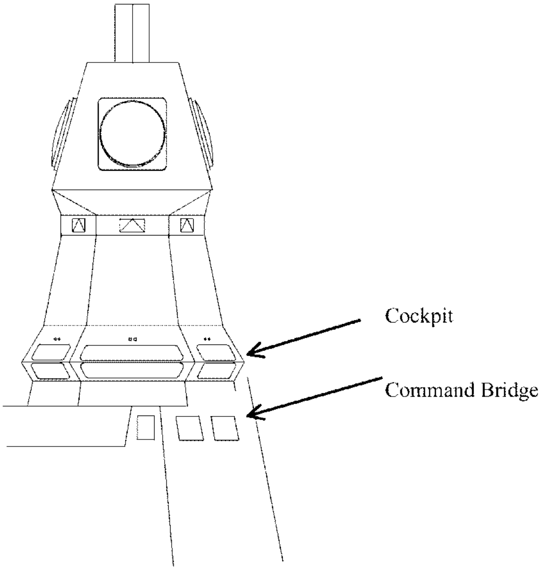


Figure 18.3. 360° bridge integrated within the mast.

Network (LAN). The survivability of this LAN is paramount and will be maximised by the use of fully distributed processing and redundant links. Further, the use of power distribution cabling for back-up data transmission may even be utilised.

18.4.3 Shockproof cabinets

The pace of technological change, especially in computing hardware far outpaces the life of naval vessels. Steps will therefore have to taken to ensure that the new vessel does not enter service with obsolete computing hardware. An additional obstacle to naval designers is that all ship's equipment has to survive high shock levels. Hence, sensitive equipment, such as computers, have to be 'ruggedised' which adds both to their unit price cost and time from design to operation. One method to counter this slow development life cycle is to utilise shockproof cabinets. Therefore, as the cabinets will absorb the shock, the computers, and other hardware installed, can be of COTS standard.

18.5 Existing technology

Figure 18.4 shows a section of the ship's control centre on a Type 23 Frigate that is the heart of the platform management system. This vessel was designed in the early 1980s and incorporated the latest technology available at that time. Although the expansive array of instrumentation may look complex at first glance, the consoles are logically divided into the traditional platform management sub-sections of Propulsion, Electrical, Auxiliaries and Damage Control systems.

On the console faces, the instrumentation is linked by stencilled lines that schematically illustrate the system being monitored. This simple facet provides the user with a prompt positional reference, which can be vital, when a rapid reaction is necessary such as in a damage control scenario. The most common complaint concerning this control centre is the lack of spacing between systems. The general spatial restraints on a densely packed warship such as the Type 23, have meant that the control of four diesel generators, four propulsion engines, numerous auxiliary air, water and lub oil systems and a complex damage control system, have all been compressed into a compartment with a floor area of just 5m by 6.6m.

The limited use of just two small visual display units, which communicate almost totally in text, is indicative of the naval engineers' reluctance to move away from traditional 'gauges, buzzers and bells'.

The forward section of the Type 23's enclosed bridge including the commanding officers' position is shown in Figure 18.5. Although this is a highly regarded naval bridge design, the detached consoles and dedicated displays, indicate its age. It should be noted that the commanding officer has no control functions at his/her position as s/he simply receives information and passes on orders verbally.

