EVALUATION OF HUMAN WORK

FOURTH EDITION

Edited by

JOHN R. WILSON

SARAH SHARPLES

CRC Press
Taylor & Francis Group

EVALUATION
OF
HUMAN WORK

FOURTH EDITION

EVALUATION OF HUMAN WORK

FOURTH EDITION

Edited by
JOHN R. WILSON
SARAH SHARPLES

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2015 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20150309

International Standard Book Number-13: 978-1-4665-5962-2 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To John. We all miss you and thank you for your friendship
and contribution to our work and lives.

To Craig, Ellen and Luke. Without you, none of this would have been possible.

xxx

Contents

SECTION I Fundamental Approaches and Tools

SECTION II Methods for Design of Work

SECTION III Assessment of Impact of Work on Individuals

SECTION IV Assessment and Design of the Physical Workplace

SECTION V Methods in Organisational and Social Human Factors

SECTION VI Implementation and Analysis of Human Factors

SECTION VII Conclusion

Preface to the Fourth Edition

It was an honour and a privilege to be asked by John to join him as an editor of the fourth edition of this text, and it is with great sadness that I have completed the task alone since John's death in July 2013. Before his death, we had discussed the revised content and structure of the book, and John had led the commissioning of the revised chapters to be submitted. He had also begun work on revising many of the chapters on which he was a co-author, and we had discussed some of the changes we would make to the structure and content of the text. We particularly discussed the introductory Chapter 1, which provides an overview of our approach to methods in Ergonomics/Human Factors (E/HF).

In reviewing the previous editions of this book (and it is worth noting that the first edition was published whilst I was still at school and I am ashamed to say I had not yet heard of ergonomics), it is clear to me that this edition is in fact the most comprehensive update that the book has had. All chapters have been revised, most of them with significant changes of authorship or content. It is interesting to see many other authors, perhaps like me, who have worked with some of the original authors of the chapters over recent years and are now contributing to the writing of this text. There are five completely new chapters, covering inclusive design (Chapter 11), situation awareness (Chapter 19), neuroergonomics (Chapter 22), ergonomics and quality (Chapter 36) and standards in ergonomics design and evaluation (Chapter 37). The final chapter (Chapter 38) is also new and considers the notion of reflective practice, which underpins our work as E/HF practitioners. Most chapters have new authors and co-authors, and much of the content is substantially changed and updated. Three chapters on which John was originally the lead author (Chapters 9, 14 and 35) have been taken on by new teams.

Much continues to change in the discipline of E/HF. John highlighted the interest in the systems perspective in the preface to the third edition of this text; this systems perspective is now prevalent within much E/HF practice and underpins many of the discussions within this book, as well as being the particular focus of Chapters 9, 28 and 33. Further trends include the increased availability and utility of physiological devices. Tools that were once solely for use in a laboratory context are now feasible to be applied in field contexts; it is therefore essential that we understand the nature of these tools, the quality of output they deliver and the relationship between these physiologically measured parameters and the theoretical concepts, such as workload, fatigue, stress or effort, to which we wish to relate them.

The name of this text is, and has been for the three previous editions, *Evaluation of Human Work*. In the preface to the first edition, the editors note that 'we are primarily concerned with people at work' but that this does not rule out 'contributions relevant to people's activities at home, leisure or on the road'. This is even more important in today's society with its blurring of the home–work boundary – not only do we now see many people who work from a home environment, but we also see how work technologies are appropriated for home use, and vice versa.

A final extremely important point to note, which reflects the increased and recognised importance of the discipline of E/HF, is the 2014 award of a Royal Charter to the United Kingdom's Institute of Ergonomics and Human Factors. This status is only awarded to those organisations that represent a field that is 'unique' and represent a tremendous endorsement of the work that E/HF practitioners do all over the world to improve the way in which people work and interact with systems and technologies.

I read the first edition of the book as a student of human factors in 1994, little expecting to one day be its editor. Its appeal to me was in its detail, opinion and guidance – at the end of reading a chapter, I felt as though I could understand the context of 'how' to practise ergonomics and what aspects of method selection and implementation I should consider. Whilst working on the book I have been both intimidated and pleased by the response of those who have commented on it being their favourite ergonomics/human factors text. I hope that this revision meets their expectations and retains the unique place the text has amongst the rich set of books that are now available for those practising E/HF.

Some of the elements of the book were very much John's and reflected the thoughts that I am sure he and Nigel Corlett, the co-editor of the first three editions, discussed and debated many times. I was fortunate to also engage in such debates with John, and there were certainly points on which we did not always agree! I have tried to reflect his views where I can but also have included views and interpretations of my own; occasionally, and particularly in Chapter 1, I note this explicitly in footnotes, sometimes interchanging between 'I' and 'we' as pronouns.

This book originally emerged from a now infamous gathering of some of the leading minds of the world of ergonomics for the Second International Occupational Ergonomics Symposium in the mid-1980s at Zadar, Yugoslavia. Just a few weeks before the submission of this volume, it was lovely to gather many other like-minded experts in E/HF, including some of those who attended that original meeting, for a conference on human factors in complex systems, organised in memory of John, at the University of Nottingham. Along with recollections of John and his contribution to E/HF, there was much lively critique and debate of many of the concepts presented in this text; some of these are reflected in the final chapter of this book.

Sarah Sharples

MATLAB® is a registered trademark of The MathWorks, Inc. For product information, please contact:

The MathWorks, Inc.
3 Apple Hill Drive
Natick, MA 01760-2098 USA
Tel: 508-647-7000
Fax: 508-647-7001
E-mail: info@mathworks.com
Web: www.mathworks.com

Acknowledgements to the Fourth Edition

It has been a difficult but immensely rewarding task to complete this book, and it would not have been possible without the support of all of the chapter authors, who have been so diligent in updating and writing their chapters and responding to my requests throughout the editorial process. In particular, Colin Drury, who has provided support and content for this volume from its first edition, has been a tremendous moral supporter and advisor to me in the final months leading up to submission of the text. The publishers at Taylor & Francis Group/CRC Press, especially Cindy Carelli, Paul Abraham and Jennifer Ahringer, have been very patient with responding to my requests and supporting me through the mammoth task of collating and submitting this volume to them.

I have had significant support from the University of Nottingham in enabling me to squeeze the time to complete this book. I particularly thank the Department of Mechanical, Materials and Manufacturing Engineering for allowing me to cover some of my teaching duties whilst working on the book and thank those who delivered my teaching and administration duties in my partial absence. The research grants EP/G065802/1, EP/G037574/1 and EP/K014161/1 provided financial support to enable this work.

I am particularly thankful to colleagues in the Human Factors Research Group for their encouragement and work to help me complete this book that has been so important and useful to so many of us over the years. Many group members helped review the content of the methods table in Chapter 1 (in particular, Sally Shalloe, Alyson Langley, Harshada Patel, Julie Waldron, Brendan Ryan, Laura Lewis and Rob Houghton provided useful input and critique). Thanks also to Cath Harvey, Michael Brown, Genovefa Kefalidou, Liz Dowthwaite, Lizzie Crundall, Anne Floyde, Richard Eastgate and Will Knight for help with proofreading. Thank you also to Lynne Mills who provided me with some aspects of the 'history' of the volume which were very useful when putting all of the elements together. Credit must also go to Heidi Freestone who has battled with my diary and protected my time as much as she could, dealing very patiently with the different demands she received, keeping my students calm and responding to my numerous last-minute requests!

I have also received informal support from many colleagues within the E/HF community, particularly Chris Baber and Peter Buckle, who have provided me with informal comments and thoughts about the book that have proved very useful to my editorial process. Thanks also to David Gilmore who, perhaps unwittingly, over a glass of wine at the annual Ergonomics and Human Factors conference in the United Kingdom in 2014, provided us with the last line of the final chapter of this book.

I had worked with John for almost 20 years (which proved helpful when having to interpret his handwritten notes on drafts of the chapters!), and this is still very much his book. His legacy to the discipline of ergonomics and human factors is primarily through his influence on the people he taught and worked with but is also through his writing. This text is probably the most important of all of the things he wrote, and I hope he would have been proud of this fourth edition.

Finally, I thank my husband, Craig, and my children, Ellen and Luke, for their love and support during what has been a very busy and demanding time. You have been extremely patient whilst I have worked in the evenings and weekends. I'm sure you will be very pleased to hear that the book is finally finished!

Sarah Sharples

Preface to the Third Edition

Ergonomics/human factors is a fast-moving discipline. The domains in which we work and can see demonstrable success from the application of our knowledge and skills are constantly expanding. At the same time we are continually improving and enlarging our methodological base, developing new theories, approaches, methods and tools as well as refining ones we have used before. This third edition of *Evaluation of Human Work* has been published now to try to reflect many of these changes. Therefore in every case bar one, the chapters from the second edition have been substantially revised, and there are several new chapters.

In general, the completely new chapters reflect the growth of interest in a systems viewpoint within ergonomics/human factors, and particularly understanding physical and cognitive work in the context of the social and organisational setting in which it takes place. Thus the new chapters include ones on sociotechnical design of work systems (Chapter 29), team design and evaluation (Chapter 30), learning from failures through a joint cognitive systems perspective (Chapter 34), and the analysis of organisational processes (Chapter 38).

In addition to these new chapters, several chapters have not just been revised in terms of updating material but with a completely new approach. Thus the chapter concerned with user trials in the second edition has been expanded into one which addresses many techniques in user-centred design (Chapter 11). The chapter on knowledge elicitation for expert systems has been rewritten to reflect the increased interest in understanding the nature of knowledge and of knowledge management in contemporary systems (Chapter 8). The opening chapters on assessment and design of the physical workplace to do with the environmental factors of climate, visual conditions and noise have been accompanied by a short chapter discussing environment surveys generally (Chapter 22). Finally, reflecting what is current practice, the chapter on accident reporting and analysis has been replaced by one on systems for near miss reporting and analysis (Chapter 33).

One thing which has remained unchanged from the first and second editions is that this text is produced *not* as a cookbook of ergonomics methods. Whilst there is a place for handbooks and manuals which describe each of hundreds of different ergonomics/human factors tools and techniques, with some form of lookup table for users to select amongst them, that is not the intention and thrust of this book. Rather the intention is to place ergonomics methodology in context, and each chapter carefully describes the background to method development in that area and to the application of methods and tools. In this way, it is intended to make the text suitable for teaching on ergonomics and human factors courses beyond those purely to do with methods of analysis and evaluation, and to try to introduce the topic of ergonomics/human factors from a "doing it" perspective. It is to be hoped that we have succeeded in this aim.

John R. Wilson

Acknowledgements to the Third Edition

First of all, I thank all the authors for their great effort in updating chapters or writing new chapters. Particularly for those authors who have been with us since the first edition, I recognise that there is little in it for their careers to publish a revised version of a chapter they have already had in print, and therefore I am extremely grateful for the diligence and care with which everyone has produced revisions. In addition, I thank the anonymous reviewers who made suggestions as to how to improve this third edition and also to all readers and users of the first and second edition who have sent comments and criticisms to the authors and editors.

For this edition, the production of the book has moved from Taylor & Francis Group in the United Kingdom to CRC Press in the United States. It is a measure of the professionalism of all involved that this transition (at the time of writing at least!) appears to have taken place very smoothly. At the publishers, we would like to thank Cindy Carelli, Helena Redshaw and Rachel Tamburri-Saunders for all their help in the production of the book. I also thank Chris Stapleton yet again for her work on the production of the index for the book.

A final thank-you goes to all my colleagues in the Institute for Occupational Ergonomics at the University of Nottingham, and those in other institutions with whom we collaborate. Without working with colleagues who have enthusiasm for the development of new knowledge in ergonomics/human factors and the application of new and existing methods, I would have little enthusiasm to produce books of this nature. I very much welcome the professionalism and friendship of all my colleagues. Finally, I must yet again thank very much my secretary Lynne Mills for all her hard and insightful work on this book, greatly enhancing its production as near to on time as possible and the quality of the content, as well as acting as a buffer between myself and the authors and publishers.

John R. Wilson

Preface to the Second Edition

Since the first edition of *Evaluation of Human Work* was published, much has happened to change the way we view ergonomics methods and techniques. This has led to the inclusion of several new chapters in this second edition, and the considerable revision of many others.

Technical, social, political and legal changes have required continual development and improvement of ergonomics methods. For instance, the ever-increasing power and prevalence of computer systems and the diversity of their user interfaces necessitate parallel improvements in methods of analysis, design and evaluation. We can see this, for instance, in human-computer interfaces in general (revised Chapter 12) and in specialised applications such as control rooms (new Chapter 13). Social and political changes, mirrored by changes in the way industrial work and jobs are organised, have increased recognition of the gains possible from greater involvement of people in what they do and from providing employees with a greater degree of control over their own activities. One manifestation of this is participation, and participative approaches (new Chapter 37) have a long and honourable tradition in ergonomics. Legal developments have had a profound influence on ergonomics in recent years, especially the health and safety regulations governing use of display screen equipment, manual handling work, work equipment and workplaces, which have come into force in Europe, Australia and to an extent in North America. Coupled with costs of compensation claims, such regulations have required structured ergonomics assessments at work (new Chapter 30) within an ergonomics management programme (revised Chapter 1 and new Chapter 35). In addition to the above, this edition includes a new chapter on measurement of physiological functions (Chapter 29) and substantial revision of the chapters on task analysis (Chapter 6), verbal protocol analysis (Chapter 7), product assessment and user trials (Chapter 10), knowledge elicitation (Chapter 14), computer aided methods (Chapter 20), mental workload (Chapter 25) and work stress (Chapter 26).

There are increasing moves towards greater professionalisation in ergonomics, for example the Board of Certification in Professional Ergonomics (BCPE) and the Centre for Registration of Ergonomists in Europe (CREE). Such moves require a recognition that the methods we choose will influence what we find from any investigation, and that methods must produce findings that are valid, reliable and generalizable; meet the objectives of the investigation; and are safe and ethical to apply. There can be little excuse for administering questionnaires that make no attempt to use previously validated scales, carry out experiments without careful piloting, or make rigid, ill-informed use of assessment checklists.

Any experimentalist should have a good knowledge of statistics. We decided, after long discussion, not to include statistics in this book – it is heavy enough already! Statistics are necessary not only for experimental design, data compression or testing of results, but for the understanding they bring to the nature of variability and the importance of interactions. Although methods are reported in a ‘stand-alone’ manner, it is rare in ergonomics to find an influence or a cause which has an exclusive relationship with an effect. Hence it is important to retain the ergonomics approach of viewing the whole person within the total environment, an approach which will make it necessary to match a selected group of methods to the requirements perceived. We hope this book will assist ergonomists to do this.

Acknowledgements to the Second Edition

For this second edition we would like to thank all our authors for their considerable efforts in updating their existing chapters or writing new ones; thanks also to reviewers of the new chapters. Many readers of the first edition have sent in comments and criticism to the authors and editors, and these have been accounted for wherever they improved the book.

Again, we must acknowledge the support of Taylor & Francis Group, and especially Richard Steele and Robert Chaundy, and thanks to Chris Stapleton for her professional service on proofreading and the index. The first editor (JW) gratefully acknowledges the Department of Safety Sciences, University of New South Wales, for the space and time to carry out most of the editing and writing of the new chapters.

Finally, we are very grateful to all our colleagues in the University of Nottingham's Department of Manufacturing Engineering and Operations Management, and especially in the Institute for Occupational Ergonomics, for their collaboration, support and a positive environment. Of these, Lynne Mills has contributed the lion's share in terms of typing, editing and organisation – of the book and of us!

John R. Wilson
Nigel Corlett

Preface to the First Edition

For a long time there existed few books on ergonomics or human factors methodology; Chapanis' *Research Techniques in Human Engineering*, published in 1959, was probably the earliest, as well as the best known. Lately, there has been a slow increase in what is available. For instance one of the contributors to this volume, David Meister, has produced two books dealing with methods (Meister, 1985, 1986) and the present editors have also been involved in two collections of conference proceedings concerned with new methods and techniques (Laboratory of Industrial and Human Automation, 1987; Wilson et al., 1987).

The books by Meister, excellent in many respects, concentrate upon investigations of large-scale (military) systems design, simulation and evaluation. The two sets of conference proceedings, whilst containing a range of methodological developments and applications, represent what was selected from the papers submitted for presentation, and do not fully represent the field. Also produced recently is the authoritative *Handbook of Human Factors*, edited by Salvendy (1987). This does have much to say about methods and techniques, both as separate chapters or as parts of other chapters; nonetheless its intention is to be a comprehensive, general text, with explanation of theories, principles, data and application, as well as of methods.

Our aim with this book on ergonomics methodology is to produce a text on methods and techniques that is both broad and deep. We intend it to be a companion to the major general textbooks on ergonomics and human factors, particularly and most recently those of Bailey (1982), Grandjean (1988), Kantowitz and Sorkin (1983), Oborne (1987), Salvendy (1987) and Sanders and McCormick (1987). All of these are well known to students, teachers and practitioners of ergonomics, as well as to many of those from other disciplines who take a personal or professional interest in ergonomics. There is, though, little opportunity in such texts to emphasise and make explicit the major part of methodology.

Therefore we have set out to produce a general text on ergonomics methodology. As the book's title implies, we are primarily concerned with people at work and with applied rather than basic research. However, the former concern has not ruled out contributions relevant to people's activities at home, leisure or on the road; nor does the latter concern invalidate descriptions of laboratory-based methods – these can have outcomes that are as practically applicable as are those from field investigations.

The contents of the book are intended to be interesting and useful for a wide range of people, including *students*, to give them a feel for ergonomics investigation and to complement their learning of theory and principles; *industrial and business personal at all levels*, to allow them to understand better what ergonomics can do for them, why and how; and *ergonomics practitioners, researchers and teachers*, to give them a compendium of methods and techniques available. For all these groups, the contributions here will also point to further sources for more detail on specific topics.

Our text on evaluating human work has brought together experts from many branches of ergonomics theory and practice, and has allowed them the space to introduce and give detail on those methods and techniques of value to them. Since ergonomics is both a science and a technology, these methods can of course be concerned with collecting data or with applying their own or others' data. The primary thrust of each contribution may be the general method (e.g. direct observation or protocol analysis), or particular fields of application for several types of methods (e.g. mental workload or the climatic environment). Whilst there will no doubt be omissions – of branches of methodology or of techniques within one area, regretted by some readers – we trust that most will find the book to be a comprehensive, readable and useful source of ergonomics knowledge and practice. Certainly we believe that for those students or readers from industry who are relatively new to ergonomics, one of the most interesting and valuable ways to learn about it is through its rich and varied methodology.

John R. Wilson
Nigel Corlett
University of Nottingham
Nottingham, United Kingdom

REFERENCES

Bailey, R.W. (1982). *Human Performance Engineering: A Guide for System Designers.* (London, U.K.: Prentice Hall), pp. 656 + xxviii.

Chapanis, A. (1959). *Research Techniques in Human Engineering.* (Baltimore, MD: John Hopkins Press), pp. 316 + xii.

Grandjean, E. (1988). *Fitting the Task to the Man: A Textbook of Occupational Ergonomics*, 4th edn. (London, U.K.: Taylor & Francis Group), pp. 363 + ix.

Kantowicz, B.H. and Sorkin, R.D. (1983). *Human Factors: Understanding People–System Relationships.* (New York: John Wiley & Sons), pp. 699 + xii

Laboratory of Industrial and Human Automatics (1987). New techniques and ergonomics. In *Proceedings of an International Research Symposium*. (Paris, France: Hermes).

Meister, D. (1985). *Behavioural Analysis and Measurement Methods.* (Chichester, U.K.: John Wiley & Sons), pp. 509 + ix.

Meister, D. (1986). *Human Factors Testing and Evaluation*. (Amsterdam, the Netherlands: Elsevier Science), pp. 424 + xi.

Oborne, D.J. (1987). *Ergonomics at Work*, 2nd edn. (New York: Wiley & Sons), pp. 386 + xvii.

Salvendy, G. (ed.) (1987). *Handbook of Human Factors* (New York: Wiley & Sons), pp. 1874 + xxiv.

Sanders, M.S. and McCormick, E.J. (1987). *Human Factors in Engineering and Design*, 6th edn. (New York: McGraw-Hill), pp. 664 + viii.

Wilson, J.R., Corlett, E.N. and Manenica, I. (1987). *New Methods in Applied Ergonomics.* (London, U.K.: Taylor & Francis Group), pp. 283 + x.

Acknowledgements to the First Edition

Our first debt with this book is to our contributing authors, all of whom have responded to our various requests with great patience, and have produced chapters of high quality within, in some cases, a very limited time. Amongst these authors we must mention those who were with us in the initial discussions about the book at the Second International Occupational Ergonomics Symposium at Zadar, Yugoslavia; they were Lisanne Bainbridge, Colin Drury, Ted Megaw, Ken Parsons, Pat Shipley and Rob Stammers. Colin Drury in particular has contributed much in terms of individual chapters, and the overall content and style of the book.

We thank our colleagues at Nottingham University for contributing to a working environment in which we feel able to embark on and complete this and other publishing ventures. One of us (JW) also thank the Department of Industrial Engineering and Operations Research, University of California, Berkeley, for allowing him time and facilities to work on this book during periods there as a visitor in 1987 and 1988.

Our editors at Taylor & Francis Group – David Grist, Sarah Waddell and, for most of the time, Robin Mellors – have been exceedingly supportive, even in the face of a project which seemed to grow exponentially! The style of the book has been enhanced tremendously by the artwork of Tony Aston and the cartoons by Moira Tracy. Despite both editors being away from Nottingham for substantial periods of time, the publication of their book has rolled on relatively smoothly; our colleagues would say this was because we left this and much else in the hands of our excellent secretaries, Lynne Mills and Ilse Browne, to whom we are immensely grateful.

Editors

John R. Wilson (Deceased) was a professor of human factors at the University of Nottingham, Nottingham, United Kingdom, and also worked for many years part time as a human factors specialist within Network Rail, the U.K. rail infrastructure provider. He was both a chartered psychologist and a chartered engineer, a fellow of the Institute of Ergonomics and Human Factors (president for 2010–2011) and a member of the HFES and Registered European Ergonomist. He was editor-in-chief of *Applied Ergonomics*. In 1995, he was awarded the Sir Frederic Bartlett Medal of the Ergonomics Society for services to international ergonomics teaching and research and was awarded the Distinguished Overseas Colleague Award of the U.S. Human Factors and Ergonomics Society in 2008. He served for many years on the Council and the Executive of the International Ergonomics Association.

Sarah Sharples is a professor of human factors at the University of Nottingham, Nottingham, United Kingdom. She leads a number of projects investigating novel technologies in workplace contexts, including transport, manufacturing and healthcare, collaborating with a wide range of industry and academic partners. She is a fellow of the Institute of Ergonomics and Human Factors and will be president of the Institute in 2014–2015. She led the Human Factors Research Group at the University of Nottingham to the award of the President's Medal from the Institute of Ergonomics and Human Factors in 2013.

Contributors

Chris Baber
School of Electronic, Electrical and Systems Engineering
University of Birmingham
Birmingham, United Kingdom

Nora Balfe
Department of Psychology
Centre for Innovative Human Systems
Trinity College Dublin
Dublin, Ireland

Ann Bisantz
Department of Industrial and Systems Engineering
University at Buffalo
The State University of New York
New York, New York

Tiffany Bisbey
Institute of Simulation and Training
University of Central Florida
Orlando, Florida

Peter Buckle
The Helen Hamlyn Centre for Design
Royal College of Art
London, United Kingdom

Gary Burnett
Human Factors Research Group
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Rebecca Charles
School of Aerospace, Transport and Manufacturing
Cranfield University
Bedfordshire, United Kingdom

Young-Jin Cho
ErgoMechanics Laboratory
Hanyang University
Ansan, South Korea

Sue Cobb
Human Factors Research Group
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Tom Cox
Centre for Sustainable Working Life
School of Business, Economics and Informatics
Birkbeck University of London
London, United Kingdom

Patrick G. Dempsey
Office of Mine Safety and Health Research
National Institute for Occupational Safety and Health
Pittsburgh, Pennsylvania

Colin Drury
Department of Industrial and Systems Engineering
University at Buffalo
The State University of New York
Buffalo, New York

Clive D'Souza
Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, Michigan

Ken Eason
Loughborough Design School
Loughborough University
Loughborough, United Kingdom

Jörgen A.E. Eklund
Unit of Ergonomics
School of Technology and Health
KTH Royal Institute of Technology
Stockholm, Sweden

Edward Elton
School of Computing, Engineering and Mathematics
University of Brighton
Brighton, United Kingdom

David Feathers
College of Human Ecology
Cornell University
Ithaca, New York

Mike Foley
Washington State Department of Labor & Industries
Tumwater, Washington

Jane Fulton Suri
IDEO
The Embarcadero
San Francisco, California

David Golightly
Human Factors Research Group
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Mike Gray
Health and Safety Executive
Sheffield, United Kingdom

Amanda Griffiths
Division of Psychiatry and Applied Psychology
School of Medicine
University of Nottingham
Nottingham, United Kingdom

Diane Gyi
Loughborough Design School
Loughborough University
Loughborough, United Kingdom

Christine M. Haslegrave
Human Factors Research Group
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Sue Hignett
Loughborough Design School
Loughborough University
Loughborough, United Kingdom

Robert Houghton
Human Factors Research Group
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Peter A. Howarth
Loughborough Design School
Loughborough University
Loughborough, United Kingdom

Harrison Kelly
Department of Industrial and Systems Engineering
University at Buffalo
The State University of New York
Buffalo, New York

Jung-Yong Kim
ErgoMechanics Laboratory
Hanyang University
Ansan, South Korea

David Kirk
School of Computing Science
Newcastle University
Newcastle upon Tyne, United Kingdom

Barry Kirwan
Eurocontrol Experimental Centre
Bretigny sur Orge, France

Kenneth R. Laughery
Department of Psychology
Rice University
Houston, Texas

Glyn Lawson
Human Factors Research Group
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Alyssa Marshall
Institute of Simulation and Training
University of Central Florida
Orlando, Florida

Christopher B. Mayhorn
Department of Psychology
North Carolina State University
Raleigh, North Carolina

Ian McClelland
Philips Digital Systems Laboratory
Eindhoven, Netherlands

Hilary McDermott
School of Sport, Exercise and Health Sciences
Loughborough University
Loughborough, United Kingdom

Ted Megaw
School of Electronic, Electrical and Computing Engineering
University of Birmingham
Birmingham, United Kingdom

Ranjana Mehta
Department of Environmental and Occupational Health
Texas A&M University
College Station, Texas

Seung-Nam Min
ErgoMechanics Laboratory
Hanyang University
Ansan, South Korea

Moira Munro
Glasgow, United Kingdom

Colette Nicolle
Loughborough Design School
Loughborough University
Loughborough, United Kingdom

Victor Paquet
Department of Industrial and Systems Engineering
University at Buffalo
The State University of New York
Buffalo, New York

Raja Parasuraman
Department of Psychology
George Mason University
Fairfax, Virginia

Ken Parsons
Loughborough Design School
Loughborough University
Loughborough, United Kingdom

Jonisha P. Pollard
Office of Mine Safety and Health Research
National Institute for Occupational Safety and Health
Pittsburgh, Pennsylvania

Jane A. Rajan (Deceased)
Rice University
Hertfordshire, United Kingdom

Emilie M. Roth
Roth Cognitive Engineering
Menlo Park, California

Brendan Ryan
Human Factors Research Group
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Eduardo Salas
Institute of Simulation and Training
University of Central Florida
Orlando, Florida

Nigel R. Shadbolt
School of Electronics and Computer Science
University of Southampton
Southampton, United Kingdom

Sally Shalloe
Human Factors Research Group
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Sarah Sharples
Human Factors Research Group
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Andrew Shepherd
Loughborough, United Kingdom

Carys E. Siemieniuch
School of Electrical, Electronic and Systems Engineering
Loughborough University
Loughborough, United Kingdom

Murray A. Sinclair
School of Electrical, Electronic and Systems Engineering
Loughborough University
Loughborough, United Kingdom

Gisela Sjøgaard
Institute of Sports Science and Clinical Biomechanics
University of Southern Denmark
Odense, Denmark

Paul R. Smart
School of Electronics and Computer Science
University of Southampton
Southampton, United Kingdom

Karen Søgaard
Institute of Sports Science and Clinical Biomechanics
University of Southern Denmark
Odense, Denmark

Tom Stewart
System Concepts
London, United Kingdom

Emile Tompa
Institute for Work and Health
Toronto, Ontario, Canada

Dwayne Van Eerd
Institute for Work and Health
Toronto, Ontario, Canada

Patrick Waterson
Loughborough Design School
Loughborough University
Loughborough, United Kingdom

Jennifer Watts-Englert
PARC, a Xerox Company
New York, New York

John R. Wilson (Deceased)
Human Factors Research Group
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Michael S. Wogalter
Department of Psychology
North Carolina State University
Raleigh, North Carolina

John Wood
CCD Design and Ergonomics Ltd.
London, United Kingdom

Paul H.P. Yeow
School of Business
Monash University Malaysia
Selangor, Malaysia

1 Methods in the Understanding of Human Factors

John R. Wilson and Sarah Sharples

CONTENTS

INTRODUCTION

It is common to introduce the idea of ergonomics, and of the importance of the human factor, by referring to one of a number of well-publicised disasters in which major loss of life and great commercial costs occurred – from Flixborough to Bhopal, from Three Mile Island to Chernobyl, from Piper Alpha to Ladbroke Grove. Many examples can be found, for instance, in Beaty (1995), Bignell et al. (1977), Casey (1993), Cushing (1994), Perrow (1984), Petroski (1992), Reason (1990) and Proctor and Zandt (2008). Recent examples include *Deepwater Horizon*, the aftermath of

Hurricane Katrina, the *Costa Concordia* capsize and the Santiago derailment and demonstrate the continued need to understand the role of design and the human factor in the management of safety and performance.

An incident that led to significant embarrassment for the Submarine Service within the British Royal Navy, but fortunately not loss of life, was the grounding of the nuclear-powered submarine, HMS *Astute*, off the west coast of Scotland in 2010. The official report into the incident revealed a number of contributing human factors, including: limited experience and lack of familiarity of the officer of the watch with the geographic location (the commanding officer was unaware of this deficit in geographical knowledge); lack of a dedicated plan or specific briefing for the upcoming transfer; absence of key members of the team (and a lack of awareness of the implications of this absence on team expertise); late-running of the submarine for a rendezvous (and the impact of this on choice of actions); missing equipment on the bridge within the submarine; faulty communications equipment and the primary radar not being used, because it was emitting noise (despite the fact that this 'did not affect the serviceability of the radar') (UK Government, 2012). The senior commanding officer (who was away from the bridge at the start of the incident) lost his post after the incident, but the fact that of the ten recommendations in the official report, four relate to personnel failure, two relate to process and four relate to equipment quality (including one specifying a need for better monitoring equipment to capture date in the case of such incidents) indicates that we are perhaps starting to move beyond the assumption that any incident that involves human factors should be wholly attributed to 'human error'.

Across a whole range of ICT (information and communications technology) implementation projects, we can find examples where: end users are not clearly identified and their needs not accounted for; where systems become excessively complex or have far more functionality than really needed; user training and support are not sufficient; interfaces hinder rather than help users; and the development process itself lacks clear objectives, is inflexible and is highly technology driven (e.g. Eason, 1997; UK National Audit Office, 1999; UK Public Accounts Committee, 2000). A very well known example, which broke most of the 'rules' of successful ergonomics, was the failed implementation of a new control system for the London Ambulance Service, which resulted in the loss of life and great financial costs (e.g. Wastell and Cooper, 1996; Beynon-Davies, 1999). The failure of the implementation of the FiReControl project in the United Kingdom (National Audit Office, 2011) wasted £469 million government money, attributed to 'underestimation of complexity of risk' of the project and the attempt to 'centrally impose a national control system on unwilling locally accountable bodies'.

We have in the past perhaps been guilty as a community of focussing on such examples of negative consequences to make the 'case' for E/HF. But the discipline has now existed for over 60 years, and E/HF has been implemented with much success in many domains. For example, the US Department of Labor maintains a list of 'success stories'* demonstrating the positive impact of E/HF interventions on employee health and performance. Identifying successful E/HF requires us to 'prove the null hypothesis' – demonstrating that an effect is absent, rather than present. We hear much less about air incidents, such as the fire after landing of Air France Flight 358 in Toronto in 2005, where all passengers successfully evacuated within 90s, than we do about an incident with a less positive outcome, yet it is highly likely that the successful evacuation was at least partly enabled by E/HF work by researchers such as Helen Muir, who conducted an extensive programme of research on aircraft evacuation after the Manchester aircraft Flight 28M fire in 1985 (Muir, 1996). The approach of focussing on when things go right, rather than when things go wrong, is embodied in the Safety II approach (e.g. Hollnagel, 2014).

Yet many examples of inappropriate design of equipment, workplaces, systems, jobs and organisations can still be found in large and small companies, in offices and factories and in physical and mental work. The common denominator in all cases is that the abilities, needs and limitations of the people working in the system or with the equipment have not been understood and accounted for in

---

* https://www.osha.gov/SLTC/ergonomics/success_stories.html.

the context of the demands of the job. On the other hand, successful products or work systems will usually show evidence that the needs of their users have been accounted for during design, implementation and operation.

Taking such account of people – or of the human factor – is the province of *ergonomics* and of the synonymous *human factors*. (The interchangeability of the terms is indicated in this chapter through use of the shorthand E/HF). This book is about the *methods* that ergonomists and human factors engineers and scientists use – in analysing, designing and evaluating equipment, tasks, jobs and organisations. Ergonomics has little value unless it is applied, and so its practical methodology is of great importance (see Dul et al., 2012 for further discussion of the 'value proposition' of E/HF).

The consequences of not applying E/HF, or of wrongly applying E/HF through inappropriate methodology, can lead to increased risk of ill-health and injury, dissatisfaction and discomfort for the workforce. For a company, the consequences at the least can be a loss of competitiveness, in terms of productivity, quality, flexibility and timeliness, as well as a loss of trust in the company, by either the public or its government. However, this book is even more concerned with the positive side of applying E/HF methodology, with the improvements in well-being that can result for employers, workforces, producers, users, engineers, designers and, indeed, people and society in general.

This application of methodology will be in both of what Kragt (1992) – in the context of manufacturing industry – distinguished as *product ergonomics* and *production ergonomics*, denoting the importance of considering design of products, objects or artefacts, as well as developing tools and methods to support the design of processes to design, produce and deliver these items. However, E/HF applies more widely than in products or production, increasingly focussing on service provision and systems. Goods and services are planned and designed by people, produced by people with processes built and managed by people, and are sold, bought, used, maintained and scrapped/recycled by people. Table 1.1 illustrates how E/HF might have a role at different stages in the traditional product life cycle. Even where automated processes are employed, these interface with people and organisational systems at some point, and once a product or system is purchased or implemented, the focus of effort may well be on maintaining and supporting a 'service' relationship with the consumer. Therefore, understanding of people, and using that understanding wisely, are central to the planning and running of successful organisations, communities and societies.

However, E/HF has an interesting challenge – its impact is often long term, rather than short term, therefore can be hard to assess on an instantaneous basis. The change from managerial capitalism to shareholder capitalism has forced companies to ignore long-term consequences to concentrate on this quarter's financial results, despite many more worthy aspirations in company documents. Even the continuous election cycle mitigates against long-term thinking.

This challenge also affects how we select, apply and interpret methods. For example, Mack and Sharples (2009) demonstrated that there were some differences between the extent to which people *thought* that E/HF (in this case, the perceived usability of a mobile phone) was important, compared to whether usability actually influenced their product choice. So increasingly perhaps, the role of the ergonomist is not only to enhance consideration of E/HF at all stages of the product life cycle, but also to clearly demonstrate to the user, purchaser or decision maker that this has been done, and what the improvement in product quality or system effectiveness was as a result.

Whatever the original roots of E/HF, as our home and work boundaries become increasingly blurred, it is now certainly not limited to the workplace, nor to 'the operator'. E/HF is relevant to all purposive human behaviour in 'designed' human–machine systems. In addition, increasingly the ability of an individual to customise, adapt and design their own work methods and workplace setup means we are unable to 'predict' the way in which people complete their work tasks. Therefore, E/HF has a role to move well 'beyond the workplace' and ensure not only that products are designed for use in a range of contexts, but that our methods consider the variety of ways in which work tasks might be completed. This means that we have to very carefully consider aspects of methods such as the generalisability and transferability of results.

TABLE 1.1
Role of Ergonomics and Human Factors throughout the Product Life Cycle

Stage in Product Life Cycle	Example Role of E/HF
Concept idea generation	Support creativity in idea generation Move beyond 'design blindness' caused by familiarity with previous products/design solutions Represent different stakeholder views Present guidelines (e.g. physical, cognitive capabilities; environmental considerations) in usable form to enable early consideration
Design	Enable collaboration and participation in design Provide tools to enable designers to appreciate user capabilities and preferences Support effective elicitation and presentation of user requirements Ensure monitoring and updating of user requirements through multiple iterations of design
Prototyping	Ensure selection of appropriate type of visualisation for stage in design process (early/late iterations) (e.g. using virtual reality, paper prototypes, physical mock-ups) Provide benchmarking against other parameters or data (e.g. a tool to enable workload of proposed system to be predicted, or a digital human modelling system that is used to measure range of population that will be accommodated by a system) Conduct user evaluations with prototypes to anticipate impact of changes on user and workplace performance, both by considering new product in isolation, and interaction between new product and other systems/artefacts in anticipated context of use
Manufacture	Design safe, comfortable and efficient manufacturing workplaces Identify points in production where errors could occur, and introduce elements to system to improve resilience Consider impact of automation in workplace, and understand how new automated and intelligent workplaces can combine with human operator's capabilities
Distribution and sale	Ensure prospective user is able to anticipate experience of product use as accurately as possible (e.g. allow prospective purchaser of a new car to visualise product quality from virtual representation) Clearly communicate product capability via product design
Use	Evaluate systems and tools *in situ* Collect opinions on use from users Measure impact of changes on operator/workplace performance
Maintenance	Design products to be 'maintainable' to minimise impact on overall work system Clearly communicate to the user which parts of a product or system are maintainable by the user themselves
Disposal/recycling	Label items and materials clearly to ensure appropriate recycling/disposal behaviours Make products easy to dismantle to make disposal/recycling as easy as possible

Note: Example activities are shown in the table, this is not intended to be an exhaustive list.

Thus, E/HF is highly relevant to, and for, people at home, on the move, in sport, at leisure as well as at work, and for school children, post-retirement populations and those with physical and cognitive impairments. Even at work, it is probable that a minority of people are working in jobs traditionally categorised as 'operators' (usually assembly line staff, machine operators or process control operators). People at work have a vast variety of roles, interact with a wide range of products and systems, have multiple goals and means and require integration of social, cognitive and physical skills, particularly in the service industries, where the customer also has a key role in the task.

Some Historical Background

Clarity of focus for E/HF has been bedevilled by the use of the different terms for the discipline itself. *Ergonomics* and *Human Factors* are now synonymous and accepted terms worldwide for the theory and practice of learning about human characteristics and capabilities, and then using that understanding to improve people's interaction with the things they use and with the environments in which they do so. The interchangeable use of both terms is demonstrated by the change in the name of the UK 'Ergonomics Society' to the 'Institute of Ergonomics and Human Factors' in 2009.

Distinctions are increasingly blurred. One difference is that people from many different disciplines will say they work in 'human factors' but generally only those with a degree in 'ergonomics' or, confusingly, 'human factors', will say they work in 'ergonomics'. There is a tendency, even amongst ergonomists, to talk about 'human (or people) factors' as being what they study, simply because this is easier grammatically in written and spoken language! The converse is that people working in the profession of human factors find it difficult to know what to call themselves, with 'human factors engineers' 'human factors scientists' and 'human factors professionals' all having been tried. Other efforts to get around the problem of name – for instance, using the title of engineering psychology – tend to lose the holistic and total systems flavour of E/HF. Interestingly, the current term used in many US bodies, such as the National Research Council, is 'Human Systems Integration' – a term that emphasises the systems approach, but perhaps distances itself from the underpinning *discipline* of E/HF.

The ease of use of term 'human factors' does have some unintended consequences though. It is increasingly being adopted by specific domain areas, thus we come across terms such as 'clinical human factors' or 'nuclear human factors'. This approach can result in a specific focus on a particular aspect of E/HF that is particularly relevant to the domain (anecdotally, in the safety critical context of healthcare, the focus of activity is often on the role of human communication, teamwork and collaboration for example). This may simply be a result of the strong influence of specific and high profile cases that have rightly emphasised the importance of human factors within these areas, but it is important that the breadth and systems approach of E/HF is not lost. The increasing profile of human factors within specific domains (particularly the healthcare domain, where there are many activities in Europe, US and UK to work with clinical professionals to increase consideration of E/HF in many types of clinical settings) is an excellent thing, and brings the active engagement not only of expert ergonomists but also clinical professionals, whose buy in is essential to ensure deep embedding of E/HF approaches within the clinical setting. This perhaps demonstrates the importance of us embracing the meaningfulness of the term 'human factors' to colleagues outside the core discipline, and making sure that the tools and methods we develop are usable by those who may not be 'trained ergonomists' but who do have vital familiarity with and tacit knowledge of their own workplace setting.

We may regret that the term 'ergonomics' lacks innate meaning or impact for clients or the public who will not be aware of its classical roots, but genies can rarely be put back into bottles; it would now be a futile and damaging exercise to move away from use of the terms ergonomist, ergonomics and human factors. In this book, we emphasise that it is what we do that matters, not what we call ourselves, and contributors will use the terms ergonomics and human factors interchangeably. We must work in teams as systems engineers, human factors integration specialists, capability and performance managers, to embed E/HF tools and practice within different industrial contexts.*

Although formal consideration of the interactions between people and their working environments can be found in writings of a hundred years ago, for instance, from Poland and Germany, the modern history of E/HF emerges from the 1939 to 1945 war. As a formal branch of learning,

---

* I have often noted that I practise 'ergonomics by stealth' (Colin Drury uses the term 'guerrilla ergonomics') – sometimes by simply being part of a team, giving anecdotal examples of the impact of ergonomics over a coffee or lunch break, or just giving a different perspective on a design suggestion, can be an important part of encouraging colleagues from other disciplines to take E/HF into account (SS).

with its own learned societies and scientific journals, E/HF has a formal history of about 60 years, for instance in Germany, the Netherlands, the United Kingdom and the United States. In the United Kingdom, the ideas and expertise from different disciplines interested in the effectiveness of human performance (anatomy, physiology, industrial medicine, industrial hygiene, design engineering, architecture and illumination engineering), and an emphasis on theory and methodology, led to the formation of the discipline of ergonomics with two strong sub-groupings: those of anatomy/physiology and experimental psychology. In parallel, the human factors profession was growing up in the United States, with strong inputs from the disciplines of psychology and engineering. In Germany, the Netherlands and across Scandinavia, a basis for ergonomics was growing out of work in medicine and functional anatomy, whilst in Eastern Europe, growth was largely from the industrial engineering profession. For much more background detail, interested readers are referred to Edholm and Murrell (1973), Singleton (1982), Stockbridge (1989) and Meister (1995b) and histories by Waterson and Sell (2006) and Waterson (2011) for the Institute of Ergonomics and Human Factors and by Kuorinka (2000) for the International Ergonomics Association.

E/HF has drawn from anatomy, physiology and psychology, and has close connections with the applied disciplines of medicine and engineering. Chapanis (1996) defines E/HF as a multidisciplinary field, with psychology (primarily experimental psychology), anthropometry, applied physiology, environmental medicine, engineering, statistics, operations research and industrial design all contributing. Wickens et al. (2013) describe how the field of human factors grew from initial considerations of performance in military contexts, but has broadened greatly in the last few decades with its various sub-domains and application contexts (they note consumer products, business, highway safety, telecommunications and healthcare). Overlapping and related disciplines and names, often influenced by academic cultures as much as linguistic meanings of terms, include experimental psychology, engineering psychology, social psychology, industrial engineering, bioengineering, cognitive science, artificial intelligence, systems engineering, human-computer interaction, interaction design, industrial design, management and statistics.

DEFINITIONS OF ERGONOMICS/HUMAN FACTORS

It is a feature of the modern world that disciplines of current relevance and value are generally multi-, inter- and trans-disciplinary and therefore less amenable to simple definition. The fact that E/HF was built upon existing fundamental disciplines should not in itself be a problem: engineering is built upon mathematics and physics, psychology is built upon biology and economic science upon a number of bases.

E/HF seeks to define itself at regular intervals – for instance, amongst many others, see Chapanis (1976, 1979), Welford (1976), de Montmollin (1992), Moray (1994) and Meister (1998). A large number of different, if overlapping, definitions of ergonomics and of human factors now exist; Wogalter et al. (1998) considered these. Most definitions stress the view of E/HF as jointly a science – providing fundamental information – and a technology – applying that information to problems of design in their widest sense (e.g. Shackel, 1996). Within this view, the E/HF sphere contains all elements of the total human–environment system, comprising people's interactions with hardware, software and 'firmware' (including space), and with other people ('liveware') both individually and as social groups.