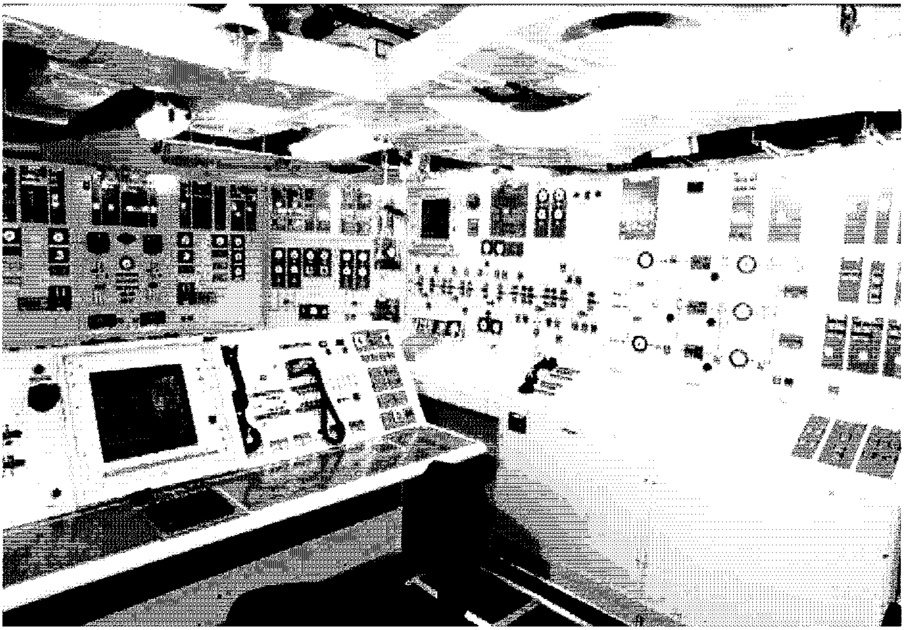


Figure 18.4. Type 23 frigate ship's control centre (Reproduced with permission from MoD/BAE SYSTEMS Marine).

Figure 18.5. Bridge of a Type 23 frigate (Reproduced with permission from MoD/BAE SYSTEMS Marine).

As already stated, in evolving from the Type 23 Bridge and Control Centre to the Future Surface Combatant equivalents, the main development area will involve deciding what information is necessary at each of these positions to ensure efficient operation of the vessel and how it is displayed.

It is expected that multi-function display screens will replace discrete consoles and instrumentation saving space and minimising hardware requirements. However, moving to a system where data will be schematically displayed on personal computer (PC) screens brings its own problems, such as colour clarity. In a combat scenario the vessel may switch from white to red lighting which can affect the way colours appear on a display screen.

One convention is for schematic objects to be shown as green when operating within normal parameters, turning to orange if a warning state occurs then changing to red if that object enters an alarm state. These three colours could easily be confused as the background light alters so a change in colour is not sufficient. One proposal will be for the shape or design of the object on the schematic to be altered, as well as its colour, to indicate a normal, warning or alarm state.

18.6 New technology

The demand to reduce the ship's complement on future vessels will necessitate multi-disciplinary crews. Subsequently, training requirements will have to be integrated across traditional naval disciplines. This blurring of engineering trade boundaries will result in a broader knowledge base but with reduced system expertise. Modern technology must therefore be implemented to reduce the demands of specific expert knowledge on the user and enhance the capabilities of the individual and consequently, the ship. The inevitable increase in workload must be offset by automation of tasks and fusion of data representation along with the improvement of Man–machine interface and human environmental conditions.

18.6.1 Ship's bridge technology

The highlighted problem of poor all round vision can be remedied on the FSC by utilising a 360° field of view bridge. As the mast is required to be the highest part of the superstructure, this all round view can only be achieved by integrating the bridge into the mast arrangement. Until recently, this was impossible for two main reasons. The weight of mast equipment required a dense support structure and bulky waveguide runs for the radar arrays took up internal space. These obstacles are being overcome by a number of design innovations including the use of next generation active phased arrays within the naval radar system, and utilising fibre optics to remove the need for large waveguide runs from the superstructure to the mast head.

The enclosed construction, utilising composite materials technology, also provides the ability to accommodate heat exchangers and cooling systems above the 360° bridge area. The combination of these technologies will minimise space requirements for vertical conduit. Thus, the bridge can be placed at the bottom of the mast with windowed sections all around.

This design will greatly improve human surveillance for the control and management of the ship during peacetime and tactical operations. Hence, one observer can potentially monitor the observation of the above deck ship's activities. As illustrated previously in Figure 18.3, this two-state bridge design incorporates a lower level enclosed bridge – the command bridge – and an upper level cockpit with 360° vision. Both the bridge and the cockpit utilise fully integrated bridge systems.