Any acceptable definition of E/HF must emphasise the need for, and the complementarity between, fundamental understanding of people and their interactions and practice in improving those interactions. Meister (1995a, p. 9) differentiates between the theoretical knowledge within ergonomics, which explains people's interaction with other things, and the instrumental knowledge, which can be utilised in design. This relationship between theory and practice, between research and application, is under examination continually (e.g. Singleton, 1994; Green and Jordan, 1999). A US National Research Council report (Rouse et al., 1997) makes strong arguments for the value of human factors initiatives in many industries, stressing its multidisciplinary, systemic, sociotechnical and user-centric orientation, but also the difficulties in the current climate of balancing needs for basic research with the high demands for applied activities. As Wilson (2000) notes 'we can see ergonomics as comprising elements of craft, science and engineering; it has aims to implement and evaluate (craft), to explain and predict (science), and to design for improved performance (engineering)'.

The International Standards Organization, in its various committees on ergonomics standards, has been using as a working definition that:

> Ergonomics produces and integrates knowledge from the human sciences to match jobs, systems, products and environments to the physical and mental abilities and limitations of people. In doing so it seeks to safeguard safety, health and well-being whilst optimising efficiency and performance.

Similarly, the International Ergonomics Association (as of 2014) has defined that:

> Ergonomics (or human factors) is the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theoretical principles, data and methods to design in order to optimise human well-being and overall system performance.

It is worth highlighting the final phrase *optimise human well-being and overall system performance* – it could be argued that this is really the 'charter' for this book, and that our goal is to measure all aspects of human well-being and system performance. The definition continues to explain:

> Practitioners of ergonomics and ergonomists contribute to the design and evaluation of tasks, jobs, products, environments and systems in order to make them compatible with the needs, abilities and limitations of people. Ergonomics helps harmonise things that interact with people in terms of people's needs, abilities and limitations.

When pushed at a party for rather more pithy definitions of ergonomics, many of us will respond that it is design for people, or designing fit for a range of human users, or fitting systems and products *to* people and not *vice versa* (Figure 1.1).

A prevalent current tendency to define and partition the work of ergonomists into specialisms, whilst understandable, can cause difficulties. Typically this is done into cognitive ergonomics, physical (or musculoskeletal) ergonomics and social (or organisational) ergonomics; other specialisations are also sometimes defined such as rehabilitation ergonomics, green ergonomics and forensic ergonomics. There are some good reasons for such partition. It can help explain a potential contribution to clients and funding bodies: for instance, an ergonomist specialising in complex control systems design may define what they practise as cognitive ergonomics; an ergonomist specialising in health and safety at work and re-design of workstations may define this as physical ergonomics. Nonetheless, there are dangers in such a parochial and molecular viewpoint. It is the very systems perspective and holistic nature of E/HF that provides its strength. The breadth of concern to cover all aspects of people's interaction with their environments and the interconnections between these interactions is what allows E/HF to define itself as a unique discipline, and is at the heart of the certification or chartership of many professional bodies associated with the discipline, such as CREE (Centre for Registration of European Ergonomists), the BCPE (Board of Certification of Professional Ergonomists) and the IEHF (Institute of Ergonomics and Human Factors).

If we do see the value in descriptions of specialisms, these might include:

- *Physical ergonomics*: fit, clearance, reach, access, tolerance, workload, manual handling, health and safety, workplace layout, displays and controls, product and equipment design, environment, tools

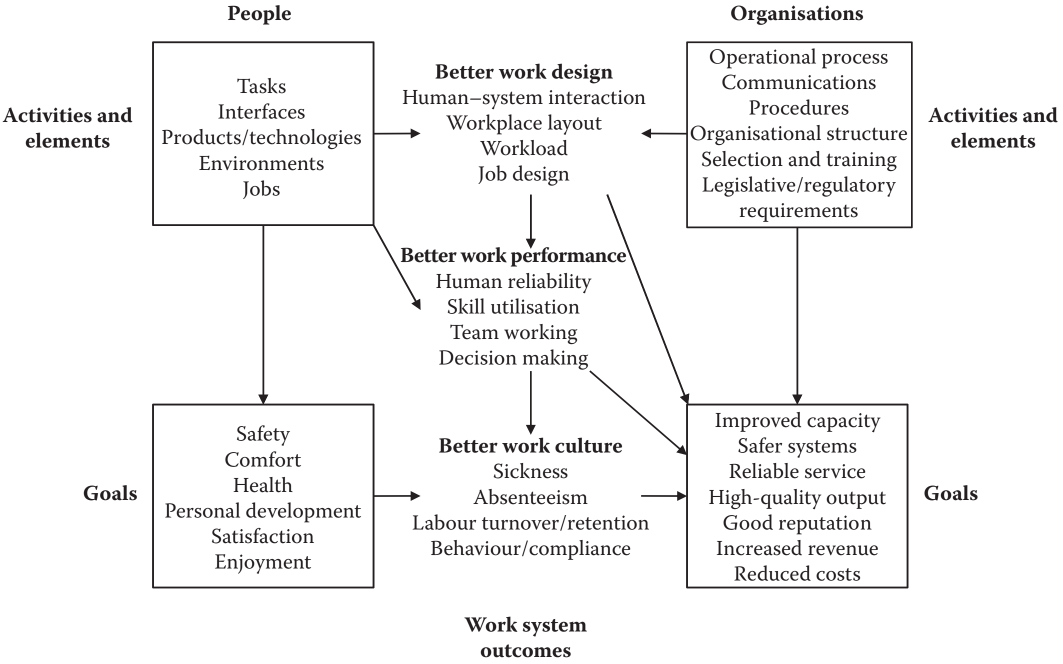

FIGURE 1.1 Aims of ergonomics/human factors. E/HF can be seen in the context of its objectives, which are the well-being of people and of organisations. These might be seen as twin aims, which are neither independent nor mutually exclusive and which have direct and systemic connections.

- *Cognitive ergonomics*: information processing, sensing, perception, decision making, problem solving, reaction, mental workload, fatigue, stress, interface design, reliability, communication, fault diagnosis
- *Organisational (social) ergonomics*: attitudes, motivation, satisfaction, job and team design, hours and patterns of work, pacing, implementation of change
- *Systems ergonomics*: taking a holistic approach to design and evaluation that integrates the physical, cognitive and organisational/social

AIMS AND FRAMEWORK FOR ERGONOMICS

In most definitions of E/HF, we will find a list of objectives or criteria that drive its application, for instance jobs, systems or products that are comfortable, safe, effective and satisfying. Aims are often divided into those of gains for the individual (employee or user), and those for the organisation (employer or producer). *These aims, however, are neither mutually exclusive nor independent.* It is not a case of having to implement either a more comfortable workstation or a more productive one for example, nor are the ways of achieving the former necessarily very different from those for the latter, and many (e.g. Becker and Gerhart, 1996) have demonstrated the positive impact of employee well-being on organisational performance. In fact, it is difficult to think of cases where work or equipment designed to meet the needs of an employee or user would detract from performance effectiveness; in general, user fit will enhance performance. To contribute such 'win-win' outcomes, and to sell them before and to prove them afterwards to managers and engineers, is one of the main tasks facing practising ergonomists. Generally, if we concentrate upon objectives of ergonomics, then we require methods that go some way to helping us 'prove' that we have met certain aims or have achieved a certain level of improvement (see Kim et al., 2001 for an example that demonstrates this approach in an assembly line context). Cost–benefit analyses (see Chapter 35), an appreciation of the relationship between ergonomics and quality (see Chapter 36) and the setting of usability metrics (see Chapter 13), are of importance here.

E/HF remains of importance in the face of the changing nature of work. As mentioned earlier, the home/work boundary is becoming blurred. Increasingly, we use our home technologies (e.g. personal mobile phones) to support our work, and our work technologies to support our home activities (e.g. accessing a website at work (during a lunch hour of course!) to organise our online supermarket shopping delivery). It would be foolish to ignore these changes. Our expertise in using work systems can be enhanced by our home practice – there is an argument for designing the interface for a system used in a workplace to control a tele-operated robot to follow interaction paradigms established by a personal games system, for example. In addition, as has been discussed extensively by Erik Hollnagel and David Woods (2005), it is no longer appropriate to consider the person and the technology/system as two separate elements, with inputs and outputs between them; instead, it can be more fruitful to consider how people work *in partnership* with technology as a 'joint cognitive system' (this approach also has synergies with concepts of distributed cognition [Hutchins, 1995] and concepts such as computational offloading [Scaife and Rogers, 1996]).

Figure 1.1 illustrates the twin aims of E/HF in the context of work systems. We see that there is a direct connection between design and development criteria for people and organisations, and an indirect or systemic one also. Demonstrating the value of considering both people *and* organisations as contributors to better work systems strengthens the argument for the consideration of E/HF.

A traditional view is that, like the epidemiological model of 'host–agent–environment' in disease control or accident prevention, E/HF is concerned with interactions between people and the things they use and the environments in which they use them. Our concern is with the contextually based (or situated) interfaces between people and the processes with which they interact, whether a toothbrush, training manual, motor car, power plant control room or school, for instance. Within this, people's interaction with other people is at the core of an E/HF framework, increasingly so as systems of interest move away from 'one person – one interface' to distributed networks (Wilson, 2000).

Traditionally, we have represented the context of E/HF using the 'onion model' (Grey et al., 1987), now a classic representation of the 'person at the centre' and used in many introductory classes to set the scene of E/HF. Figure 1.2 shows the 'onion revisited' – a revised look at the interactions between factors relevant to the application of E/HF in work design.

The centre shows the three overlapping elements representing 'individual interaction' – a person, their technologies and artefacts that they are using. If we consider an example of the Air Traffic Controller (ATCO), they would work with, at the very least, communications technologies and a representation of aircraft location (usually via a radar system of some sort), will use paper or virtual flight strips to record communications and instructions and will be in a room/at a desk where artefacts include clocks, diagrams of airspace or schedule information. The ATCO will complete discrete tasks (e.g. approve a flight level change request) and have goals (e.g. maintain safe separation of aircraft, ensure planes depart according to schedule) which influence their completion of those tasks and their interactions with technologies and artefacts.

Task completion takes place in a personal physical and virtual workplace – we need to move 'beyond the desk' when we consider a person's workplace. Increasingly we store our personal documents and information in the 'cloud' – a distributed network of servers that enables us to access stored data resources remotely, regardless of our physical location. Although this may be in a physically different place to an individual, it is still their 'personal' space in this virtual system. We should not of course overlook the continued importance of local workspace design (e.g. see Chapter 14 on control facilities design for further consideration of this matter).

Similarly, the wider work environment takes on a physical and virtual form. Environmental considerations remain a significant contributor to workplace performance and satisfaction, and are increasingly automated and 'intelligent' (and we have all sat in freezing cold 'intelligent' environments, wondering how we can break into the air conditioning system), presenting challenges

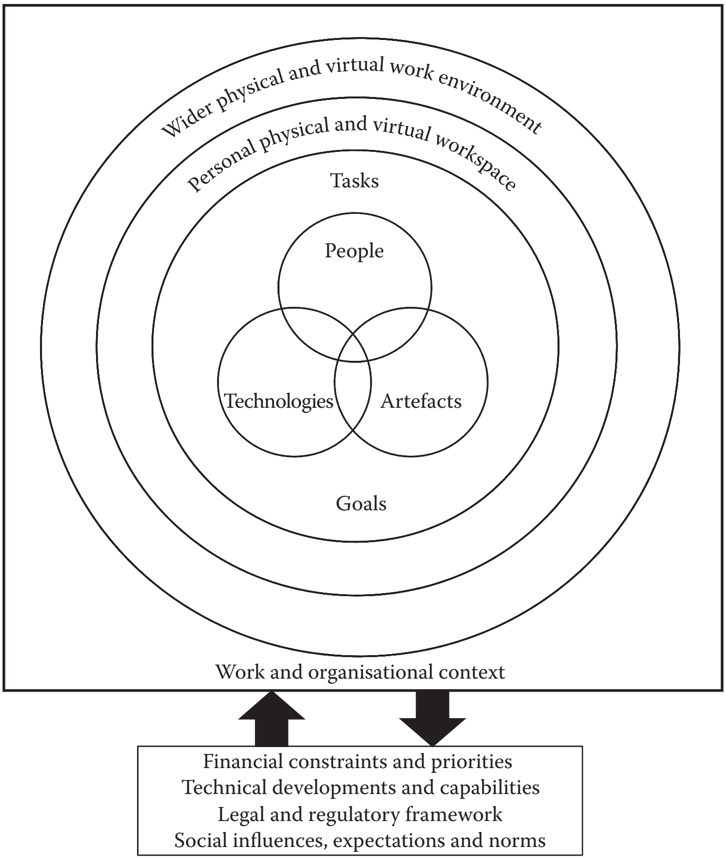

FIGURE 1.2 The 'onion model' revisited – Interactions of factors relevant to the application of E/HF in work design. At the centre is the consideration of people, technologies and artefacts. These are placed in the context of their tasks and goals, physical and virtual workspaces and organisational context, and are influenced by financial, technical, legal and social considerations.

regarding individual workspace control and ownership. Increasingly we work with our virtual systems, whether it be a holiday request online system, video conferencing technologies or a shared file storage repository. The role of organisational goals and culture (as highlighted earlier in Figure 1.1) also play a part here. All individuals within the organisation, and the policies and practices of the organisation itself, are also influenced by financial, technical, legal and social factors.

Argument for and against E/HF

It was implied earlier that it is not always easy to sell the approach of ergonomics and human factors and the aims to jointly improve work and task performance as well as worker and user well-being. We must be prepared for the criticisms or arguments against. Helander (1999) and Pheasant (1986, p. 8) identified some of these, and the authors have directly come across others in their own work. Arguments from engineers, designers and managers can include:

1. 'I can use it' – I have done my user testing on myself, what is the problem? Some software developers will try their own product out and assume that it can be used by all; whilst such expert walkthrough is a very powerful early assessment technique (see Chapter 10), it is almost always not sufficient. It is particularly difficult to 'unlearn' expertise in interacting with systems, and what may seem logical to a designer or engineer who has not

only designed the 'user-facing' elements of a product or system but also been involved in designing the underlying architecture will not automatically be logical to an end-user.

2. 'We designed it for the average' – I know we should understand the user, but we were not sure how to do this or even who this was, so we plumped for something in the middle. There are two problems here. First, there is no such thing as an average person; one person might be in a middle range of, say, stature, but generally would not be for leg length, head circumference, shoulder breadth, etc. as well; still less will they be in the middle range of intelligence, strength, vision, hearing, reaction time and many other characteristics as well. Second, where a single characteristic of people is the critical one, and we can identify an average value, it is not usually appropriate to design to it: a doorway for people only of average height, a display that can be read only by people of average eyesight – these are not sensible design choices. Equally, we will not always design for the absolute extremes either – we would not want the cost and inconvenience of houses with door frames 225 cm high nor all instructions in nursery school level language. The 'user of choice' is a critical consideration in any design activity and one that will be influenced by many factors, including the nature and business context of the task/product; this must be balanced with accessibility expectations and legislative requirements, ensuring that accommodations of those with additional requirements are either incorporated or feasible with small adjustments.
3. 'People are too variable to account for' – I understand the ideal of accounting for variability across people for a whole set of critical characteristics, but there are too many of these and too much variability. In fact, very early on in any ergonomics contribution to design, we should establish exactly *which* are the critical characteristics that must be accounted for. Moreover, unless there is good reason not to, we will not design for 100% of people but to best suit 95% or 97.5% of the population, whether through establishing the critical end of the range or through adjustability or by fitting to the average in certain circumstances only (see Chapter 27).
4. 'People can adapt so why be worried?' – Why should we bother with all the hassle of ergonomics, people will cope. In this, human beings have been their own worst enemy. We are very adaptable – to poor physical design of workstations, to poor information and instructions, to poor design of jobs. Everywhere, workers and consumers are managing with, or working around, the poor designs of equipment, environments and jobs that they have been given, to keep the company working or to survive in daily life (in fact, when encountering a workplace for the first time, an ergonomist will often glance around for where workers have made adjustments – a label here or an improvised footrest there can be a good clue to where design problems are occurring within the work environment). But such adaptability can come at a long-term cost, to health, satisfaction, spare capacity and reliability if unexpected events or emergencies arise.
5. 'It is the workers that make mistakes' – Let us try to automate everything we can and blame the user if we cannot. Yes, human beings are fallible, they get distracted, tired and forget things. But often those errors are made almost inevitable at some time by the planning, organisation, equipment, job and procedures design and training programmes that have gone before (see Reason's [1990 and 1997] idea of latent failures and organisation pathogens – see also Chapter 30). Also, to borrow from the legal concept of strict liability, even misuse by individuals is often likely and can often be foreseen by a prudent design team, and should be part of the design decisions that are made. The concept of resilience engineering directly addresses this matter by examining how we can design multiple elements within a work system to anticipate, mitigate and prevent the occurrence and impact of human error. Moreover, it is often the attempts by engineers to design people out of the loop as far as possible that leaves badly designed jobs and circumstances that actually lead to unreliability, by not making the most of the capabilities of people, as Bainbridge (1987) notes in her ironies of automation.

FIGURE 1.3 The presence of three separate formal and informal sets of instructions (and a threat of removal of entitlement of door use!) suggests that something may be amiss with the design of this set of door controls.

6. ‘OK, it is a bad design, but we train them’ – Let us increase the adaptability of people to our poor design. Yes, instructions, cues and training are important, but after well-thought-out human-centred design, not instead of it. It is also normally the case that you cannot cure a hardware problem with software changes. Again, sometimes an indication of poor design can be found by the presence of additional instructions (see Figure 1.3) but, where possible, good design should take advantage of pre-existing user skills and experience, and provide intuitive cues and affordances, to encourage correct behaviour (see Norman, 1988 for an extensive discussion of the role of design in encouraging appropriate user interaction).
7. ‘People should just do as they are told’. – We have set out perfectly clear instructions or processes; why do not people follow them? Again, we can find a clue here – maybe those processes do not fit into users’ expectations of normal behaviour, (if an individual in a maintenance crew is expected to pass all communication via the foreman, but the foreman is 100 m away, and the person who they need to help is right next to them, why should they make the request via the foreman?) or may seem cumbersome or time consuming (e.g. there is much discussion about adherence to and engagement with the WHO surgical checklists that require a pre-operation discussion amongst all the surgical team (Sparks et al., 2013). A better approach is to think *why* are people not adhering to instructions or rules, and how can we ensure safer, more efficient behaviour through a holistic design approach?
8. ‘Human factors costs too much’. This is an old argument, and efforts are frequently made to apply cost–benefit reasoning to ergonomics (see Chapter 35). Yes, a contribution to take account of human factors will increase the cost of any design project, implementation scheme or operational overhead – in staff, equipment, study costs and any extra time needed in the system life cycle. But, as we saw from the view on aims of ergonomics earlier, these costs will be outweighed by the saving made on the costs of not getting it right and also the financial and less tangible benefits which will accrue (see Table 1.2).

TABLE 1.2
Potential Costs and Benefits of Ergonomics/Human Factors

Potential Costs of E/HF	Potential Costs if *No/Poor* E/HF	Potential Benefits from E/HF
Personnel resources	Accidents, incidents and disruption	Improved workforce/customer health and safety
Administrative and contracting resources	Design retrofit, alteration	Improved workforce performance
Technical and equipment resources	Product withdrawal	Improved workforce satisfaction
Cost of analysis, design and evaluation process	Loss and dissatisfaction staff	Improved client satisfaction and perceived value
Capital costs of redesign features	Criminal prosecutions	More effective systems
Increased system development, design and test time	Compensation claims	Legal compliance improved
Disruption to normal activities	Insurance premiums increase	Better use of resources and capacity
	Absenteeism, labour turnover, cover	Better labour relations
	Recruitment and retraining	Positive company image
	Lower product sales	Repeat sales
	Poor company reputation	Market differentiation

9. 'Ergonomics is all common sense'. The answer here is yes and no. Some is common sense – but which often seems in short supply without a structured framework within which to identify what needs to be known, apply it and deal with trade-offs (see human factors integration plans – Chapter 9). On the other hand, much of ergonomics and human factors is clearly not common sense; some research findings or guidance may even be counter-intuitive (e.g. the negative impact of underload on attention and task resumability). It is use of the approaches and application of the methods described in this book that will enable us to address the unusual and opaque issues and to rationally bring together into a coherent process and plan the more obvious and straightforward issues.

CONTEXT FOR APPLICATION OF E/HF METHODS

Since E/HF must have relevance for real settings, the understanding of findings and their application in practice requires a good grasp of context. This same context will also affect society's view of ergonomics and its value and place in the modern world. It also impacts strongly on choice and application of methods. John Wilson identified the following six features that are particularly important to consider in systems E/HF in a paper written in early 2013 (published as Wilson, 2014) that provide an overview of the context for application of E/HF methods:

1. We need to adopt a *systems focus*, making sure that we move beyond the 'person-technology' dyad and considering human interaction within the context of man-made and naturally present systems. One way of making this more manageable is to consider a 'system of systems' where we consider an overall system (the example of a hospital is given in Wilson, 2014) encompassing a series of sub-systems, such as beds, ambulances or equipment sets. When selecting appropriate methods, we need to clearly understand the breadth of system that we are considering, and the range of elements within that system that will interact with the designs, workplaces or products that we are specifically examining.
2. We need to remember that behaviour and performance happens in a *context* (see Moray, 1994; Waterson, 2009). This has implications for our selection of methods; if we wish to examine the effect of context, we must capture context and thus, wherever possible,

examine behaviour *in situ*. However, there is still a role for controlled studies in a 'laboratory' type context if we wish to capture detailed measures of performance under circumstances which we wish to manipulate (e.g. changing the level of demand in a task simulated to influence perceived workload). What is important is the extent to which we generalise or transfer our results, both to underlying theories, and to real-world contexts.*

3. A system comprises a number of *interacting parts*. These could be 'human, technical, information, social, political, economic and organisational components' (Wilson, 2014). Therefore, our methods need to move beyond examination of linear patterns between discrete elements, to capture complex relationships between people, technologies, artefacts and context. Of particular importance here is the appreciation of not only the range of users a particular product or system might have, but the needs of different stakeholders within that system – in a novel data recording system in a manufacturing context, for example, the machine operator may require data in one form; the line manager may require different data presented and captured over a longer time frame and a production manager may need data from different lines, from different products and machines, in a comparable form.
4. Systems should be considered in a *holistic* manner. This is above and beyond simply considering the context, but acknowledging that people do not work in a social vacuum. Whilst our methods may align to the categories of E/HF outlined earlier, such as physical, cognitive or organisational, we must appreciate that the impact of 'people on the performance of the human–machine system… and impacts of system design on the well-being of all stakeholders' (Wilson, 2014) is captured in our methods as needed.
5. Once systems are implemented in a working context, different *emergent properties* will develop. These can be a positive development – for example text messaging was not originally included as a 'killer app' for mobile phones in the early twenty-first century (145 billion texts were sent in 2013), (*The Guardian*, 14 January 2014) but, once restrictions on sending texts between different mobile networks were lifted, rapidly it became the most prevalent interaction of users with mobile phones. Conversely, unanticipated behaviour of users can cause people within systems to behave in a different manner to that which had been anticipated by the system designers (as seen in many failed implementations of national IT systems). Therefore, user requirements methods should try to anticipate as much of this emergent behaviour as possible, and capture/manage this behaviour when it does occur.
6. Finally, the place that E/HF has within an organisation has a large influence on the success of *embedding* ergonomics practice, and ultimately influencing change according to E/HF needs. Different organisations place their E/HF expertise (if they have it) in different departments/sections, and some prefer to only use external collaborators to provide an 'ergonomics view'. This may influence the way in which we design our methods; methods may be more accepted if appropriate and familiar terminology is used, for example. Our own extensive experience in the rail industry has allowed us to become familiar not only with the necessary acronyms and work procedures, but also to appreciate the appropriate levels of complexities and acceptability of our methods in different situations; for example if we wish to conduct some in-depth observation of a night maintenance operation, or a complex signalling task, it is important that our presence is accepted and not seen as an inconvenience, and that we understand enough about the work environment that we

---

* John and I had many lively discussions about the role of laboratory studies in E/HF. In his 2014 paper, John states 'laboratory research has its place, but not a primary one'. I would not state this so strongly, and in a footnote in his paper, he notes 'I do recognise that for some this may be a little extreme, and must point out to the reader that this may not be a universal view in E/HF'! (SS).

can interpret it without needing to interrupt the operators and ask for any explanations of terms. We joke that once we have been told where the snack drawer is in a workplace, we know that our presence is accepted; achieving this non-intrusive access, whilst maintaining an external analysis perspective is extremely valuable if we are to capture E/HF factors in a systems context.

ROLE OF E/HF METHODS

Although we looked at definitions of ergonomics and human factors earlier in this chapter, it is more important that E/HF should be seen as an approach (or as a philosophy) of taking account of people in the way we design and organise, in other words, as designing for people. In this view, E/HF itself is primarily a process, to an extent a meta-method, which makes the clear understanding and correct utilisation of individual methods and techniques even more important. This supports the need for this book.

Back in 1986, the NRC Committee on Human Factors reported the development of applied methods to be one of the major research needs for human factors (National Research Council, 1986), and nothing has occurred in the years since to change this. Ergonomics is a science, craft, discipline, practice and a technology and thus has need of techniques for both data collection (basic or functional data), analysis and application. The debts we owe to other disciplines are obvious. However, as we have gained experience and confidence and as our armoury of knowledge and methodology have grown, so the debt to other disciplines has been, and is being, repaid. Methods developed or adapted within E/HF are employed by psychologists, engineers, computer scientists, management scientists or health care professionals in turn, just as we are constantly enlarging our own human performance database through results of human–machine systems evaluation.

Bearing in mind the applied nature of ergonomics, we have both general methods and approaches (e.g. observation, subjective assessment, data logging and experiments) that can be applied at many stages of the design process, and then specific tools that are used when we have specific types of input to the design process that we wish to make. These can be classified into five groups: data about people, systems analysis and development, evaluation of system performance, assessment of effects on people and the organisation of E/HF management programmes.

METHODS FOR THE COLLECTION OF DATA ABOUT PEOPLE AND THEIR NEEDS

Our first methodological need is for data about people and this can cover all characteristics: physical size and strength, physiological capability, sensory characteristics, mental capacities and psychological responses. In addition to collection of data about these human characteristics, it is also important that we establish users' motivations and desires – this is particularly valuable when we are trying to balance priorities of, for example usability and aesthetics. Just as important as the collection and reporting of data is the generation of design and evaluation criteria from these. For example given data on the population range of arm reaches in different directions, what advice can be given on placement of frequently used rotary controls? Or again, given data on working memory limitations, can these be adapted to form design guidance on numbers of different codes that should be used in a coding system? Methods used to produce data about people comprise much of the scientific base of E/HF.

METHODS USED IN SYSTEMS ANALYSIS AND DESIGN

The second input of E/HF is its contribution to the analysis and design process. This relates specifically to methods that assist in the analysis and design stages of development of equipment, workplaces, technologies, jobs, work processes or buildings. In essence, we need methods to *analyse* current or proposed systems (analysis strictly meaning to resolve the system into its constituent elements and critically examine these) and then to *synthesise* data (i.e. build up a coherent whole by putting elements back together) into ergonomically sound concepts, prototypes and final designs (see Chapter 9). Specifications produced out of this process must be capable of transference into design requirements, criteria and data, and also have reasoned justifications, in order that ergonomists can work sensibly with engineers and designers.

METHODS TO EVALUATE HUMAN–MACHINE SYSTEM PERFORMANCE

In part, analysis at the start of development may involve evaluations of an existing system's performance. Certainly, we must evaluate system performance during and at the end of development to support cost–benefit analysis and enable benchmarking of system performance. Many measures can be defined for this, and one challenge today is the search for measures of system performance that extend beyond the typical direct (but often sterile) measures of times and errors other than. Manufacturing system performance, for instance, may be assessed by means of production output rates and product quality levels, but we could also use machine utilisation rates, minimisation of finished stocks or work-in-progress, raw material wastage, speed of response to changed schedules, accident rates, sickness or other absence, and job attitudes and job satisfaction measures. Similarly, although a computer interface can be assessed in terms of time taken and errors made in performing a sequence of tasks, more interesting measures might be 'extent of system explored', 'willingness to change direction', 'quality of finished work', etc. (Johnson and Wilson, 1988). New technologies, such as detailed data recording and sensing equipment, along with the increased capacities and reduced intrusion of video tools, make detailed longitudinal data collection of complex interactions much more feasible than in the past, and can support these more detailed and valid approaches to analysis of system performance.

Any evaluation of subsequent system performance where there is a human factors input into system development is also, in an interesting closing of the loop, an evaluation of just how well E/HF was applied to the design.

METHODS TO ASSESS DEMANDS AND EFFECTS ON PEOPLE

As we have seen, E/HF has twin aims in its contribution to design and development: improvements for the job-holder or user *and* improvements for the producer, employer or organisation. As a result, any system assessment should be carried out in terms of the demands made on people and of effects on their well-being, as well as accounting for system performance. There is an argument that the demands made on people should in fact be viewed as an implicit part of system performance and as such, their assessment should not be seen as separate; however, on balance, it is more persuasive that such a distinction will emphasise the twin, yet interdependent, aims of E/HF. Many methods can be applied to assess the effects that different environments, jobs or equipment have on people. Such impacts might be medical, physical or psychological in nature, and methods will vary from recording of measureable phenomena (e.g. heart rate) to observation of people's affective states (e.g. boredom). Again, technological developments have dramatically changed the extent to which such recording methods are feasible to implement in a workplace or complex simulation – for example it is now possible to capture a person's expression and automatically infer their emotional state (e.g. el Kaliouby and Robinson, 2005). In almost all circumstances however, the data collected are not useful by themselves but must be interpreted and any effects for the overall system design inferred by looking at the context of the effect and the interacting factors that may have contributed to that demand or effect; this is again a large part of the ergonomist's input. Moreover, if assessment methodology is developed appropriately, then data obtained can be generalised to become part of our first input, the basic data on people.

METHODS TO DEVELOP E/HF MANAGEMENT PROGRAMMES

Our fifth input concerns the management of E/HF programmes. Methods are required here for two situations, although there is not a clear distinction between them. Firstly, there are ergonomists who are working within companies, and they are often doing this in very small groups or even on their own. Secondly, we have what is termed 'devolving' (Wilson, 1994) or 'giving away' (Corlett, 1991) ergonomics expertise. It is unrealistic to expect all enterprises to employ only ergonomists (either as employees or outside consultants) to handle their ergonomics; it is inappropriate also in many circumstances. In many areas – for instance health and safety, job redesign, workplace layout – the ergonomics profession must provide methods which allow the development of appropriate strategies and which support the management of programmes which can be run as a part of normal company activities. Nobody is advocating that untrained staff handle all human factors in a company. However, design engineering, systems engineering, interaction designers, health and safety advisors, line management and production workers can all make considerable contributions; the methods and support that ergonomists give to them must include enabling them to recognise when specialists must be brought in. This fifth input is certainly the 'messiest' area of E/HF methodology, and it embraces aspects of all other inputs as well as participatory ergonomics, systems implementation and so on (see Chapters 28, 32 and 34). Nonetheless, it is an increasingly important area as E/HF moves more and more into real companies with real problems as well as maintaining its position within research laboratories and universities.

CLASSIFICATION OF METHODS

Classification of ergonomics methods in terms of the parts or stages of any models of E/HF is difficult. Meister (1985) attempts a gross distinction and divides his behavioural methods into the analytic techniques employed during the development of systems and the measurement methods employed to evaluate functioning systems. He does, though, recognise overlaps, particularly that many measurement methods are also used during system development. Similarly, at a first level, Sharples and Megaw, in Chapter 18 of this book, distinguish analytical methods or measures – broadly those based on theory, modelling and prediction – from empirical ones – those based on observation or experiments with an actual situation.

The following tables are adapted from one which was published in the earlier editions of this book, and provides an overview of the different types of methods that are needed to address all aspects of a work or interactive system. Six core types of methods are identified, represented within each of the sub-tables. The six methods types are:

a. *General methods*: Representing approaches of data collection and/or analysis that can be used with a range of different design or evaluation goals, across a work system, and normally within any of the method categories (b to f).
b. *Collection of information about people, equipment and environment*: Typically 'baseline' and descriptive measures of either an individual or environment's state or characteristics. For an individual, this could be an inherent state, such as their anthropometry or age, or one that is transient and affected by external factors, such as their attitude; similarly for an environment, this could be a core characteristic, such as the size of a room, or its current thermal state.
c. *Analysis and design*: Tools that move beyond description to provide links between the data collected and underlying influences, actively supporting the design process.
d. *Evaluation of user and system performance*: This will often use methods of type (b) as a baseline to understand both the performance outputs of a system, and the factors that may influence those outputs.

e. *Evaluation of demands on and effects experienced by people*: Capturing data from people about their physiological and psychological responses to systems, experiences and tasks.
f. *Management and implementation of ergonomics*: Toolkits to support the integration of E/HF, normally focussing on early consideration of E/HF requirements in work/system design.

The tables are not intended to be exhaustive, and categories are not mutually exclusive, but aim to provide an indication of the types of methods that an ergonomist should consider using, and their outputs. One notable difference between the contents of these tables compared to the similar table in third edition of this book is the emergence of complex modelling and technology-enabled data collection approaches, which have become both technically feasible, cost-effective and of reduced intrusion compared with a decade earlier (Table 1.3).

DEBATES AND DISCUSSIONS AROUND METHODOLOGICAL APPROACHES

As far as methods are concerned, according to an old expression, the proof of the pudding is in the eating. A method which to one researcher or practitioner is an invaluable aid to all their work may to another be vague or insubstantial in concept, difficult to use and variable in its outcomes; of course, individual E/HF practitioners and teams develop expertise in and familiarity with particular tools. More than this, the validity, reliability and sensitivity of methods will be application specific. We may need examination of utility and generalisability also in order to select or prioritise between methods. However, a number of traditional 'debates' persist within the discipline, and we consider some of them here.

Once all possible measures are considered, we begin to see some of the problems of classification and selection. Kantowitz (1992) lists 46 possible indicators of nuclear power plant safety, split into seven categories (e.g. operations, quality programmes, management/administration) and reports that 'no single indicator was by itself an adequate measure… [nor were any] optimal for predicting plant safety' (p. 391).

It is not surprising therefore that we will often look to use more than one, and often several, methods and measures in any one study. Indeed, as E/HF has a bounty of techniques and methods to draw upon, the issue should not be one of A versus B but of selecting a subset of our broad range of available methods that will be appropriate to the issue at hand.

This is particularly so when we are carrying out evaluations in the field. Technically, if this is done formally, it is known as triangulation (see Denzin, 1970 or Webb et al., 1972) (and, despite the name, requires only two or more methods or approaches to be applied independently). Guion et al. (2011) identify five types of triangulation: Data (using different source of information); Investigator (using different people in the analysis process); Theory (using multiple perspective to interpret data); Environmental (Using different location settings or contexts to capture data); and the most commonly considered type of triangulation, Methodological Triangulation. Methodological Triangulation refers to the independent use of multiple distinct qualitative or quantitative methods to address the same research question. The application of triangulation will improve the effectiveness of a study or our confidence in any findings; weaknesses in one method can be balanced by strengths in another. To take one simple example, only by questioning and observing operators in complex systems *and* also recording their concurrent or retrospective verbal reports for subsequent protocol analysis can we begin to understand something about their decision making activities. A multiple-methods study may utilise a mixture of qualitative and quantitative techniques, in field and laboratory settings. Such an approach in the field is often known as contextual inquiry.

Only by use of several of these methods may a full evaluation be possible in any one situation, and thereby effective suggestions be made for redesigning job content, tasks, workstations and environments. However, a word of caution is in order. It is all too easy to fall into the trap, once an investigation has started, of measuring everything possible 'just in case'! This can lead to results

TABLE 1.3
Classification of Methods, Techniques and Measures Used with Ergonomics/Human Factors

Method		
Group	**Example Techniques**	**Typical Measure/Outcome**
a. General methods		
Direct observation	Human recording: observation checklists, expert rating, photographs, diagrams	Event frequency; sequence, strategy analysis; performance time, error, accuracy; participant behaviour
	Technology-led recording: videos, audio, photos, data and event logging, location tracking	Interaction patterns; link/sequence analysis; emotions
	Performance measures	Time, error; strategy descriptions
Indirect observation	Archival analysis, artefact analysis, data mining	Event counts; covariation analysis; critical incident analysis; temporal sequence analysis; absenteeism, sickness records
Participant reports	Verbal protocol, surveys, questionnaires, rating, ranking, scaling, diaries, checklists	Attitudes; perceived effort; preferences
Practitioner–participant communications	Group discussions, interviews, participant-led observations	Themes; thoughts and preferences; ideas
Work system instrumentation	Screen capture, eye tracking, voice recording, data logs	Scenario description; eye movement analysis
Experiments	Inferential and correlational designs, multiple and individual independent variables, analysis of covariables	Causative analysis; likelihood of relationships compared to chance; statistical trends and patterns
Literature and data interpretation	Description, statistical analysis	Patterns and prevalence; links, networks
Standards and recommendations	Standard adherence check (e.g. BSI [British Standards Institute], ISO [International Organization for Standardization])	Official or unofficial quality statement
Prediction and modelling	Analytical hierarchical process, human reliability assessment, simulation, cognitive work analysis	Predicted behaviours and probabilities
Multiple methods	Triangulation, mixed methods	Validity and reliability indications
b. Collection of Information about people, equipment and environments		
Physical measurement	Anthropometry, 3D scanning, ultrasound, fitting trials, photographs	Dimensions, percentiles, population descriptions
	Biomechanics, goniometry, stadiometers	Spinal shrinkage, angles, strength
Physiological measurement	Sway measurement, posture and gait, heart rate variability, muscle contraction, brain activity (e.g. electroencephalography, functional near-infrared spectroscopy, magnetic resonance)	Brain activity, posture, gait, movements
Environmental measurement	Hygrometer, thermometer, sensors, light/sound meter, accelerometer	Humidity, temperature, vibration, light/sound levels
Perceptual/cognitive assessment	Visual acuity, colour blindness, stereopsis.	Perceptual ability
	Psychophysics (method of limits, constant stimuli, adjustment)	Thresholds and levels of perception, sensitivity
	Mental or cognitive tests (e.g. intelligence, spatial ability), response times, skill tests	Predictions of performance, individual characteristics

(*Continued*)

TABLE 1.3 (*Continued*)
Classification of Methods, Techniques and Measures Used with Ergonomics/Human Factors

Method		
Group	**Example Techniques**	**Typical Measure/Outcome**
Social and organisational measures	Cultural probes, photo diaries, historical analysis, network analysis, social media analysis, sentiment analysis	Communication patterns, habits, interactions, networks, feelings, team behaviours
Knowledge	Written, audio, video records, interviews, protocol analysis, conceptual mapping, goal decomposition	Rules, practices, reasoning, explanations
Models	Artificial intelligence models, biomechanical models, logical mathematical models	Processes, sequences, limits and capabilities
Task analysis	Hierarchical, tabular, link analysis	Task completion methods, action sequences
c. Methods to support analysis and design		
Task analysis	Hierarchical, tabular task analysis, requirements analysis, cognitive task analysis, task action grammars, operation sequence diagrams, flow charts	Requirements for people and systems, data to support prediction of consequence of task change
Expert analysis	Guidelines, heuristics, standards, cognitive walkthrough, Delphi technique, likelihood matrix	Weightings and priorities, qualitative reports, critical issues, risk analysis
Introspection/protocol analysis	Repertory grid, concurrent/retrospective verbal protocol, diaries, shadowing	Behaviour transitions, rules, strategies, expertise, knowledge models
User models	Mental models, task action grammar, GOMS (Goals, Operations, Methods, Selection)	Strategies, beliefs, priorities, goals
Statistical analysis	Signal detection theory, information theory, reliability assessment	Performance measures, likelihood of outcomes
Models	Rich pictures, soft systems methodology	Performance predictions
Simulation	Mathematical and computer models, virtual reality, computer-aided design, personas	System capacity and capability
Method study	Link analysis, layout analysis, process, flow, time charts; micromotion	Movements, times, actions, frequencies
Work measurement	Time study, activity analysis, synthetic analysis	Times, standards, task sequence, simultaneity, frequency, demand
System-level representation/ modelling	Influence diagrams, STAMP (System-Theoretic Accident Model and Processes), soft systems methodology, Accimaps, cognitive work analysis	System interactions, boundaries and emergent behaviours
Prototyping	Wireframing, paper prototypes, virtual prototypes	Design representations, user/stakeholder views on designs
Creativity techniques	Scenarios/use cases, stop/start animations	Novel design ideas, representations of user perspectives
Participative methods	Co-design, drama in design, role playing, design decision groups, co-operative evaluation, living labs	User-informed design suggestions, user requirements

(*Continued*)

TABLE 1.3 (*Continued*)
Classification of Methods, Techniques and Measures Used with Ergonomics/Human Factors

Method		
Group	**Example Techniques**	**Typical Measure/Outcome**
d. Evaluation of user and system performance		
Work systems analysis	Cognitive work analysis, checklists, walkthrough, expert assessment, critical decision method	Time, reaction time, strategies, processes
Usability evaluation	Observation, protocol analysis, expert analysis, interaction analysis, screen recording	Accuracy, errors, opinions, attitudes, satisfaction, workload
Work system instrumentation	Environmental measures, frequency of use (e.g. footfall measures), performance logs, eye tracking	Measurement against comparison, baselines, norms; fit to requirements; acceptability
Participant reports	Scaling, rating, questionnaires	Comfort, annoyance, acceptance, pleasure
Performance measures	Work rate, waste, errors, communications analysis, behavioural markers, secondary task measures	Measurement against comparison, baselines, norms
Modelling and simulation	Critical path analysis, keystroke level model	Strategies, task sequences
Self-recording	Gripe button, diaries, event recorder, crowd-sourcing	Problems, incidents, potential improvements
Text analysis	Readability formulae, judgements, scan/read tests	Normative scores, age ratings
Human reliability analysis	Error modelling, error prediction, error and incident classification	Error analysis, causal analysis, error probabilities
Accident and safety reporting and analysis	Event analysis, incident analysis, Accimaps	Incidence, severity, aetiology
e. Evaluation of demands on and effects experienced by people		
Physical response	Subjective assessment, perceived exertion, sickness ratings, physical changes (e.g. stadiometer), psychomotor performance	Ratings, performance change, physical measures
Psychological response	Subjective response, performance measures, task strategies	Experienced load, perceived demand
Posture and activity analysis	Automatic posture analysis (e.g. CODA, computer vision), Postural assessment tools (e.g. RULA [Rapid Upper Limb Assessment], REBA [Rapid Entire Body Assessment]), force platform	Posture descriptions, rating of risk of injury rating, body part angles and positions
Physiological measures	Heart rate variability, galvanic skin response, oxygen uptake, brain activity	Objective comparative data
External demands	Time pressures, frequency of demands	Temporal analysis of activity patterns
Fatigue measures	Critical flicker frequency, task performance, reported fatigue, blink rate	Visual fatigue scores
Environmental response	Sweat rate, body temperature, hearing impairment, contrast sensitivity, sensation loss	Measurements to compare to norm/baseline
Stress measures	Galvanic skin response, subjective report techniques	Measures to compare to norm/baseline, subjective ratings
Job and work attitude measurement	Rating scales, interviews	Satisfaction, user needs, job characteristics

(*Continued*)

TABLE 1.3 (*Continued*)
Classification of Methods, Techniques and Measures Used with Ergonomics/Human Factors

Method		
Group	**Example Techniques**	**Typical Measure/Outcome**
f. Management and implementation of ergonomics		
Ergonomics management, organisation analysis, project management	Human factors integration, early human factors analysis, lifecycle management, MANPRINT (Manpower and Personnel Integration)	Early consideration of E/HF in system/product design lifecycle, inclusion of E/HF in design/operational process
Implementation	ISO tools, heuristics, embedded E/HF experts in organisation	Change to implementation approach
Cost–benefit analysis	Cost estimates, activity analysis	Financial analysis of impact of E/HF
Participative methods	Co-design, focus groups, elicitation techniques	User influence in system/product design
Ethics in ergonomics	Guidelines	Adherence to guidance

that are difficult to interpret or, worse still, an analysis phase that uses large amounts of study resources to no clear or useful purpose.

In selecting an appropriate approach and set of methods for any particular need, we will need to weigh up the relative merits of a quantitative or qualitative approach, so-called objective or subjective methods, and studies in the field or laboratory.

QUALITATIVE/QUANTITATIVE DEBATE FOR METHODOLOGY IN ERGONOMICS/HUMAN FACTORS

Qualitative approaches and methodology are being increasingly used in E/HF and sister disciplines such as organisational management, product design, psychology and engineering, and chapters throughout this book refer to its much increased acceptability and use. Although there has been considerable debate, even rancour, about its place in disciplines such as psychology, there is good reason to think that E/HF will benefit from the explicit inclusion of qualitative methods and tools in teaching, fundamental research and practice. This, in turn, would help to support the consideration of social, emotional and philosophical factors when conducting both theoretical and practical studies (Hignett and Wilson, 2004a,b). E/HF can be located on the cusp of the sciences and the humanities, and in the centre of the qualitative–quantitative continuum; it has much perhaps in common with anthropology, where the unit of analysis is interaction, in contrast to psychology where the unit of analysis is the individual.

Moray (1994, 2000) has recognised that sociologists, anthropologists and ethnographers offer a methodology more sensitive to the context of analysis of work, rather than methodologies which result in generalisations based on quantitative analysis, but that there is need to generalise findings of qualitative study to understand other different systems. There should be no claim made for a general superiority or even a preference for either of the qualitative or quantitative philosophies, approaches, methods and measures. Each should be used as needs and circumstances dictate, and we need to be clear about differences and similarities in scope, approach and outcomes – see Table 1.4 from Hignett and Wilson (2004a).