18.6.2 Command bridge

The lower level command bridge, illustrated in Figure 18.6, will have accommodation for four seated personnel and is based on the efficient bridge layouts being put forward by agencies such as Lloyds and Det Norsk Veritas. Each position is electronically and operationally interchangeable and designed for use principally during non-combat situations. This concept is analogous to the aircraft 'glass cockpit' design.

The forward port position includes full conning controls, while the two rear positions feature flat screen multipurpose work-stations. These stations will feature a combination of ergonomically laid out touch sensitive screens, soft key controls for efficient interface, and hard-wired controls for essential services in case of electronic breakdown. This layout maximises the display of tactical picture information.

It is suggested that touch screen technology is not suitable for naval use as the action of touching one particular section of mimic screen may become extremely difficult in a heavy sea state or combat situation. However, replacing large metal control panels with multi-function screens will remove large sections of hardware, saving valuable space and weight. Therefore, a compromise is required, accommodating the versatility of soft key multi-screen control while avoiding the hazards of bad commands due to fingers slipping on the screen. The solution appears to be to utilise work-stations featuring arrays of hard keys adjacent to the screen that can be attributed soft functions dependent on the screen in use.

18.6.3 Cockpit

During combat scenarios, bridge manning will be reduced to a cardinal staff of two who will transfer to the cockpit while other essential crew disperse to distributed command headquarters. The cockpit, illustrated in Figure 18.7, is designed to accommodate just the OOW and a fellow officer to act as a back-up observer. Utilising modern communications technology, these are the only personnel

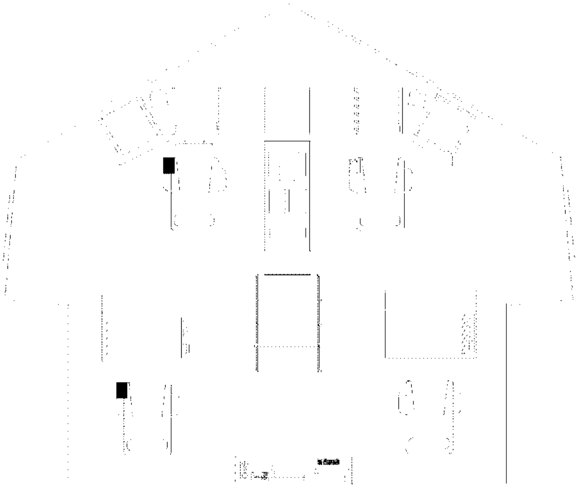


Figure 18.6. Command bridge layout.

required for the safe ship bridge operations, leaving lower bridge manning unnecessary.

The IPMS will facilitate contact between the cockpit, the commanding officer, and the distributed command centres. The actual cockpit has been arranged to give the OOW full control of the ship including navigation, steering, machinery control and communications. This has additional benefits with combat scenarios, such as potentially freeing the command bridge for training while maintaining operational control from the cockpit.

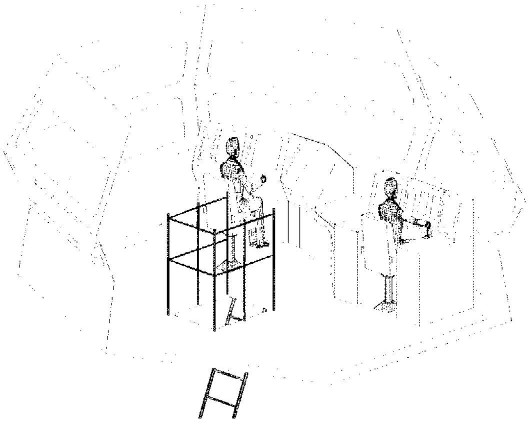


Figure 18.7. Cockpit.

The OOW will have in front of him/her, three multi-function colour displays with pre-set display configurations dependent upon the present ship's operating mode. Configuration will be controlled by a bridge computer that is operating a knowledge-based system to determine optimum display configuration. The importance of the knowledge-based system is in filtering data so that, in this instance, the OOW is fed information that is critical to his/her duties at that time.