'OBJECTIVE' AND 'SUBJECTIVE' METHODS

The recent debates over the relative merits of quantitative and qualitative approaches interact with debate over so-called objective and subjective measurement (Wilson and Nichols, 2002). This has been an issue from the earliest days of the ergonomics and human factors discipline, with early

TABLE 1.4
Dimensions of Qualitative and Quantitative Methodologies

Qualitative Dimensions	Quantitative Dimensions
Words, understanding	Numbers, explanation
Purposive sampling, inductive reasoning	Statistical sampling, deductive reasoning
Social sciences, soft, subjective	Physical sciences, hard, objective
Practitioner as a human being to gather data, personal	Researcher, descriptive, impersonal
Inquiry from the inside	Inquiry from the outside
Data collection and analysis intertwined	Data collection before analysis
Creative, acknowledges of extraneous variables as contributing to the phenomenon	Predefined, operationalised concepts stated as hypotheses, empirical measurement and control of variables
Meanings of behaviours, broad and inclusive focus	Cause and effect relationship
Discovery, gaining knowledge, understanding actions	Theory/explanation testing and development

Source: Adapted from Hignett, S. and Wilson, J.R., *Theor. Issues Ergon. Sci.*, 5, 517, 2004b.

views or implications that if subjective measurements do not match objective measurements, then it is the former which are biased (e.g. Poulton, 1975).

However, what do we mean by objective measurement? Different forms of apparently objective measurement may show many elements of possible observer bias and of subjectivity in the actual selection of measures, criterion levels and methods. This was illustrated for us in work on mental model identification and representation; the form of mental models reported in the literature often seemed dependent upon the scientists' choices of experimental and measurement method, in turn based on their preconceptions about the form of mental models (Rutherford and Wilson, 1992).

Ergonomics and human factors deal with many concepts which are difficult to define explicitly and therefore have no clear agreement on method of measurement. We work with the complexity of human beings, individually and as social groups. In cases of mental workload, situation awareness, fatigue, mental models and presence for instance, it is quite possible that what we are measuring with objective (performance or physiological) methods may be different – if related – entities to what is measured with subjective methods (participant ratings, interviews etc.). This means that such concepts should be clearly, often operationally, defined in relation to the measurement method being used, rather than independently of it. Kanis (2014) contributes to this debate in his discussion of validity and reliability in E/HF tools, highlighting the perils of 'measurement by fiat' (the imposition of numerical values by arbitrary definition after presuming a relationship between obtainable observations and the property of the subject under investigation [Torgerson, 1958]). We would argue that subjective methods are extremely valuable firstly as tools to obtain information about a task or system quickly, with minimal intrusion, and without the need for specialist equipment, but secondly, and probably more importantly, to capture the user's perception about the system, as this perception may not only affect the user's attitude, but also their selected behaviours and interactions (see Chapter 18 by Sharples and Megaw). Yes, this opinion may be biased in some way (and of course, we should use techniques of scale design and good practice to ensure that this bias is minimised [see Chapter 4]), but this 'biased' opinion has a direct influence on the user's interactions and behaviours, and is important to capture.

Indeed, the word 'subjective' is a challenging one, and is often associated with an implied derogatory tone, assuming that in some way, a method that is subjective is inherently inaccurate. This is not surprising when we consider two alternative definitions of subjective from the Oxford English Dictionary: The first definition, which might be paraphrased as 'of the subject (i.e. participant)' is *Relating to a person who is subject; belonging to or characteristic of a political subject; submissive, obedient.*, yet a further definition is *Of, relating to, or proceeding from an individual's thoughts,* *views, etc.; derived from or expressing a person's individuality or idiosyncrasy; not impartial or literal; personal, individual. OR Existing in the mind only, without anything real to correspond to it; illusory, fanciful.* (Oxford English Dictionary, 2014).

The first definition simply uses the adjective to denote the source of information (the 'subject'), whereas the second places the notion of subjective data in opposition to that collected through 'objective' means. In fact, much of our skill in development of methods allows us to collect 'subjective' data in an 'objective' manner (see Zhang et al. (1996), Jian et al. (2000) and Pickup et al. (2005) for examples of such approaches). Annett (2002a) in a target paper, and several commentary papers in the same issue of ergonomics, covers very well the topic of objectivity and subjectivity in ergonomics measurement and assessment.

FIELD AND LABORATORY STUDY

Within the sporadic debates about the nature of ergonomics and human factors over the past decade, there has been much discussion of the merits of field and laboratory study, contrasting formal and informal methodological approaches and setting the advantages of control against those of veridicity (see Wilson, 2000).* Moray (1994) believes that ergonomics only makes sense in the full richness of the social settings in which people work, and Wilson (2014) returns to this theme. The influence of contextual factors – such as interactions between individuals, the formation and relationships of teams, individual motivations, computer-based support systems and organisational structures – on work in practice must be understood before we can decide on techniques for the measurement and prediction of the outcomes of people's work, or on recommendations for the design of systems. Our question is therefore how do we capture this contextual influence, but retain the power that can be obtained from the structure and control that we can have the opportunity to impose in a laboratory context.

Before we consider the relative merits of the laboratory and field, it is important that we distinguish the laboratory from 'experiments'. The experimental paradigm, the description of elements as independent or dependent variables, and the structure imposed on measurement, which may then lead to use of statistical analysis in some form, are useful tools that underpin scientific research (see Chapter 2). But this paradigm does not need to be restricted to a controlled laboratory context. For example if we are collecting data in a naturalistic driving study (such as was done in the 100 car study, Neale et al., 2005) we may specify one influencing ('independent') variable as being driver experience, and examine the number of driving errors or incidents recorded for drivers with varying numbers of years of driving. This allows us to utilise not only the experimental paradigm and infer causality (the influence of experience on driving performance) but *also* capture data in context. So the use of an experimental paradigm does *not* automatically imply the collection of data in a tightly controlled laboratory environment.

Laboratory studies have long been the dominant approach in, for instance, physical ergonomics research. This is a vital source of information and insight about isolated work variables, but may not be the most valid approach to understand work in practice and certainly will not replace the need for field study. The very nature of the traditional controlled laboratory environment means that the complexities and uncertainties of work environments are, by definition, not being simulated.

Sometimes, however, a field context simply does not yield the appropriate types of data, in an appropriate form, to enable us to study the research question that we wish to study. The context may be safety critical, or the methods which we wish to apply (e.g. concurrent verbal protocol, or frequent subjective ratings) may be intrusive and change the nature of task completion. If, for example, we wished to implement a new interface for representation of text-based communications in ATC,

---

* As mentioned earlier, this was an area about which John and I had much discussion. This section was written after John's death, and I fear he would disagree with some of the points here (or at least would have argued that they should be stated differently), so for a pure 'Wilson' view, the reader is referred to the third edition of this book. (SS).

it would be sensible to use a laboratory study to examine some of the basic issues associated with screen layout or complexity before implementing the tool in an active control context. A contribution that the field of E/HF has to make to the conduct of laboratory studies is in establishing approaches that allow us to *transfer* between the lab and the field; ensuring that we fully understand the extent to which we can generalise the data we collect in a laboratory context (so ensuring we report clearly what was and was not controlled for example) and designing those laboratory contexts and tasks to an appropriate level of complexity that allows the difference between behaviour in a field and laboratory context to be as close as possible. Simulation technologies are now extending to not only be able to capture aspects of individual control elements of work, but also the communications and some contextual elements (Sharples et al., 2011).

It is also our responsibility to ensure that any laboratory studies that we do design manipulate and examine the appropriate variables that are of importance in a field context. It is then also important to return to the field, to appreciate the roles of extraneous and influencing variables (e.g. environmental disturbances, work flow, individual motivation, inter-personal relationships and team relationships (Wilson et al., 2003) that have been controlled in a laboratory setting.

Work is difficult to study in practice, but ‘when the unit of analysis is interactions, then field research is arguably the main methodological approach for ergonomics’ (de Keyser, 1992). ‘Methodologically explicit field research is vital for the core purpose of ergonomics, investigating and improving interactions between people and the world around them’ (Wilson, 2000, p. 563).

We must be aware that field study naturally includes context from the environment, supervision, motivation and disturbances, but the very presence of the investigator will create an additional context which, without care, can invalidate the study. Well-planned, sensitive and long-term studies can minimise any influences (Webb et al., 1972; Bernard, 1995; Hammersley and Atkinson, 1995; Denzin and Lincoln, 1998). Ethnomethodologists talk of the ‘participant observer’, in special cases when the observer becomes a member of the group being studied and also generally, because they also participate as an actor in the workplace by their very presence as a researcher.

Our argument is for careful consideration of the use of field and laboratory contexts, and understanding of the value and contribution of both. The worst-case scenario is a field study that captures the wrong or insufficient data to allow detailed understanding of the research question, or a laboratory study that either fails to control variables, or is so complex that it is not possible to infer the influence of a useful isolated independent variable. If time and resources allow, an excellent approach can be seen as: starting in the field to identify the problem statement; collecting data using a mixed methods approach, using the field or laboratory as the question or researcher expertise demands and returning to the field to reflect upon the impact or context of findings. This is certainly not an argument against laboratory research – it is important to acknowledge that both laboratory and field studies are necessary – but is an argument for a better balance. Nor is there usually a stark either/or choice to be made.

QUALITY OF METHODS AND MEASURES AND THEIR CHOICE

In a recent commentary at the end of a special issue on ergonomics models and measurement, Annett (2002b) gently reminded us that ergonomics texts (the second edition of this one included!) do not always treat issues of the quality of ergonomics methods to the extent needed. In particular, he suggests that we have not properly considered the reliability and validity of our methods and measurement. Kanis (2014) with responses by Hignett (2014), Stanton (2014) and Vink (2014) continues this discussion. In fact, there are a number of ways in which we might judge the adequacy and quality of methods and measures, and therefore, we have a number of selection criteria for methods. These are discussed in the following text and the reader should see also any good research methods texts, such as Dane (1990), Frankfort-Nachmias and Nachmias (1996) or Robson (2011). It should be pointed out that we will rarely, if ever, be able to satisfy all of these criteria but equally we will rarely if ever need to for any particular study or investigation.

VALIDITY

The concept of validity can be defined as whether something measures what it claims to measure. This encompasses both whether a method looks like it measures what it actually measures, and demonstrates not only the importance of using carefully developed methods but also being very clear about what a method actually does or does not measure. E/HF presents a particular challenge here, as many of the terms we use, like workload, attention or teamwork, may have colloquial meanings that are understandable to the general public, but also specific meanings within the context of E/HF science. Therefore, one of our responsibilities, as referred to in the earlier discussion about devolving ergonomics, is to ensure that we clearly communicate the capabilities and appropriate implementation context of any method that we develop.

There are many different forms of validity; some types are embedded within others, some are very similar but are given quite different names.

Construct validity represents the extent to which a measure represents the concepts that it should, and particularly that it does not represent those that it should not. *Content validity* is complementary to construct validity and asks whether a measure is complete and measures all relevant parts of the concept it claims to measure.

Within construct and content validity, we might look for: *convergent validity* (the measure is consistent with outcomes from two or more different measures – a form of triangulation); *divergent validity* (the measure does not correlate closely with measures of different concepts); and *face validity* – see the following text.

Internal validity is the extent to which a measure allows a plausible demonstration of causal relationships (i.e. the identified cause does in fact lead to the assumed effect) and *external validity* relates to the generalisability of results or findings to other situations, times or contexts.

Criterion validity – the extent to which a measure is able to predict an outcome for one variable, based on data from another variable, has two forms: *concurrent validity* – whether the new measure compares well to outcomes of an existing 'accepted' measure (a limited form of and related to convergent validity) and *predictive validity* – how well the measure predicts what we eventually see happen in the real world.

Related to criterion validity, but important when transferring findings from a simulated or laboratory context to the real world, is *behavioural validity*. This describes whether behaviour observed or recorded in one context is consistent with that which would be found in a real-world context.

Finally, *face validity* is, literally, the validity of a measure or method 'on the face of it'. Sometimes seen as a form of construct validity, face validity considers the degree of consensus that a measure actually represents a particular concept and is strongly influenced by the look or appearance of a measure. It can also denote how acceptable a measure or method is to a particular set of stakeholders, which can be an important consideration when implementing measures in a real working context.

RELIABILITY

It is sometimes said that reliability is a special sort of validity, and certainly, it is required in order to be sure of validity. Reliability comes in many different forms and manner of assessment as well, but essentially asks whether we would get the same results and interpretations if we repeatedly use a method or measure; it is directly related to the amount of bias or error we may have allowed to creep in. Bias and error may be from the investigator (and her arrangements) or the study participant, and can have conscious or sub-conscious origins. We may look for assessment of reliability through inter-rater comparisons, use of test/retest procedures, application of different forms of measure at different times or on different groups, or through split half techniques whereby the measure is somehow divided into two for each participant (e.g. divide the questions in a scale into two halves), and compare the scores within participants.

GENERALISABILITY

Literally this refers to how well a method, measure or measurement, and data from that method, will generalise into other domains, situations, settings or populations. Influences and limitations on generalisability of methods may include the use of domain-specific terminology (e.g. in a specially designed questionnaire for health professionals) or the skills and experience of the participants being studied with respect to those to whom the results will be generalised. Generalisability of data is affected by the influence of context-specific variables on the behaviours or activities of interest in both the original domain and the domain to which the data is being applied. This can be difficult to determine, as it is not always clear whether extraneous variables have an influence on the data being collected.

INTERFERENCE

It is difficult to think of any ethical measurement of people's performance that does not affect or change the participants or their environment in some way, no matter how trivial. However, we would want to be sure that the reaction to the method and measurement process is not too great for the study purposes, and, critically, is known, acknowledged and reported. Rich participant observer accounts are often quite clear about the effect data gathering has had on the focus group, but the value of the data and interpretations frequently outweigh the scientific drawbacks. In an experimental setting, on the other hand, we would expect efforts to minimise any reactivity, to minimise bias and error and enhance reliability.

SENSITIVITY

We would not want to use a wooden ruler to measure changes in stature due to exposure to vibration – it would just not be sensitive enough to the size of changes we expect. On the other hand, we would not want to go to the expense of recruiting 500 participants in an experiment when the effects we are interested in will be shown by almost all people and certainly will show up with 10 participants (e.g. some visual functions or performance). Methods and measures should be selected which are sensitive to the right degree to the sorts of effects we are examining, whether to confirm or reject our hypotheses. Related to this, we need to select measures which can provide the level of detail needed for the particular enquiry, but not too much greater than this (and with the range of new technologies and sensors that are now available, there is considerable temptation to 'over-measure' a situation).

FEASIBILITY OF USE

Sometimes, it is said that methods should be simple to use – a little like product interfaces! This is true of methods which ergonomists have developed for use by others (engineers, managers, and health and safety specialists) in practical settings, but is not necessarily required for methods used by human factors and ergonomics specialists in laboratory or field. Some methods such as ethnographically informed observation, or specific techniques such as cognitive work analysis (Vicente, 1999 and see Chapter 3), require considerable training and experience in order to be able to apply them in a comprehensive form. However, we do want the methods used and the measures taken to be feasible for use in the particular circumstances we define, whether this is in relation to participants' understanding of questions, access points for observation to be set up, or enough data to populate a model meaningfully.

ACCEPTABILITY AND ETHICS

Strongly related to the idea of face validity and also to feasibility, the methods and measures must be acceptable – to the client, the study population to whom they are applied and to our peers and colleagues in the wider scientific community. Such acceptance may be in terms of resources required, interference with the activities of the population under study, or matters raised which the client would prefer not to be (e.g. worker opinions on matters outside the focus of the study). Very much related is the idea of ethical research and practice, where we should follow codes of conduct and guides to ethical practice such as produced by professional bodies (e.g. the Institute of Ergonomics and Human Factors or the UK Economic and Social Research Council). Key aspects of these are the way in which we treat our participants, how we represent ourselves and how we report on our findings (see Chapter 4 for further discussion of research ethics).

Resources

As with any endeavour, the resources required will be an important determinant of the methods and measures we use. Not only that but it will be the resources needed set against the importance of the study and the potential value of its outcomes that will be critical. Resources are not just to do with financial cost, although this is a central consideration; we must also assess the resources needed in terms of people (whether as investigators/researchers or as participants), equipment and also time – for the analysis as much as for the data collection, as is seen regularly in observation studies employing video analysis.

CONCLUSIONS

E/HF has as its field the interactions of people with all other people and artefacts within an environmental context; its goal is the well-being of individuals, organisations and national economies. This requires appropriate theory and practice, models and methods. The thin line which we must tread when we become involved in discussion of methodology is summed up in two views: on the one hand, '[psychology] should not allow itself to be driven by obsession with method to the exclusion of the human problems that are its province' (Barber, 1988, p. 7, reporting Maslow, 1946); on the other hand 'anyone who wishes to reflect on how they practise their particular art or science, and anyone who wishes to teach others to practise, must draw on methodology' (Cross, 1984, p. vii). Certainly, ergonomists are driven by human problems, or driven to improve the quality of people's lives, but also we are always concerned to educate others (designers, engineers, politicians, accountants, managers, public, and media) in our approach and the necessity for it. If we truly believe that we are, above all, promoters of an approach and of a process, then it behoves us to pay great attention to the roots and current and future state of our methodology. The remainder of this book is but one step in doing this.

Earlier in this chapter, we explored frameworks for ergonomics and human factors, and in doing so suggested that the systems with which we should now be concerned cannot be modelled and understood by reference to a single person at the centre of a set of interactions with equipment and environment, but by representations of groups of people acting in social networks. In fact much work and life of the future, and many of the challenges facing ergonomics and human factors, will be found in complex interacting and distributed sociotechnical systems. The activities of interest are no longer just those of command and control, with individual operators working with single control display interfaces, but collaboration, co-ordination and integration, with constantly shifting and multiple teams of people, co-located and virtual, interacting with each other directly and via various forms of information and communication technologies (see Edwards and Wilson, 2004).

Because this focus for ergonomics is increasingly important in the future, then we will need to extend our range of methods to deal with it. We certainly will need to draw from fields such as ethnography (e.g. Hammersley and Atkinson, 1995; Engstrőm and Middleton, 1996), social network analysis (Wasserman and Faust, 1994), and interaction design (Rogers et al., 2011), and utilise the best new developments in cognitive work analysis and design (e.g. Hollnagel, 2003). We continue to be in an exciting time for ergonomics methods in this area, and it will be a considerable challenge to find ways in which we can truly understand peoples' behaviour in such rich distributed sociotechnical systems, and at the same time provide data and findings with enough rigour to be applicable in future design and implementation.

This book is about methods in ergonomics and human factors. In introducing the reader to how we study people's interactions with artefacts, environments and other people and how we gather information, form models, develop recommendations and designs and enhance understanding, we are also introducing them to ergonomics and human factors in general. Many other books, as noted throughout this chapter, do this as well, to different levels of detail, and in doing so introduce human factors methodology. The aim of *this* book is to discuss and consider the goals of different methods in different work contexts, to understand how we can best combine and apply the range of tools and techniques that we have available in a manner that ensures that we effectively support design of work systems and tools.

REFERENCES

Annett, J. (2002a). Subjective rating scales: Science or art? *Ergonomics*, **45**, 966–987.

Annett, J. (2002b). A note on the validity and reliability of ergonomics methods. *Theoretical Issues in Ergonomics Science*, **3**, 228–232.

Bainbridge, L. (1987). The ironies of automation. In: J. Rasmussen, K. Duncan and J. Leplat (eds.), *New Technology and Human Error*. London, U.K.: John Wiley.

Barber, P.J. (1988). *Applied Cognitive Psychology*. London, U.K.: Methuen.

Beaty, D. (1995). *The Naked Pilot*. Shrewsbury, U.K.: Airlife Pub. Co.

Becker, B. and Gerhart, B. (1996). The impact of human resource management on organizational performance: Progress and prospects. *Academy of Management Journal*, **39**(4), 779–801.

Bernard, H.R. (ed.). (1995). *Research Methods in Anthropology*: *Qualitative and Quantitative Approaches*, 2nd edn. Walnut Creek, CA: Altamira Press.

Beynon-Davies, P. (1999). Human error and information systems failure: The case of the London ambulance service computer-aided dispatch system project. *Interacting with Computers*, **11**, 699–720.

Bignell, V., Peters, G. and Pym, C. (1977). *Catastrophic Failures*. Milton Keynes, U.K.: Open University Press.

Casey, S. (1993). *Set Phasers on Stun*. Santa Barbara, CA: Aegean Pub. Co.

Chapanis, A. (1976). Keynote address: Ergonomics in a world of new values. *Ergonomics*, **10**(3), 253–268.

Chapanis, A. (1979). Quo vadis, ergonomia. *Ergonomics*, **22**, 595–605.

Chapanis, A. (1996). *Human Factors in Systems Engineering*. New York: John Wiley.

Corlett, E.N. (1991). Some future directions for ergonomics. In: M. Kumashiro and E.D. Megaw (eds.), *Towards Human Work: Solutions to Problems in Occupational Health and Safety*. London, U.K.: Taylor & Francis Group.

Cross, N. (1984). *Developments in Design Methodology*. Chichester, U.K.: John Wiley.

Cushing, S. (1994). *Fatal Words*. Chicago, IL: University of Chicago Press.

Dane, F.C. (1990). *Research Methods*. Belmont, CA: Wadsworth.

Denzin, N. (1970). *Sociological Methods.* New York: McGraw-Hill.

Denzin, N.K. and Lincoln, Y.S. (eds.). (1998). *Strategies of Qualitative Inquiry*. London, U.K.: Sage Publications.

Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W.S., Wilson, J.R. and van der Doelen, B. (2012). A strategy for human factors/ergonomics: Developing the discipline and profession. *Ergonomics*, **55**(4), 377–395.

Eason, K.D. (1997). Understanding the organizational ramifications of implementing information technology systems. In: M. Helander, T. Landauer and P. Prabhu (eds.), *Handbook of Human-Computer Interaction*, 2nd edn. Amsterdam, the Netherlands: Elsevier.

Edholm, O.G. and Murrell, K.F.H. (1973). *The Ergonomics Research Society: A History 1949–1970*. Loughborough, U.K.: The Ergonomics Society.

Edwards, A. and Wilson, J.R. (2004). *Guide to Implementation of Virtual Teamworking*. London, U.K.: Gower Press.

el Kaliouby, R. and Robinson, R. (2005). Real-time inference of complex mental states from facial expressions and head gestures. In: *Real-Time Vision for Human-Computer Interaction.* Berlin: Springer-Verlag, pp. 181–200.

Engstrőm, Y. and Middleton, P. (eds.). (1996). *Cognition and Communication at Work*. Cambridge, U.K.: Cambridge University Press.