The operator will at all times have access to a manual display override via keyboard or tracker ball control. In the cockpit, the ship's wheel is replaced with an ergonomic joystick and all engine controls and emergency communications are at arms' length. To reduce the human workload further, a voice control will be provided enabling the OOW to orally select a communication channel or change to an alternative configuration option of the multi-function displays.

The cockpit has been designed specifically for maximum naked eye visibility but this may be enhanced using a Helmet Integrated Display System (HIDS). The HIDS may be worn by the OOW as s/he sits in an optimal viewing position. The relative positioning of the OOW's head within the conning bridge can be monitored in real time by an optical helmet tracking system and the information relayed to the bridge computer for correlation with surveillance equipment data.

The HIDS is a second-generation binocular, wide field of view display system. Raster graphics imagery and symbology display can offer all weather, day and night capability to superimpose the IR surveillance picture upon the wearer's field of view. Target designation will also be possible for light calibre gun control and this system will also have the capability to provide replication of the OOW's field of view to any integrated terminal on the ship. This will be most useful for relaying real time visual information to the captain. Conversely, ship sensor operators will be able to use the helmet display to direct the OOW's field of view onto the position of a detected object requiring visual conformation.

18.7 Integrated platform management system

The IPMS network, illustrated in Figure 18.8, will integrate all the ship's subsystems by utilising a high-speed fibre optic Asynchronous Transfer Mode (ATM) backbone architecture connected via bridges to individual system LANs. These will also utilise high speed ATM. To maximise survivability, each of these LANs will be capable of functioning individually if isolated from the backbone due to fault or damage.

The LANs will use technology such as the SONET STS12c, at 622Mbps, and ATM switching whose 53-byte fixed-packet transmission facilitates efficient data transfer due to the short packet size. Using ATM, the delay due to the passing of other packets is so short that the multiplexed data 'appears' uninterrupted upon reception. This ensures virtual real-time linkage. ATM switching ensures data transmission occurs only along the links between the transmitting and receiving nodes. This form of switched architecture ensures that every network node has contention-free access.