Frankfort-Nachmias, C. and Nachmias, D. (1996). *Research Methods in the Social Sciences*. London, U.K.: Arnold.

Green, B. and Jordan, P.W. (1999). The future of ergonomics. In: M.A. Hanson, E.J. Lovesey and S.A. Robertson (eds.), *Contemporary Ergonomics,'99*. London, U.K.: Taylor & Francis Group, pp. 110–114.

Grey, S.M., Norris, B.J. and Wilson, J.R. (1987). *Ergonomics in the Electronic Retail Environment.* Slough, U.K.: ICL (UK) Ltd.

Guardian. (2014). OMG! Number of UK text messages falls for first time. Published online 13 January 2014.

Guion, L.A., Diehl, D.C. and McDonald, D. (2011). *Triangulation: Establishing the Validity of Qualitative Studies.* Gainesville, FL: Department of Family, Youth and Community Sciences, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, FCS6014, at http://edis.ifas.ufl.edu/pdffiles/FY/FY39400.pdf. Accessed 24th March 2014.

Hammersley, M. and Atkinson, P. (1995). *Ethnography*, 2nd edn. London, U.K.: Routledge.

Helander, M.G. (1999). Seven common reasons not to implement ergonomics. *International Journal of Industrial Ergonomics*, **25**, 97–101.

Hignett, S. (2014). Commentary on the paper by Heimrich Kanis entitled: 'Reliability and validity of findings in ergonomics research'. *Theoretical Issues in Ergonomics Science*, **15**(1), 50–54.

Hignett, S. and Wilson, J.R. (2004b). Horses for courses – but no favourites. A reply to three commentaries. *Theoretical Issues in Ergonomics Science*, **5**(6), 517–525.

Hignett, S.M. and Wilson, J.R. (2004a). The role for qualitative methodology in ergonomics: A case study to explore theoretical issues. *Theoretical Issues in Ergonomics Science*, **5**(6), 473–493.

Hollnagel, E. (ed.). (2003). *Handbook of Cognitive Task Design*. Boca Raton, FL: CRC Press.

Hollnagel, E. and Woods, D.D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton, FL: CRC Press.

Hollnagel, E. (2014). Safety-I and Safety-II The Past and Future of Safety Management. Farnham, UK: Ashgate.

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

Jian, J.-Y., Bisantz, A.M. and Drury, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, **4**(1), 53–72.

Johnson, G.I. and Wilson, J.R. (1988). Future directions and research issues in ergonomics and advanced manufacturing technology (AMT). *Applied Ergonomics*, **19**(1), 3–8.

Kanis, H. (2014). Reliability and validity of findings in ergonomics research. *Theoretical Issues in Ergonomics Science*, **15**(1), 1–46.

Kantowitz, B.H. (1992). Selecting measures for human factors research. *Human Factors*, **34**, 387–398.

de Keyser, V. (1992). Why field studies? In: M. Helander and M. Nagamachi (eds.), *Design for Manufacturability: A Systems Approach to Concurrent Engineering and Ergonomics*. London, U.K.: Taylor & Francis Group, pp. 305–316.

Kim, S.-W., Drury, C.G. and Lin, L. (2001). Ergonomics and quality in paced assembly lines. *Human Factors and Ergonomics in Manufacturing*, **11**(4), 377–382.

Kragt, H. (1992). *Enhancing Industrial Performance.* London, U.K.: Taylor & Francis Group.

Kuorinka, I. (ed.). (2000). *History of the International Ergonomics Association*. Boca Raton, FL: IEA Press.

Mack, Z. and Sharples, S. (2009). The importance of usability in product choice: A mobile phone case study. *Ergonomics*, **52**(12), 1514–1528.

Meister, D. (1985). *Behavioural Analysis and Measurement Methods.* Chichester, U.K.: John Wiley.

Meister, D. (1995a). Divergent viewpoints: Essays on human factors questions. Personal publication.

Meister, D. (1995b). Human factors – the early years. *Proceedings of the Human Factors Society 39th Annual Meeting*, San Diego, CA, pp. 478–480.

Meister, D. (April 1998). Twenty-first-century challenges to ergonomics. *Ergonomics in Design*, **6**, 33–34.

de Montmollin, M. (June 1992). The future of ergonomics: Hodge Podge or new foundation? *Le Travail Humain*, **55**(2), 171–181.

Moray, N. (1994). "De Maximum non Curat Lex" or how context reduces science to art in the practice of human factors. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, Nashville, TN, pp. 526–530.

Moray, N. (2000). Culture, politics and ergonomics. *Ergonomics*, **43**(7), 858–868.

Muir, H.C., Bottomley, D.M. and Marrison, C. (1996). Effects of motivation and cabin configuration on emergency aircraft evacuation behavior and rates of egress. *The International Journal of Aviation Psychology*, **6**(1), 57–77.

National Audit Office. (2011). The failure of the firecontrol project. HC 1272 Session 2010–2012. Department for Communities and Local Government. http://www.nao.org.uk/wp-content/uploads/2011/07/10121272es.pdf. Accessed 24th March, 2014.

National Research Council. (1986). Research needs for human factors. Report of the Committee on Human Factors. Washington, DC: National Academy Press.

Neale, V.L., Dingus, T.A., Klauer, S.G. and Sudweeks, J. (2005). An overview of the 100-car naturalistic study and findings. Paper 05-0400. National Highway Traffic Safety Administration. http://www-nrd.nhtsa.dot.gov/pdf/ESV/esv19/05-0400-W.pdf. Accessed 24th March, 2014.

Norman, D. (1988). *The Psychology of Everyday Things*. New York: Basic Books.

Oxford English Dictionary. (2014). www.oed.com. Accessed 24th May, 2014.

Perrow, C.B. (1984). *Normal Accidents: Living with High Risk Technologies*. New York: Basic Books.

Petroski, H. (1992). *To Engineer Is Human: The Role of Failure in Successful Design*. New York: Vintage Books.

Pheasant, S. (1986). *Bodyspace*. London, U.K.: Taylor & Francis Group.

Pickup, L., Wilson, J.R., Norris, B.J., Mitchell, L. and Morrisroe, G. (2005). The Integrated Workload Scale (IWS): A new self-report tool to assess railway signaller workload. *Applied Ergonomics*, **36**(6), 681–693.

Poulton, E.C. (1975). Observer bias. *Applied Ergonomics*, **6**(1), 3–8.

Proctor, R.W. and Zandt, T.V. (2008). *Human Factors in Simple and Complex Systems*, 2nd edn. Boca Raton, FL: Taylor & Francis Group.

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

Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Aldershot, U.K.: Ashgate.

Robson, C. (2011). *Real World Research*. Oxford, U.K.: Blackwells.

Rogers, Y., Sharp, H. and Preece, J. (2011). *Interaction Design: Beyond Human-Computer Interaction*. Chichester, U.K.: Wiley & Sons.

Rouse, W., Kober, N. and Mavor, A. (eds.). (1997). *The Case for Human Factors in Industry and Government, Report of a Workshop*. Washington, DC: National Academy Press.

Rutherford, A. and Wilson, J.R. (1992). Searching for the mental model in human–machine systems. In: Y. Rogers, A. Rutherford and P. Bibby (eds.), *Models in the Mind: Perspectives, Theory and Application*. London, U.K.: Academic Press, pp. 195–223.

Scaife, M. and Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, **45**, 185–213.

Shackel, B. (July 1996). Ergonomics: Scope, contribution and future possibilities. *The Psychologist*, **9**(7), 304–308.

Sharples, S., Millen, L., Golightly, D. and Balfe, N. (2011). *http://www.swetswise.com/common/images/spacer.gif*The impact of automation in rail signalling operations. *Journal of Rail and Rapid Transit. Proceedings of IMechE, Part F*, **225**(2), 179–191.

Singleton, W.T. (ed.). (1982). *The Body at Work: Biological Ergonomics.* Cambridge, U.K.: Cambridge University Press.

Singleton, W.T. (July 1994). From research to practice. *Ergonomics in Design*, **2**(3), 30–34.

Sparks, E.A., Wehbe-Janek, H., Johnson, R.L., Smythe, W.R. and Papaconstantinou, H.T. (November 2013). Surgical safety checklist compliance: A job done poorly! *Journal of the American College of Surgeons*, **217**(5), 867–873.

Stanton, N.A. (2014). Commentary on the paper by Heimrich Kanis entitled 'Reliability and validity of findings in ergonomics research': Where is the methodology in ergonomics methods? *Theoretical Issues in Ergonomics Science*, **15**(1), 55–61.

Stockbridge, H.C.W. (1989). *The Ergonomics Society: A History 1971–1989*. Loughborough, U.K.: The Ergonomics Society.

Torgerson, W.S. (1958). *Theory and Methods of Scaling*. New York: Wiley.

UK Government. (2012). Report of the service inquiry into the grounding of HMS Astute on 22 October 2010. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/27118/astute_grounding_si_report.pdf.

UK National Audit Office. (27 October 1999). The passport delays of summer 1999. Report by the Comptroller and Auditor General. London, U.K.: The Stationery Office.

UK Public Accounts Committee. (2000). Improving the delivery of government IT projects. Report of the House of Commons Public Accounts Committee. London, U.K.: The Stationery Office.

Vicente, K.J. (1999). *Cognitive Work Analysis*. Mahwah, NJ: Lawrence Erlbaum.

Vink, P. (2014). Reaction on the paper 'Reliability and validity of findings in ergonomics research' of Kanis (2013). *Theoretical Issues in Ergonomics Science*, **15**(1), 47–49.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, U.K.: Cambridge University Press.

Wastell, D.G. and Cooper, C.L. (1996). Stress and technological innovation: A comparative study of design practices and implementation strategies. *European Journal of Work and Organizational Psychology*, **5**, 377–397.

Waterson, P.E. (2009). A systems ergonomics analysis of the Maidsone and Tunbridge Wells infection outbreaks. *Ergonomics*, **52**, 1196–1205.

Waterson, P.E. (2011). World War II and other historical influences on the formation of the Ergonomics Research Society. *Ergonomics*, **54**(12), 1111–1129.

Waterson, P.E. and Sell, R. (2006). Recurrent themes and developments in the history of the Ergonomics Society. *Ergonomics*, **49**(8), 743–799.

Webb, E.J., Campbell, D.T., Schwartz, R.D. and Sechrest, L. (1972). *Unobtrusive Measures: Non-reactive Research in the Social Sciences.* Chicago, IL: Rand McNally.

Welford, A.T. (1976). Ergonomics: Where have we been and where are we going: I. *Ergonomics*, **19**, 275–286.

Wickens, C.D., Hollands, J., Banbury, S. and Parasuraman, R. (2013). *Engineering Psychology and Human Performance*, 4th edn. Boston, MA: Pearson.

Wilson, J.R. (1994). Devolving ergonomics: The key to ergonomics management programmes. *Ergonomics*, **37**, 579–594.

Wilson, J.R. (2000). Fundamentals of ergonomics. *Applied Ergonomics*, **31**, 557–567.

Wilson, J.R. (2014). Fundamentals of systems ergonomics/human factors. *Applied Ergonomics*, **45**, 5–13.

Wilson, J.R., Jackson, S. and Nichols, S. (2003). Cognitive work investigation and design in practice: The influence of social context and social work artefacts. In: E. Hollnagel (ed.), Handbook of cognitive task design. Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, J.R. and Nichols, S.C. (2002). Measurement in virtual environments: Another dimension to the objectivity/subjectivity debate. *Ergonomics*, **45**(14), 1031–1036.

Wogalter, M.S., Hancock, P.A. and Dempsey, P.G. (1998). On the description and definition of human factors/ergonomics. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, pp. 671–674. Santa Monica, CA.

Zhang, L., Helander, M.G. and Drury, C.G. (1996). Identifying factors of comfort and discomfort in sitting. *Human Factors*, **38**(3), 377–389.

Section I

Fundamental Approaches and Tools

The first section of this book describes a set of general approaches and methods that might be employed in any ergonomics/human factors (E/HF) enquiry or investigation, whether research driven or practice led. These methods help us to improve our fundamental understanding of people, whether in a laboratory or field context, and generally focus on capturing their activities, responses, reactions and thoughts. They might fit with an empirical approach, where we collect quantitative or qualitative data to record behaviours or opinions, or an analytical approach, where we create models or abstractions of people and what they do.

Many of the terms used within this field are heavily loaded, and there has been detailed and careful discussion about the titles of some of the chapters within this section. Distinctions such as direct/indirect or subjective/objective have been made in the past to classify methods, but these terms can be confusing or problematic in implying some element of 'bias' in subjectivity or interpretation of data. Broadly, the E/HF practitioner should normally be able to select the method appropriate for the question or situation, and be aware of the opportunities and limitations associated with that method.

The methods presented here include those which enable direct observation or capture of behaviours and interactions, and those which are indirect, and thus rely on interpretation from either the participant themselves (e.g. in terms of the way in which they respond to a question using a rating scale) or the experimenter (when classifying a behaviour whilst using a structured observation framework).

Although the direct observation group of methods is sometimes seen as 'objective' and somehow 'true', it is important to note that even data collected by direct observation will be re-analysed, summarised and abstracted by the observers, from their memory, notes, tape or video recordings, and then interpreted by the investigators, introducing some possible bias or at least subjectivity. This is even becoming apparent in very recent work which considers 'big data' analysis, where very large data sets, collected by, for example, commercial organisations, or technology-led companies such as Google, provide the apparent opportunity to no longer need to 'sample' and instead apply analysis approaches to the 'complete population'. This needs to be done with caution – even in these data sets drawn from exceptionally wide sources, we cannot be confident that the complete population has been captured, and we should not forget the value of the statistical approaches that have underpinned our interpretation of experimental data for many years (as discussed in Chapter 2).

Similarly, a well-designed and implemented ‘subjective’ tool can elicit extremely valuable and reliable reports that simply could not be captured using ‘objective’ means.

What remains of course is the need for any E/HF practitioner to understand fully the nature of the method that they are applying, its limitations and the opportunities for the insight that it provides. The methods presented in this section enable readers to understand the underpinning considerations in selection of these broad methodological approaches, which can then be applied in a range of contexts, from laboratory to field.

Whilst much of the focus of this book is on methods that can be applied in the field context, it remains important for the E/HF practitioner to be aware of the fundamentals in experimental design and analysis. Controlled experiments allow us to isolate individual variables and, if extraneous variable are appropriately managed and robust measures applied, infer causative relationships between independent and dependent variables. Chapter 2, by Drury et al., presents a detailed description of the considerations involved in experimental design and analysis, with particular focus on the statistical consequences of different experimental design and analytical approaches. Whilst much of our work as E/HF practitioners is focussed in the field, or in tasks where it is simply too complex to control all of the influential variables, data to inform much of the underlying theory that informs our thinking, whether it is our response to an environmental variable, or the physical effort exerted in a task, is most effectively captured in a laboratory setting. Laboratory approaches allow us to test repeatability of results and understand the response of variables in isolation. So that such studies and experiments do not end up as costly, sterile exercises, great care must be taken in their design and conduct. The trade-off – being able to isolate the variables of greatest interest but by doing so perhaps losing the influence of setting and context – must be understood and assessed carefully before any choice to carry out experimental research.

Chapter 3 presents a consideration of the capture of complex, cognitive work. Whereas many of the later chapters in this volume use examples of physical and physiological responses, which can often be directly measured by instrumentation, this chapter focusses on methods used to capture and interpret tasks that are primarily cognitive. Such tasks are inherently harder to observe, and rely on a knowledgeable and experienced researcher capturing and analysing data that has been collected. Methods considered include structured interview approaches such as critical decision method, observational methods and generative methods, that are particularly useful when establishing user needs.

The general considerations that should be made when applying methods to capture or elicit responses from individuals more generally are presented in Chapter 4. In this chapter, a range of different methods and specific guidance to their design and approach are presented. This chapter covers a very wide range of general ergonomics techniques, from rating scale and questionnaire design, to developing interview protocols, to the ethical considerations of applying methods to capture and observe participant responses. A number of phenomena encountered whilst applying these methods, such as participant bias, are described, and approaches to minimise the effects of these biases on data collected are suggested. In addition, the psychophysics approach is described in this chapter. This is a method that underpins much of our understanding of how people perceive different sensations, and provides a specific approach to the quantification of responses. It is included in this book partly to ensure it does not become a ‘lost art’ but also to demonstrate the value and requirements of such a structured approach in eliciting specific responses from participants.

A specific chapter on qualitative methods was first introduced in the third edition of this book; the use of such methods now underpins much E/HF work both in research contexts and real world practice. Whilst in some situations, there is still a preference for, or perhaps confidence in, data that can be reported in a quantitative form, and the notion of statistical significance is still particularly important, the value of qualitative data, particularly when explaining *why* or *how*, is increasingly acknowledged. Hignett and McDermott, in Chapter 5, present the different philosophies associated with applying qualitative methods, and consider the different techniques that can be used to support such approaches.

A method that remains core to E/HF is task analysis (Chapter 6). General opinion within the discipline emphasises that task analysis is, as the name suggests, not just description but also analysis of tasks; within task analysis, data are collected, represented in an appropriate description form, and are analysed to assess task requirements for the person, their expected behaviour, any constraints and the task and environmental demands on them.

Task analysis is widely used, since it can be applied during the analysis of existing systems, the design of new ones and the evaluation carried out subsequently. Information gained is useful both in development and as criteria against which to assess what is developed. What is represented and analysed is what must be done in order to fulfil certain goals, within constraints from the task environment and from individual or general human limitations. Task analysis and its techniques were originally distinguished from method study both by underlying purpose but also by their concentration upon operator decisions as much as upon actions. This chapter also touches upon user modelling approaches, which use outputs of task analysis as the basis for predictive models of user performance to inform and test prospective designs.

One particular goal of task analysis and analysis of cognitive work is to understand the nature, manifestation and elicitation of knowledge, taken up by Shadbolt and Smart in Chapter 7. Human factors have a role to play in addressing the bottleneck in knowledge-based systems – acquiring the knowledge necessary to build the system. Generally, the major source of such knowledge will be the experts themselves; knowledge elicitation is the name for gathering the relevant information about people's knowledge. This chapter also describes semantic web approaches to representation of knowledge elicited from individuals in machine-readable form, moving beyond dedicated 'expert systems' to a generic approach that allows knowledge to be captured and stored in a structured and searchable manner. In some ways, parallels can be drawn between this and chapters on involving people in design research and participatory ergonomics, later in this volume, in that the techniques described there are to elicit information, opinion and insight from users or user 'representatives', to be applied in design.

Eliciting expertise uses a number of methods from other domains but has also seen development of a number of methods explicitly for the purpose of understanding the knowledge and skills that people hold, which is why this chapter appears in this general section. Moreover, the focus is now on the support for knowledge management, knowledge-based systems of all kinds and for design of jobs and organisations that make the most of the skills and knowledge of the staff.

Finally, in Chapter 8, the approach of user modelling in simulation and digital human modelling forms is discussed. Many tools now exist to represent users and their potential work contexts in a three-dimensional form – if designed and applied appropriately, they provide a rich resource to analyse the impact of a design intervention or modification earlier in the design phase than would previously have been possible. However, as with other methods, it is vital that a practitioner is aware of the limitations of these approaches; in the case of digital human modelling for example, a model is only going to be as good as the source from which anthropometric data is drawn and the biomechanics built within the computer representation. If the underpinning representation is 'fit for purpose' (i.e. it has the accuracy and granularity required for the topic under question), then this can be a very powerful and valuable predictive technique, not only for analysing the cognitive, physical and environmental 'fit' of a future design, but also for communicating the output of an E/HF analysis to key stakeholders and decision makers.

A message throughout this book is that methods and techniques are only tools, to be selected and applied with a clear understanding of the design or evaluation objectives. However, although skilful use of methods requires practice, we should beware of sticking rigidly to standard methods. Imagination, as well as scientific rigour, as stated in the third edition of this text, remains a key ingredient of successful ergonomics.

Sarah Sharples

2 Experimental Design and Analysis

Colin Drury, Victor Paquet and Harrison Kelly

CONTENTS

INTRODUCTION

This chapter provides the ergonomics/human factors (E/HF) practitioner and researcher with practical advice on how to design studies and analyse the resulting data to achieve effectiveness and efficiency. In the 1930s, there was a major revolution in how experiments are designed. We moved from the traditional physical sciences model of 'vary one factor at a time; keep all other factors fixed' to a philosophy that emphasises varying multiple factors in the same experiment. What made this revolution possible was the development of sophisticated statistical techniques (e.g. analysis of variance [ANOVA]) that allowed for the parsing of the effects of each factor, and their combinations, and the testing of each effect against the normal variation experienced in any experiment. This is the model for experimental design and analysis we still use in E/HF.

Within this chapter, we use a number of terms as if they were already understood by researchers and practitioners. Examples are 'experiment', 'factor', 'effect' and 'statistical techniques'. We can do this because most E/HF professionals have had courses with titles such as 'design of experiments' (DoE) at some point in their careers. This chapter will cover some familiar ground, such as multifactorial experiment designs and analysis of variance, but will hopefully extend the topic in a number of ways. First, we will provide some definitions that broaden the topic to other E/HF studies. Next, we will delineate statistical techniques for data that does not fit the usual normality assumptions, and finally, we will add material on power calculations, effect magnitudes and meta-analysis that will at least provide sources for further reading.

WHAT IS AN EXPERIMENT?

An experiment is 'a scientific procedure undertaken to make a discovery, test a hypothesis, or demonstrate a known fact' according to the Oxford Dictionary (www.oxforddictionaries.com, 2014). Experiments differ from all other techniques in that they directly change a system, normally in the form of a manipulation of the independent variables. These independent variables are derived from research hypotheses and enable the measurement of the degree to which each independent variable impacts and affects the dependent variable of interest. The research hypotheses are usually stated in the form of a null hypothesis, which states that the independent variable has no effect, and an alternative hypothesis, which states that the independent variable has an effect. The convention is to use statistical results to either 'reject' or 'fail to reject' the null hypothesis with a degree of certainty (sometimes also reported as supporting the alternative hypothesis). Hypotheses are discussed in more depth later in this chapter.

Alternative explanations of the results in terms of coincidence are most unlikely, because the experimenter determined when and how to change the system. Thus, experiments are able to detect and infer causality with some degree of confidence. A typical designed experiment is that of Drury et al. (2008) who studied the effects of workplace posture on performance and comfort in a security screening task. They used three postures at the airport workplace for screeners viewing x-ray images of baggage potentially containing threat objects. Screeners could be either standing, sitting on a high chair or sitting on a normal-height chair. The design of the experiment used three factors: posture, run order (to measure any learning/fatigue) and participant (12 experienced screeners). (The findings were no effects of posture on performance, but large effects of posture on discomfort and of run order on performance.) The results suggest that 'posture' can affect discomfort and 'order' can affect learning, but no conclusions can be reached about the effects of 'posture' on performance or 'order' on discomfort.