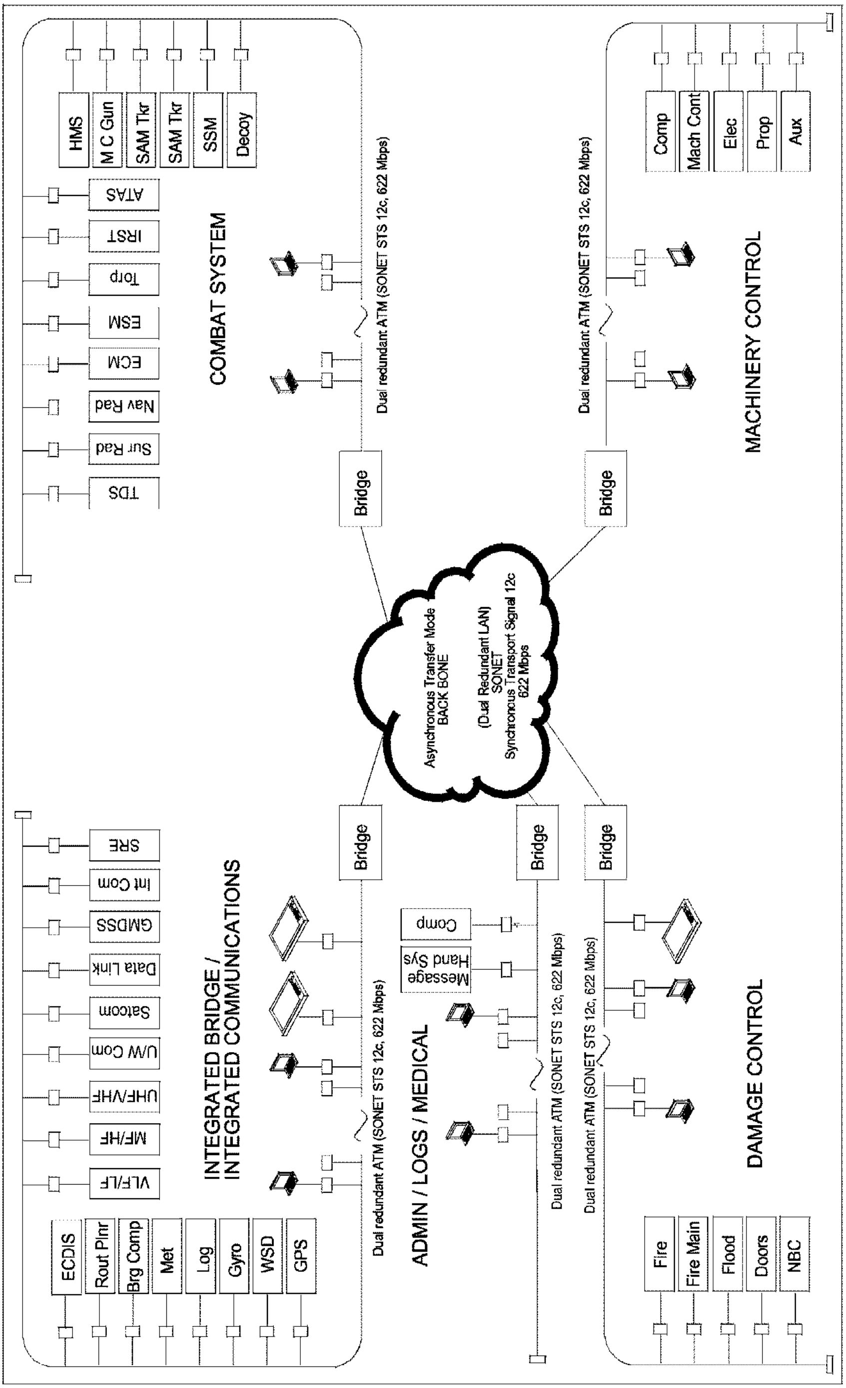


Figure 18.8. Integrated platform management system.

All user terminals and displays will have the potential to access any information available across the network. Access level security will be imposed in the form of operator authentication codes. The physical format of the display/control device will be determined by its operating environment, primary role and ergonomic requirements. A major modification to the IPMS will be the incorporation of multiple tap-in points. Thus, equipped with some form of portable work-station, personnel will be able to access the network from a multitude of points spread throughout the ship simply by plugging in and supplying a security password.

The Combat System will incorporate fusion of ship sensors and external data link tactical information in virtual real time. As stated, the Integrated Bridge System will supply both bridges with the required navigation, conning and machinery control information, at any multi-function display, suitably filtered by a knowledge-based control system.

Other networks will be attached to the backbone – such as Administration, Medical and Logistics – with appropriate priority levels for data transfer allocated to each facet dependent on the situation. It is proposed that a satellite data communications link will be utilised to remotely access databases for diverse information such as personal email, condition-based monitoring of machinery, medical history records, *etc.*.

In addition to the Integrated Bridge system, the Machinery and Damage Control systems will be monitored from strategically placed Distributed Command Information Units (DCIUs). These DCIUs represent the concept of distributed control. During peacetime the ship will be monitored and controlled from the bridge with perhaps one active subsidiary DCIU sited conveniently for whatever maintenance or training is being performed. In combat situations, however, control and surveillance will be spread to a number of DCIUs strategically placed throughout the vessel. Thus, the traditional single Command Information Centre is replaced by a series of smaller Command Information Units dispersed to different locations on the ship.

These DCIUs will co-ordinate ship tactical operations and each unit will be electronically compatible, equipped with multi-function terminals. Therefore, being interchangeable, they will be less vulnerable to major damage from one direct hit when under attack. The vessel's commanding officer will be able to maintain overall control of the ship from any one of these DCIUs. Again, knowledge-based systems will be utilised to minimise dissemination of surplus data.

To minimise the risk of loss of life, it is proposed to replace life rafts with fully enclosed, self-righting, escape pods as presently used on oil production platforms. One option being considered is the siting of DCIUs inside one or more of these evacuation pods. Hence, a critical control unit can be permanently based in an evacuation pod and control the ship until they actually eject from the vessel. This pod may even be equipped with certain remote control capabilities for 'last gasp' operation of the vessel even after it has been evacuated. Traditionally, it is assumed that a vessel will only be abandoned when all hope of saving it is lost. However, steering a badly damaged vessel away from the rescue zone or manoeuvring her around to provide cover from weather or hostile fire, could enhance the safety of the abandoning crew.

18.8 Conclusions

This chapter has attempted to illustrate how the disciplines of human factors design are being applied to hitherto very traditional areas of naval engineering. Human factors engineering is now well established as an invaluable tool in the quest to develop faster and more efficient warships. More than ever before, achieving budget constraints is the main target for new ship designs. This extends beyond the expense of initial construction to the through-life costs. Personnel present the major through-life cost and therefore any reduction in ship's crew will provide substantial dividends in the battle to reduce expenditure.

Reducing crew has the supplementary benefit of placing fewer human lives at risk but highlights the problem of operating more sophisticated vessels using fewer staff. Although aimed at a ship that will not go into service for eight years, the proposals outlined in this chapter are based on technology which is in an advanced stage of development at present. It is the authors' opinion that in the coming years, hardware will not alter radically but the operating software will become faster and more versatile. The area to investigate is therefore not how to build faster computing systems but how to fashion future control systems such that the human operators can operate them faster and more efficiently.

As a footnote, the June 2000 issue of *Jane's Navy International* carried a photograph of the French Navy's Georges Leygues-class frigate *Dupleix*. The *Dupleix* has just recently – in late 1999 – completed an 'air defence upgrade' featuring the addition of a second tier large command structure above the bridge. Furthermore, the RN's newest vessel, the Type 45 Destroyer, is being designed with a Platform Management System featuring an ATM-type fibre-optic LAN, multiple plug-in points, multifunction consoles and distributed control positions in all fire zones. The RN is currently developing new training regimes to cope with the increase in ship technology.

The views expressed in this chapter are those of the authors and do not necessarily represent those of BAESYSTEMS Marine Ltd.

Further reading

1 CRAMPIN, T., FOXHALL, S., and CORBRIDGE, C.: 'Human factors design of future ship control centres'. Proceedings of 11th Ship Control Systems Symposium, 1997, **2**, pp. 93–103

2 SMYTHE, C.: 'Networks and their architectures', *Electronics and Communication Engineering Journal*, 1991, pp. 18–28

3 SQUIRES, M. A., and MATTICK, Cdr D.J.: 'Survivability technology for the 21st century warship', *INEC 98, Paper 3*, 1998

Index

People in Control

Human factors in control room design

As industrial processes have become more automated, there is increasing concern about the performance of the people who control these systems. Human error is increasingly cited as the cause of accidents across many sectors of industry.

This book provides state-of-the-art information on various aspects of human–machine interaction and human-centred issues encountered in the control room setting. Subject coverage includes vigilance and human error in control room situations, analysis and training of control room activities, and control room design including alarm systems.

Based on a successful multi-disciplinary IEE conference and illustrated with useful casestudies, this book is essential reading for all students, engineering professionals and managers interested in human–machine interaction and human performance and ergonomics within the control room setting.

Jan Noyes is Professor of Human Factors Psychology at the University of Bristol. Her research interests include the human factors of advanced and emerging technologies. She is a Fellow of the Ergonomics Society and an Associate Member of the IEE. In 1999, she was awarded the Otto Edholm medal for her contribution to ergonomics application and research. She has written over 120 publications including five books, and was awarded the IEE Informatics Premium Award for her paper on 'engineering psychology and system safety'. She was also Chair of the 1999 and 2001 IEE People In Control conferences.

Matthew Bransby was a leading authority on control room alarm systems. His career included four years lecturing in Control Engineering at Sheffield University, 22 years in the electricity supply industry and three years as an engineering consultant. He was involved in implementing novel control systems and running major process control projects. His work on alarm systems included carrying out research for the Health and Safety Executive in the UK, and writing an industry best practice guide. He acted as an Expert Witness in process control and was a Fellow of the IEE. He died in 2000.

ISBN 978-0-85296-978-6
9 780852 969786 >

The Institution of Engineering and Technology
www.theiet.org
0 85296 978 3
978-0-85296-978-6