Experiments are not the only form of E/HF enquiry, as other chapters in this book amply demonstrate. However, they do have the most obvious impact on determining causality. Also, the same statistical design techniques used in experiments can also be used in other forms of study such as observation studies (where, as discussed in Chapter 1, we can sometimes impose an 'experimental paradigm'). We can choose which factors and which levels of these factors to observe, and apply similar statistical techniques to the data to understand the separate effects. As an example, Chang and Drury (2007) studied human interaction with doors by observing 1600 people as they used different doors. The doors (both push and pull types) were chosen to give four levels of physical difficulty, characterised by the restoring torque in N m needed to move the door. The people observed were characterised by their strength, determined by their gender and stature (observed against a set of marks on the door). The measurement was the use of body weight, rather than just one hand, to help open the door. Even though this was an observational study rather than an experiment, a statistical multifactorial design was used, one recognisable by anybody taking a DoE course. (Chang and Drury found, as expected, that task demands [door difficulty] and human capability [stature used as a proxy for upper body strength] jointly determined the likelihood of using body weight to open the door.)

As noted in Chapter 1, an experiment is NOT automatically a laboratory study. While laboratory studies are often experiments, so are field studies. For some classes of enquiry, such as examining the details of visual search or biomechanics, a laboratory may be the most convenient place to control the factors of interest and measure their effects precisely. But a valid experiment can take place in a field setting, for example measuring the effects of changes in workplace design or aircraft cockpit advances. In a typical field study, Latorella and Chamberlain (2001) asked experienced pilots to fly a twin-turbo-prop aircraft towards convective weather fronts to measure the effectiveness of three different weather displays on the pilots' decision points for avoiding the weather front. Note that the experimenters made deliberate changes to the system (three displays, nine quantised distances from the weather front) and were in a position to measure both the pilots' decisions and their responses to weather questionnaires. This was not a simplified laboratory simulation, although much prior research derived from laboratory simulations was used in the detailed design of the experiment.

Any experiment changes the system under study. Thus, it is reactive on the system. Participants know the experiment is taking place and most details of how it will affect them. In any academic or research institution or government organisation, an Institutional Review Board (IRB) or ethics committee will have vetted the study before it is allowed to proceed. This helps ensure that the experiment is ethically designed and executed, protecting all participants to the extent possible, but it also emphasises that an experiment does impact the system and its participants. An IRB review might not always be used in an enterprise setting (business, industry) where the enterprise employs the experimenters and participants, although it should always be the norm.

Direct intervention can be both costly and potentially dangerous. Given that, what does this reactivity buy for the E/HF professional? In addition to the major advantage of determining causality, the three other advantages of a more highly reactive design are as follows:

1. The ability to be in the right place at the right time to observe. This is particularly important in human factors studies where the system behaviour observed is rare and unexpected (e.g. accidents or breakdowns).
2. The ability to use more obviously invasive, but information-rich, measurement techniques. For example, in inspection research, the response to each individual item inspected can be observed in considerable detail in an experiment (e.g. Drury and Sinclair, 1983), whereas in the real situation, only a simple accept/reject response is often given.
3. The ability to control or manipulate other variables not of primary interest that would otherwise alter the results of the investigation. Variables that are unaccounted for and contribute to variability in the measured effects of a study can mask the effects of the independent variable. In cases where such unaccounted variables are also correlated with the independent variable, the results of a study become 'confounded' by the unaccounted-for or spurious variable and it is impossible to determine the relative contributions of the independent variable and the confounding variable. Experiments, when designed properly, account for these variables so as to limit the spurious variability of the effects and minimise confounding. For example, in the security screening study, the order of presentation of conditions to screeners would be likely to have an effect, so this variable was controlled by explicitly including it as a factor in the study.

If the E/HF professional gains in experimental control and measurement detail by using highly reactive designs, what is lost? The major loss is in face validity. If we observe a system in its natural state, those associated with the system and possibly those who commissioned the study can be convinced that the study is realistic. An experiment, particularly one performed in a laboratory with artificial stimuli and non-representative participants, requires much more persuasion on the part of the E/HF professional to gain acceptance. The lead author was once involved in two studies of fork-lift truck control. One (Drury and Dawson, 1974) involved real drivers using real fork-lift trucks in a real warehouse to study lateral control behaviour. The other (Drury et al., 1974) involved real drivers controlling a toy train in a laboratory to study longitudinal control behaviour similar to Fitts' law tasks. It is obviously much easier to quote the former study to convince warehouse managers of its design implications.

FACTORS AND LEVELS

What scientists refer to as independent variables are characterised in DoE as factors. They are the things the experimenter varies in order to measure their effects. In the doors study example, the factors were 'physical difficulty', 'gender', 'push vs. pull' and 'stature'. For the weather display study, the factors were ‘display’, ‘proximity to weather front’ and ‘participant’. In the security screening study, the factors were ‘posture’, ‘run order’ and ‘participant’. These are examples of factors designed into the experiment with specific values, called levels. For example, in the doors study, the factor ‘gender’ was at two levels: male and female. The factor ‘physical difficulty’ had four levels of door restoring torque: 30, 46, 55 and 72 N m. In the weather display study the factor ‘display’ was at three levels: aural only, out-the-window view plus aural and a graphic weather display plus aural. In that study also, there were six participants and the factor ‘proximity to weather front’ was at six levels: every 20 nautical miles from 120 down to 20. In the security screening study, ‘posture’ was at three levels: standing, high chair and normal chair, and so on. The combinations of levels studied in an experiment are called experimental conditions.

Statistical DoE emphasises factors and levels, almost to the exclusion of other ways of treating independent variables. In any study, there are a potentially infinite number of variables that could possibly affect the outcome. To reduce the normal variation between data points, the E/HF professional must deal with all of these variables, even though that sounds impossible in theory. Some independent variables become part of the experiment by including them as factors at a number of levels (>1), as in the earlier examples. The problem with allowing each variable to take on numerous levels is that in most experimental design solutions, the total number of trials is determined by the *product* of the number of levels of each factor. At the other extreme are what a physical scientist would call ‘nuisance variables’, which could affect the outcome if not controlled in some way.

The most useful ways of controlling these extraneous variables are to fix them at a single level or to use random assignment so that they do not *systematically* bias the outcome and conclusions. Fixing a variable at a single level is the most obvious way to control it. In the doors example, the whole study took place on a campus, eliminating some variability in age. For the weather display study, all participants had to be instrument-rated pilots with minimum experience defined by specific numbers of flights and flight hours to eliminate much variability due to skill and experience. Trained security screeners were used in the security study for the same reason. There is a price to be paid for fixing a variable at a single level in that the results may strictly apply only to that level of that variable. Thus, the doors study outcomes would not necessarily apply to, for example, people in senior citizen housing, in an elementary school or in a private home. In the weather display study, we could not generalise to, for example, novice pilots or military pilots.

Using randomisation to prevent, for example, all of the older participants being tested on a Monday is a powerful tool that prevents systematic bias by ensuring that any uncontrolled variability contributes only to the ‘normal variation’. In statistical terms, this prevents bias to the mean, possibly at the cost of increasing the residual variance. Randomisation is a safe way to control unwanted variation, but the cost of the study might increase, because larger sample sizes are needed to reach the same level of certainty in the conclusions. The security screening study used three sets of baggage images to avoid screeners recognising particular bags: these three sets were presented in a different random order to each participant. A potential alternative to randomisation for reducing order effects (learning and/or fatigue) is to counterbalance the order in which factor levels are presented to each participant. It works best when any change over time or order is linear, which is rarely the case. Other order-balancing designs, such as Latin Squares, are presented later under DoE.

There is another way to treat an independent variable, part way between treating it as a factor at several levels and fixing or randomising it. If we know of an independent variable that is likely to affect the outcome of the experiment, we can measure that variable and treat it formally in the design as a covariate rather than a factor. As an example, in studies of inspection tasks such as that studied by Drury and Sinclair (1983), much evidence has been found that the cognitive ability of inspectors helps determine their effectiveness (Drury et al., 2009). One good measure is that of field independence, measured by an embedded figures test, first developed in the 1950s (Witkin, 1950; Jackson, 1956). Using the individual scores of field independence as a covariate to account for the effects of different cognitive abilities when different groups of people are tested for different conditions removes some of the random error variance, thus increasing the power* of the study.

The E/HF professional designing a study must trade off the increased power coming from the reduced variance due to fixing a variable at a single level against the decreased applicability of the results beyond that single level. There is no rule for this; it depends upon the aims of the study. For example we typically sample participants from some population of interest (e.g. supermarket workers, retired women or current air traffic controllers). The list of all such potential participants is called the sampling frame. There are correct statistical procedures both for sampling safely and for treating the results so that we can generalise to the whole population of interest. The basic choice is between systematic sampling and random sampling. We start with the sampling frame and use either a systematic method (e.g. every fifth person) or a random method (e.g. choosing via random number table or programs) to determine the members of our sample (e.g. Section 3.3.3.2 of the *NIST Statistics Handbook* http://www.itl.nist.gov/div898/handbook/). In general, randomisation is statistically safer, although stratified sampling by randomly choosing from different defined strata of the sampling frame can be useful where more specific results are required. Occasionally, we are only interested in the specific participants who have taken part, particularly where there is a very small population from which to choose (e.g. active cosmonauts). This was the case in the fork-lift truck control study referenced earlier (Drury and Dawson, 1974) where only four drivers had been trained on all of the trucks tested and no more training was planned. In this case, the four participants were technically a fixed factor (at four levels of course) rather than the usual random factor. This changed the statistical treatment later in the analysis.

We have laid out some of the statistical and design issues in choosing factors and levels. However, the main considerations are more practical (i.e. design or resource driven) than statistical. DoE texts (e.g. Winer, 2012) or online expositions (e.g. *NIST's Statistics Handbook* http://www.itl.nist.gov/div898/handbook/) give general guidance on choosing factors, but cannot be specific when aimed at a diverse audience. Much traditional teaching of DoE uses data-driven insight, as if the process were truly a black box that needed to be understood from existing data or preliminary studies. Thus, exploratory data analysis is seen as a first step in process understanding. In E/HF we can provide more specific guidance. The key is E/HF insight, that is what is important in a situation or process. There is no substitute for E/HF knowledge, whether from textbooks, journal papers or applying E/HF insight into a specific process through direct observation. E/HF has over half a century of accumulated knowledge, on top of hundreds of years of insight from component disciplines such as psychology, physiology or occupational epidemiology. In DoE, we ignore this knowledge at our peril.

E/HF insight means understanding which variables affect individuals' and systems' behaviour, health and performance, and having some feel for the relative magnitude of the effects of these factors. Some factors are obvious, even to lay persons: there will be differences between individual participants in the experiment, there will be learning effects when a task is relatively new to the participant, larger participants will often be stronger, and so on. Some are more obvious to an E/HF practitioner: Task overload will often lead to degraded performance, non-neutral postures held for long periods will result in discomfort, performance may be worse on night shifts or with insufficient sleep.

Within this, some variables are almost mandatory and unavoidable. For example individual differences between participants or temporal changes within the study (learning, fatigue) could affect behaviour, health or performance. These have been discussed earlier, but we return to them later under DoE. Unless we deal actively with such variables, our conclusions will potentially be flawed. We cannot just ignore such factors and hope they will not affect our results. Study sponsors and journal editors will both catch such design problems, at the point where it is too late to deal with them. Other factors of importance are dependent on the particular study. Physical ergonomics studies would

---

* Power denotes the extent to which an experimental design is able to detect an effect that does indeed exist. A study with a power of 0.8 has an 80% chance of detecting the effect, if there is actually such an effect.

suggest that age, body size, strength and endurance are individual factors that could be important when assessing health (or physiological effects) and physical performance, and probably need to be dealt with by one of the techniques discussed earlier for reducing unwanted effects of individual differences. But other factors such as temperature, humidity and altitude would potentially be important too, as would workplace layout and task pace. We would expect some, but not huge, effects of cognitive variables on physiological effects and physical task performance, and no effects of such outlandish factors such as eye colour or moon phase. Similarly, cognitive ergonomics would suggest that age, cognitive style and task training would be very important to assess behaviour, health and performance during cognitive tasks, as would display layout strategies and task pacing. Physical factors would have less importance, but certainly measurable effects (see review in Drury et al., 2008).

In teaching experimental design, we have often found that the choice of sensible factors and number of levels is the most difficult aspect to communicate. Perhaps this is because traditional experimental design textbooks tend to ignore it, leaving more detailed knowledge to domain expertise. But the statistical and domain-specific aspects of DoE are rarely brought together in domain teaching (e.g. in E/HF subjects).

EFFECTS

The basic classification of E/HF-relevant effects comes from the definition of Ergonomics provided by the *International Ergonomics Association* (IEA) as:

> Ergonomics (or human factors) is the scientific discipline concerned with the understanding of the interactions among humans and other elements of a system, and the profession that applies theoretical principles, data and methods to design in order to optimize human well-being and overall system performance… (http://www.iea.cc/)

Clearly what E/HF professionals are ‘about’ are the twin groups of measures: system performance and human well-being. Just as in a road test of a car, there is no use knowing its speed and handling characteristics without also understanding its fuel consumption and reliability; in E/HF, we typically need to measure both the system performance and the cost to the human of achieving that performance.

‘Performance’ measures can involve the overall system, but are also possible for sub-systems even down to the individual human or group of humans. They can be simply classified into measures of speed and accuracy, or ‘time and errors’, although it is possible to argue that sub-standard speed performance is in itself an error so that we really only need to measure errors (e.g. Drury, 1994a). Speed measures are any measures with time in the numerator or denominator. In the numerator, they can be performance times, cycle times, reaction times at the individual, sub-system or system level, or even broader measures such as system down-time or systems availability. In the denominator, they are speed measures rather than time measures (e.g. output per shift, bits-per-second, rate of progress or miles per hour, as in driving). Accuracy is the positive aspect of performance (e.g. number of hits on target, quality level of a production process, percentage of driving task spent within the desired roadway). The negative aspect is errors (e.g. number of Methicillin-Resistant Staphylococcus Aureus [MRSA] infections per month at a hospital, fraction defective in a production process, percentage of missed threats and false alarms in a security process). Errors can either have a time-rate aspect (infections per month) or an event-rate aspect (fraction defective). If the time or event horizon is pre-specified and constant, perhaps the duration of the study, then the raw numbers of errors can be counted. As with the complementarity of performance and well-being measures, so speed and accuracy often need to be measured together. How can we praise the speed of a Formula 1 driver who has frequent crashes? How can we praise the output quality of a process operator who misses all deadlines?

Speed and accuracy do trade-off (although they may appear not to in current quality literature: Drury, 1997), often enough that many have studied the ‘speed-accuracy trade-off’ or SATO (e.g. Drury, 1994b). This is also known by the more accurate descriptor speed-accuracy operating characteristic [SAOC] (e.g. Pew, 1969), because the plot of accuracy versus speed is in the form of a statistical operating characteristic (OC) curve. Typically, more accuracy demands less speed (e.g. in Fitts' Law; Hoffmann, 1992), or in visual search tasks (Drury and Forsman, 1996). Also different errors can trade off against each other, the most obvious example being correct detections (hits) against false alarms in a signal detection task (e.g. Fisher et al., 2012). Thus, choosing measures of performance needs good E/HF models of how the suite of performance measures might co-vary. It is assumed that there will be a suite of measures to capture both system and human performance (see Chapter 1 for the range of measures that might be selected).

Well-being measures can range from the 'soft', such as discomfort ratings, to the dramatic, such as fatalities. They tell of the positive and negative effects of the system, and the human role in the system, effects on the humans within the system, or even those whom the system impacts in use. Typical measures cover workload (e.g. NASA Task Load Index, TLX, scale), internal state of the operator (e.g. discomfort, fatigue), physiological/biomechanical stress, freedom from long-term diseases, health status, negative incidents (e.g. near misses), actual accidents and injury/equipment damage/fatalities. Most of these are covered elsewhere in this book (see Chapters 15, 18, 20 and 31 for examples).

The measures of effects must themselves measure up to scientific adequacy. No measure is useful unless it can be measured with sufficient reliability and actually represents a parameter of interest to those commissioning or reading the study (refer to Chapter 1 for a more in-depth discussion of reliability and other considerations when selecting methods). As with much of what is presented in this chapter, initial resolution of these issues with those who commission the study can save much grief when the study findings are finally presented.

HYPOTHESES

In designing an experiment, or other E/HF study, a critical challenge is turning a research hypothesis into one or several statistical hypotheses to guide future actions. With the revolution in statistical DoE and statistical testing came a profound change in how we interpret results. The philosophy is that we pre-specify which outcomes of the study will lead to which conclusions. Thus, the E/HF professional could in theory give the designed experiment to a competent subordinate and be assured that whatever the data outcome, the conclusions will be exactly as planned. (This rather rigid statistical approach of course contrasts with more grounded theory or emergent approaches as are described in Chapters 3 and 5.) Such a rigorous scientific methodology is often followed until the last step of conclusions, where experimenters have been known to hedge their bets somewhat when faced with conclusions they do not like. For this reason, we spell out the transformation of a research hypothesis into a testable statistical hypothesis and provide an example of following through to conclusions.

The example is the study of security screening outlined earlier (Drury et al., 2008). This did not start out as a study of security screening, but as an examination of 'the inter-relationship between physical ergonomics and cognitive performance'. The experimenters were in fact a graduate class performing a rigorous E/HF practicum (Drury et al., 2007). The aim was to review the extensive literature on the interactions of physical and cognitive work, and test key aspects of this experimentally. (The idea of using security screeners did not arise until later, when it was found to be a convenient domain for testing the research hypotheses.)

The process involves several steps, adapted from the text by Siegel and Castellan (1988), which are illustrated here for the security screening study, concentrating on one of the hypotheses tested.

1. *State the research hypothesis*: Here the research hypothesis was that there is a measurable effect of the posture enforced by the physical workplace on performance in a cognitive task. That is not particularly new, as some authors have found such an effect in the past, but others have not (e.g. Mozrall and Drury, 1996).
2. *State the null and alternative hypotheses*: For a statistical test, we must turn the research hypothesis into a null hypothesis that states 'nothing was found'. The negation of this null

hypothesis is the alternative or experimental hypothesis. The data will eventually show which hypothesis we can conclude is true. The use of a null hypothesis is important as we can find the sampling distribution of any test statistic fairly easily when there is no effect. Here the null hypothesis was that there is no difference in any performance measures (hits, false alarms, time per bag screened) between the three workplace postures tested, that is 'there is no effect of the workplace postures on the performance measures'.

3. *Choose the statistical test*: In a multifactorial experiment such as the screener study (where the factors were posture and participant), the most logical test is the F-test for many cases when the effects are measured on interval or ratio scales. A test statistic (such as t or F) is a dimensionless quantity that is typically calculated as the ratio of the size of an effect to the appropriate variability. The test statistic increases in absolute magnitude as the size of the effect increases and as the variability of the effect within a condition decreases. Size of an effect is represented by the difference between two means (or the variance between several means for an F test) and is thus a fact, although subject to sampling error. The appropriate variability for a test statistic is the standard error of the difference between two means, which in simple cases is

$$\text{SE} = \frac{\text{Standard deviation}}{\sqrt{}\left(\text{Number of data points}\right)} = \frac{\text{SD}}{\sqrt{}\,\text{N}}$$

where SD is the standard deviation of a set of N data points. Thus the test statistic is

$$\text{Test statistic} = \frac{\text{M1} - \text{M2}}{\text{SE}} = \frac{\left(\text{M1} - \text{M2}\right)\sqrt{}\,\text{N}}{\text{SD}}$$

where M1 and M2 are the means of the conditions being compared. For any statistical hypothesis, there are typically alternative tests with different assumptions. Non-parametric alternatives, which make fewer assumptions about the distribution, type and source of the measurement values, are possible. However, these tend to only be suitable for simple designs involving only one or two independent variables (e.g. Siegel and Castellan, 1988).

4. *Find the sampling distribution of the test statistic under the null hypothesis*: The F statistic's distribution is well known, tabulated in almost any statistics or DoE text and an integral part of most statistical software available for DoE and analysis. It does, however, rely on the assumption of normally distributed data, homogeneity of variance, independence of sampling and use of interval or ratio data. The assumptions and how violations of the assumptions impact results are also described in most DoE texts.
5. *Select the level of significance*: The level of significance or 'p value' is our threshold for concluding that the alternative hypothesis is true. This defines the likelihood of concluding that an effect exists when it truly does not exist, an erroneous conclusion. In statistical texts, this called a Type I error, and it is important in experimental research to ensure that this likelihood is small. In research, we typically choose a level which would rarely be found by chance (e.g. 1 in 10, 1 in 20, 1 in 100 or 1 in 1000). In this way, we limit our false alarm rate *when many tests are performed.* We do not eliminate false alarms, just make them rarer. If a researcher publishes 100 studies in a career, each with a single test at 1 in 20 ($p = 0.05$), then about 5 false alarms would be expected. A couple of notes are in order here. First, there is no generally accepted p value – it depends upon the circumstances. A p value of 0.01 may be needed if the consequences of false alarm are very high. A p value as low as 0.1 may be acceptable when there is a limited population to collect data from, but a decision is still required. Second, in theory, the person who commissions the study should

choose the significance level. This is fine when the commissioner is a government, scientific or medical agency. However, most managers, most lawyers and many public servants are quite unused to the idea of probability and balk at choosing a level because of their lack of statistical knowledge. Here, E/HF researchers must help them reach an informed opinion, just as they must help a participant understand the risks involved in study participation. In industry, which should be quite used to probabilities after years of Six-Sigma programs, E/HF professionals still get asked ‘what is the right answer?’ when they attempt to involve managers in the choice of level of significance. Taking some responsibility to help our colleagues from other fields is in fact helping to fit the task to the decision maker. In the screener study, we chose a level of significance of $p = 0.05$, a very traditional value in science, and the normal minimum level of significance adopted in E/HF journal publications, because rejecting the null hypothesis with only a 5% chance of a ‘false alarm’ would convince others in our profession that we really had found an effect.

6. *Determine the region of rejection*: If we know the sampling distribution of the test statistic and the probability we will accept for a false alarm, then we can split those values of the test statistic into a set for which we will accept the null hypothesis and a set for which we will reject the null hypothesis. Thus, we have mapped all possible outcomes (values of F in our example) onto the conclusions we will draw before running the experiment. From here on, the process is purely mechanical. The region of rejection of the F test in the screener study was those values of F beyond which lay only the upper 5% of the probability distribution. The exact F value depended upon the number of degrees of freedom* in the numerator (2 in this example, because there were three levels of the posture factor) and the denominator, which could vary, because there was more than one effects variable, and each had a different number of measurements.
7. *Run the study and determine the value of the test statistic*: This is the most time-consuming step, but collecting the data and calculating the F value are issues covered later. Setting up the screener experiment and collecting the data took several weeks of effort by all six authors of that study.
8. *Determine whether or not the data support the alternative hypothesis based on the criteria established in Step 6*: In the screener example, the null hypothesis was accepted for all performance measures.

What we have done is to force ourselves to pre-select our ultimate actions based on the statistical outcome. As with our discussion of factors, levels and the qualities of measures chosen, this discipline forces us as experimenters to think about the process of experimentation through to the final conclusions *before* we begin even recruiting participants or building equipment. In this way, there should be fewer studies that fail to meet the expectations of either client or experimenter.

EXPERIMENTAL DESIGN ALTERNATIVES

The Basis of Factorial Experimentation and Analysis

Any standard DoE text (e.g. Winer, 2012) will provide literally hundreds of potential designs for statistical experiments, with detailed instructions on the choices available and the correct analysis techniques. Clearly, such a treatment is inappropriate here, so we shall concentrate on issues of most interest to the E/HF practitioner. Most texts start with simple comparisons of two levels of a single factor using a t-test or non-parametric equivalent. Because we are using people as participants in our experiments, and differences between participants are non-trivial, we are rarely able to use such basic designs. Unless we wish to run the study on a single participant, we always have one factor

---

* Degrees of freedom are related to the number of values that can vary in a calculation.

of participant with differences between participants as a factor in our analysis. Even a comparison of two levels of a factor of interest will result in a multifactorial experiment, so that is where we must start.

Underlying all multifactorial designs is the analysis concept of ANOVA, again covered in depth in DoE texts. This in turn rests on the mathematical property of a variance: the variance of the sum of, or difference between, two means is equal to the sum of the variances of the separate means:

$$\mathrm{Var}(x_a + x_b) = \mathrm{Var}(x_a) + \mathrm{Var}(x_b) \quad \text{and} \quad \mathrm{Var}(x_a - x_b) = \mathrm{Var}(x_a) + \mathrm{Var}(x_b)$$

This simple assertion (again, proved in most statistical tests) extends to a linear combination of variables so that the variance of their sum is the sum of their variances. Thus, if we run a two-factor model with Factor A at a number of levels denoted by [i] and Factor B at levels [j], we can say that any single data point at level i of Factor A and level j of Factor B is composed of:

The overall mean of the whole data set
+ the difference between the overall mean and the true effect of Factor A at level i
+ the difference between the overall mean and the true effect of Factor B at level j
+ the random error of the combination of Factor A at level i and Factor B at level j

$$X_{ij} = \mu + A_i + B_j + \varepsilon_{ij}$$

This is the structural model of the experiment. Taking variances, which are additive, and noting that the variance of a constant such as μ is zero, we have

$$\mathrm{Var}(X_{ij}) = \mathrm{Var}(A_i) + \mathrm{Var}(B_j) + \mathrm{Var}(\varepsilon_{ij}),$$

provided that the error variance is the same for all [i,j] combinations of Factor A and Factor B. In this way, we can take the overall variance of all of the data ($\mathrm{Var}(X_{ij})$) and split it into components due to Factor A ($\mathrm{Var}(A_i)$), Factor B ($\mathrm{Var}(B_j)$), and the error variance ($\mathrm{Var}(\varepsilon_{ij})$). We can thus see how much variability is uniquely associated with each factor, called here a 'source of variance'. We can also compare this variance with the error variance to form a test statistic, as in the previous section, because a test statistic is an effect size divided by its appropriate variance. We have just partitioned variance into its components, so it is unsurprising that the technique is called ANOVA. If we can do it for two factors, then we can do it for any number of factors, provided that their effects are additive. Note that this paragraph has two uses of 'provided that', implying that to use ANOVA, we must make assumptions about homogeneity of variance and additivity of factor effects. We shall return later to alternative ways to analyse the data if these assumptions are not met.

Multifactorial experiments combine the factors at each level, so that the total number of combinations (i.e. conditions) studied must be the product of the number of levels of each factor. In the security screening experiment, we collected data on the three posture levels for each of the 12 participants, giving 36 combinations. Note that we did not use each run order for each screener, but balanced the three sets of images across participants so that each participant saw a different image set in their three postures.

MAIN EFFECTS

For each experiment, there is a structural model, written as in the example earlier ($X_{ij} = \mu + A_i + B_j + \varepsilon_{ij}$). This example is a very simple model where the separate effects of A and B are strictly additive. That means that the effect of Factor A is the same at all levels of Factor B. This is called a 'main effect'. If we plotted a graph of our measure against the level of Factor A, then there would be parallel graphs for each level of Factor B.

INTERACTIONS

Experimental results are not usually limited to main effects. The effect of one factor may well depend on the level of another factor: the graphs need not be parallel. To take an example from physics, the combined gas law relates the pressure (P), the temperature (T) and the volume (V) of an ideal gas by

$$PV = kT$$

or

$$T = \frac{PV}{k}$$

Thus, the effects of P and V on T are not additive but multiplicative. If we plotted T against P at different values of V, we would get lines that were converging rather than parallel. This is a simple example where it is just a different operator relating the two factors P and V. In more complex situations, there may be different joint effects of the two factors. All non-additive combinations of factors are called interactions (which can be described as the effect of one independent variable on the *effect* of another independent variable on the dependent or measured variable), and are one of the main reasons for performing multifactorial experiments.

An example from the doors study mentioned at the beginning of the chapter is shown in Figure 2.1. It is obvious that the lines joining the data points at each level of participant stature are not parallel, although they do form a pattern that should be familiar to those with E/HF training. Compared to shorter people, taller (and presumably stronger) individuals do not need to use their body weight to open doors until a much higher level of door restoring torque.

In the design and analysis of multifactorial experiments, interactions can be treated statistically by including an extra term in the structural model to represent the combined effect of

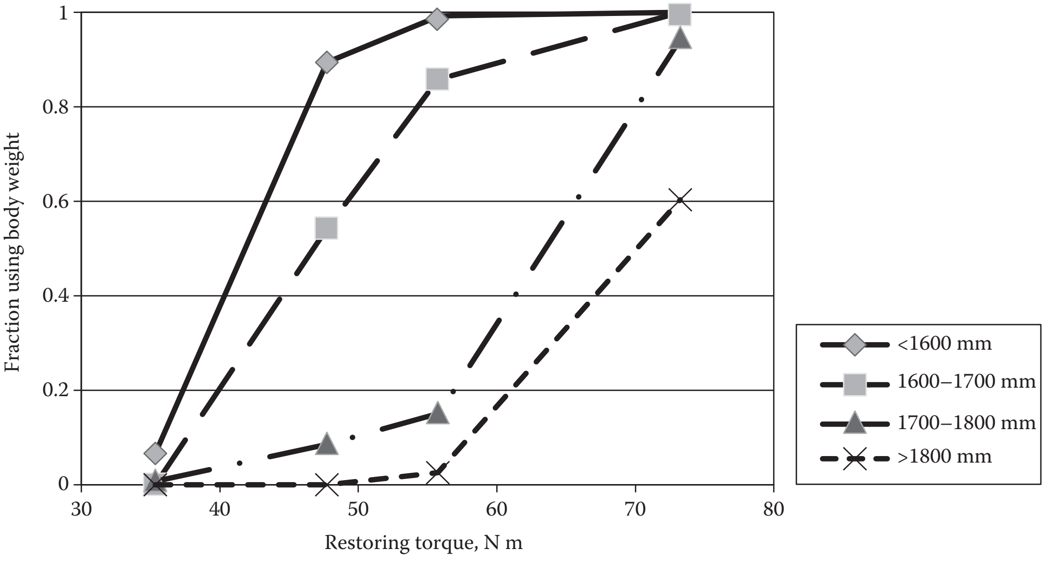

FIGURE 2.1 Example of an interaction from Chang and Drury (2007). The lines reflect four groups of people having different stature.

two factors that is not predictable from their individual additive effects. This is the term AB_{ij} in the following equations:

$$X_{ij} = \mu + A_i + B_j + AB_{ij} + \varepsilon_{ij}$$

$$\mathrm{Var}(X_{ij}) = \mathrm{Var}(A_i) + \mathrm{Var}(B_j) + \mathrm{Var}(AB_{ij}) + \mathrm{Var}(\varepsilon_{ij})$$

Note that the structural model is still additive, so the variances are also still additive. The additional interaction term can now be tested for significance like all of the other terms against its appropriate error variance. Note also that we need some independent measure of the error variance. This is typically accomplished by repeating measurements under nominally identical conditions. In the doors study, multiple people were observed in each of the 16 combinations, but because the measure was only use/non-use of body weight (a nominal measure), it could not be easily used to estimate error variance. In the case of all three example experiments in this chapter, other assumptions were made in order to obtain estimates of error variance. The safest way is always direct replication, but this may not be possible for all experiments. For example, in the weather displays study also mentioned earlier in this chapter, the same weather front conditions could not be replicated so that only a single measure was possible for each combination of factor levels.

We can obviously extend the multifactorial idea to experiments with more than two factors. With each additional factor, we can find more interactions. For two factors, we can measure the effects of

$$A, B, A \times B.$$

When we add a third factor C, this becomes

$$A, B, A \times B, C, A \times C, B \times C, A \times B \times C.$$

Interactions are of great importance in E/HF, often because they represent multiple sources of stress on the human in a system, and each additional source may cause a more-than-additive effect as the limits of human capability are reached. Figure 2.1 is an example of this: at the lowest restoring torque, door opening is easy for all of the population, but as restoring torque increases, the impact is greater among the less strong members. Interactions also have importance in the theoretical underpinnings of our discipline. For example, in time-sharing between two tasks, the lack of an interaction implies that the two tasks must be processed serially, while the existence of an interaction implies that some parallel processing is possible (Wickens and Carswell, 2012). Also in visual search tasks, the rate of increase of reaction time with increasing background complexity depends upon the difficulty of discriminating a target from its background. Thus, if target/background discriminability is one factor and number of non-targets in the search field is another, there will be an interaction between these factors on reaction time (Treisman, 1986). In the extreme, if one target is pre-attentive, then the number of non-targets will have no effect on reaction time for that target, leading to an extreme interaction effect. Finding interactions, and finding situations where there are no interactions, are primary objectives of practical E/HF and can form the basis for DoE principles.

DESIGN OF MULTIFACTORIAL EXPERIMENTS

The basic design of a multifactorial experiment, known as a complete factorial, is to test all combinations of levels of all factors. We just add factors such as D, E, F, etc. beyond the aforementioned A, B and C. This was used in the doors study, where 4 levels of restoring torque were combined with 4 levels of participant stature to give the 16 conditions plotted in Figure 2.1. Also, in the weather display study, 6 pilots were combined with 6 weather front distance and 3 displays to give 108 conditions. In contrast, the security screening study did not use all combinations of 3 postures, 12 participants, 3 run orders and 3 image sets, using only a single image set for each run order. This gave a design comprising 3 postures × 12 participants and so could only find main effects of posture, participant and run order.

TABLE 2.1
Design Tableau of a 3 × 3 × 2 Complete Crossed Factorial Experiment with Three Replications

Factor A	Factor B	Participant	Replication 1	Replication 2	Replication 3
Level 1	Level 1	P1			
Level 1	Level 2	P1			
Level 1	Level 1	P2			
Level 1	Level 2	P2			
Level 1	Level 1	P3			
Level 1	Level 2	P3			
Level 2	Level 1	P1			
Level 2	Level 2	P1			
Level 2	Level 1	P2			
Level 2	Level 2	P2			
Level 2	Level 1	P3			
Level 2	Level 2	P3			
Level 3	Level 1	P1			
Level 3	Level 2	P1			
Level 3	Level 1	P2			
Level 3	Level 2	P2			
Level 3	Level 1	P3			
Level 3	Level 2	P3			

In E/HF, the factor of participant can require special treatment. An example of a complete factorial experiment, with 3 levels of Factor A, 2 of Factor B, 3 of participant and 3 replications (i.e. repetitions) of each combination, is given in Table 2.1. This of course requires a minimum of 3 × 3 × 2 × 3 = 54 measurements.

This design can be used to calculate the variance component of all three main effects (A, B, P), all two-way interactions (A × B, A × P, B × P) and the three-way interaction (A × B × P) as well as a true error variance. Note that each participant is tested at all three levels of Factor A and both levels of Factor B. This is known in E/HF as a 'within participants' (also known as repeated measures) design, as both factors are tested on the same participants (P1, P2 and P3). It is clearly a 'good' design if there are large inter-participant differences as the ANOVA can calculate the effects of A, B and A × B independent of participant. In other words, differences between participants do not contribute to the error variance associated with A, B and A × B effects. But what if we cannot test each participant more than once? An example is in comparison of learning technologies where a participant can only learn the task once.

So far, we have only considered designs where all combinations of levels were tested, known in DoE as crossed designs. There is another class of designs for complete factorial experiments known as nested designs, or in E/HF as 'between participants' (or independent samples) designs. In these designs, one factor (participant) is nested under other factors so that different participants are tested under different conditions. Table 2.2 shows a design equivalent to Table 2.1, but with participant nested under both Factor A and Factor B so that the 18 different combinations of A × B are tested with different participants rather than re-using the original 3 participants of the crossed design in Table 2.1. Such designs have the advantage of minimising confounding due to fatigue or learning (order effects), but differences between participants groups contribute to the error variance, making it more difficult to identify A, B and A × B effects.

WITHIN-PARTICIPANT VERSUS BETWEEN-PARTICIPANT DESIGNS

The main criterion for choice between these two design structures is whether or not the participants change during the course of the experiment. Clearly a participant's strength will not change much over

TABLE 2.2
Design Tableau of a 3 × 3 × 2 Complete Nested Factorial with Three Replications

Factor A	Factor B	Participant	Replication 1	Replication 2	Replication 3
Level 1	Level 1	P1			
Level 1	Level 2	P2			
Level 1	Level 1	P3			
Level 1	Level 2	P4			
Level 1	Level 1	P5			
Level 1	Level 2	P6			
Level 2	Level 1	P7			
Level 2	Level 2	P8			
Level 2	Level 1	P9			
Level 2	Level 2	P10			
Level 2	Level 1	P11			
Level 2	Level 2	P12			
Level 3	Level 1	P13			
Level 3	Level 2	P14			
Level 3	Level 1	P15			
Level 3	Level 2	P16			
Level 3	Level 1	P17			
Level 3	Level 2	P18			

any experiment of reasonable duration, while the same participant's task knowledge and skill will surely change unless very experienced participants are chosen and tested using familiar conditions. Thus, in the doors study, we could have re-used participants, although we did not. Also in the weather display study, highly experienced pilots were used, so that changes between multiple trials on how to react to weather fronts would be unlikely to change, so again, the same participants can be re-used in a crossed design. However, in the security screening study, we found a definite learning across trials on the image sets used, so that even our experienced screeners did change over time. We used a partially within-participants design, but were able to measure and remove any change (learning) effects from our comparison of posture effects. To give other examples where different designs were used, consider the following

> Laughery and Drury (1979) used a between-participants design in a study of optimisation skills because it was suspected that techniques learned during the solution of one type of optimisation problem might transfer in an inconsistent manner to other problems, with an adverse effect on bias and variability. Thus five participants were used in each condition, which meant that any comparison between conditions had to be made against between-participant variability. The groups were kept reasonably homogeneous (engineering students) but this in turn limits the generalisability of the results.
>
> Drury et al. (1989a,b) studied the biomechanics and physiology of handle positions on boxes used ten participants, each performing a box holding task using ten handle positions. The within-participants design eliminated the influence of individual differences on the effects, allowing the effects of handle positions on boxes to be detected despite the limited sample size.

No changes to the participant were expected during the box holding experiment, but changes were expected in adaptation. Change occurs in humans in the short term as they fatigue and in the long term as they adapt or learn. With appropriate rest periods, no fatigue was expected (or found) in the box holding task and certainly an hour or two of experimentation on a well-practiced task is unlikely to change either a participant's body strength (adaptation) or box holding technique (learning). Hence, a biomechanical and physiologically limited task is unlikely to exhibit what Poulton's famous (1974) paper called asymmetrical transfer effects. The same cannot be said for most intellectual skills. What you learn in first solving one optimisation problem is quite likely to affect your performance in solving the next. The transfer can be positive, if the same solution techniques are useful in both problems, or negative, if the solution to the first problem is inappropriate in solving the second. An optimisation task is a priori likely to be closer to an intellectual task than to a biomechanical one, hence the choice of a between-participants design. Any human functions, even anatomical ones, will adapt or change given sufficient time, but the key question is not whether or not change will occur but whether enough will occur to bias the experimental comparison. We can minimise change during an experiment by choosing participants already highly skilled, but as noted earlier that worked for pilots but not for security screeners. We can also provide extensive training in the task so that the typical negative exponential or fractional power law learning curve reaches enough of an asymptote to prevent further changes in task performance. Such techniques would allow the greater power of a within-participants experiment (less variance in comparisons). Extensive task training also helps when the participant pool is limited (astronauts) or non-existent (operators of an entirely novel system). Finally, we may be interested in the response of each individual participant rather than the overall distribution of performance, so that a within-participant design must be used.

There is no reason that an experiment must be entirely a crossed design or entirely a nested design. We can have useful designs that are partially crossed and partially nested, called mixed model designs. However, wherever different conditions are tested on the same participant in a factorial study, we must use a form of analysis called repeated measures ANOVA to correctly capture the contributions of the independent variables and interactions on the variability of the measurements.

The final word on between- vs. within-participant designs is that a between-participant design is always the safer alternative, but may not be practical within resource constraints. They are also subject to the risk that differences between individuals may mask influences of factors, particularly when the number of subjects within each condition is small. Within-participant designs need steps to ensure absence of carry-over effects and a different type of ANOVA but, if designed carefully, help to manage the potentially confounding effects of individual differences on the measurements.

SAMPLE SIZE, EFFECT SIZE AND POWER

We have seen that the significance level chosen for any statistical test determines the probability of the test giving a false alarm, that is concluding that an effect exists when it truly does not (Type I error). As any E/HF professional can guess, it is not possible to discuss false alarms without also discussing the complementary error, that of failing to conclude that an effect exists when it truly does (Type II error). We have two hypotheses, null and alternative, so that if

$$\text{Significance level} = \text{p(conclude Alternative} \mid \text{Null is true)},$$

then its complement, power, is defined as

$$\text{Power} = \text{p(conclude Alternative} \mid \text{Alternative is true)}.$$

Whereas we usually look for a very *low* significance level (0.10, 0.05, 0.01, 0.001), we would like a *high* value of power (0.90, 0.95 etc.).

There are four inter-related factors that need to be considered: significance level, power, effect size and sample size. The effect size is the magnitude of the difference between two means, or the variance between several means. The larger the effect size we are looking for, the easier the statistical testing is, so that for a large effect size, we can have both a low significance level and a high power. It is possible to manipulate the anticipated effect size with DoE. If, for example, age is related to a particular performance measure, a more powerful experimental design would be to compare groups that are dramatically different in terms of age, as opposed to groups that are closer in age. Because a test statistic is the ratio of an effect size to its standard error, we can also reduce the size of the standard error for any given effect size by taking more samples. Recall that the standard error is calculated by dividing by the square root of the sample size, and it becomes obvious that increasing the sample size is quite an inefficient way to increase power.

In designing our experiment, we have so far not mentioned sample size, except when we have given examples of numbers of participants or numbers of replications per condition. But both of these experimental parameters must be chosen before we can proceed, and we can only logically do this on the basis of the other three underlying variables. If we need to know sample size, we must first decide on effect size, significance level and power of the test. The usual statistical textbook advice is to work with the study commissioner to find values of these parameters, but this is more taxing than asking a client for a significance level alone. One of the authors recently had to produce tableaus of effect sizes and sample sizes for an aviation security study so that the client could make better informed decisions about how to set up an experiment. Such an approach is facilitated by web-based software (e.g. http://homepage.stat.uiowa.edu/~rlenth/Power/) and advocated in related journal publications (Lenth, 2001).

The mechanics of calculating sample size given the other three parameters are not simple beyond the 'toy' designs of comparing two samples with no other sources of variation, but help is provided in many statistical packages. In addition to the web-based application noted earlier, one author has used the PASS (power analysis and sample size) software to determine sample sizes for experimental designs to be analysed by ANOVA. A much-used package (MINITAB) performs power/sample size calculations for many ANOVAs and some non-parametric statistical tests (e.g. chi-square) but not for complex mixed models. But in the airport security study described at the beginning of this chapter, only nominal data could be collected (e.g. number of threats found, number of false alarms) so that contingency tables and the chi-square test took the place of ANOVAs. The sample sizes were computed manually, although it was later found that they were part of the MINITAB statistical package.

FRACTIONAL FACTORIAL DESIGNS

The complete factorial design is a powerful and often-used tool in E/HF, but also a costly one. It can find important interactions, but if interactions are known (or assumed) not to exist, it is wasteful of resources. There are special designs that trade off knowledge of interactions for reduced experiment size. We have already seen this in the security screening study where the information on the posture × participant interaction was sacrificed to allow a study that would fit the available resources. More formal methods are available, known as fractional factorial designs. As their name implies, they only test a fraction of the combinations used in a complete factorial design.

The simplest fractional factorial for E/HF is probably the Latin Square design. It uses three factors, all at the same number of levels, and counterbalances their appearance so that only n^2 instead of n^3 combinations are tested. The Latin Square ensures that each combination of the levels of each factor occurs once and only once in the $n \times n$ tableau of the design. Clearly something must be lost for such a large saving in effort and indeed it is. All interactions are sacrificed so that only the three main effects can be calculated and the error variance term is merely the left-over variance when the three main effects are calculated. Even the term 'sacrificed' does not capture the whole loss: any interactions are confounded in a complex way with the main effects. A Latin Square should only be used where no interactions are expected or the interactions are known from prior research not to exist. This makes a long list of assumptions that should really be attached to the experimental conclusions, although they rarely are. The main use of Latin Squares in E/HF is in presentation order of different conditions to participants in a within-participant design. Thus, if we have six levels of the Factor (A, B, C, D, E, F) and six participants (1, 2, 3, 4, 5, 6), we can present them in the following trial order (Table 2.3):

In this Latin Square, each level of the factor (usually called a treatment) follows a different order for each participant, helping minimise the effects of unwanted transfer between treatments.

TABLE 2.3
Example of a 6 × 6 Latin Square Design to Eliminate the Potentially Confounding Effects of Trial Order

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
Participant 1	A	B	F	C	E	D
Participant 2	B	C	A	D	F	E
Participant 3	C	D	B	E	A	F
Participant 4	D	E	C	F	B	A
Participant 5	E	F	D	A	C	B
Participant 6	F	A	E	B	D	C

(The same pattern can be used for any even number of treatments and participants.) This use of a Latin Square is an excellent alternative to randomisation of treatment order across participants.

More complex fractional factorials have been advocated and used in industrial experimentation, often as part of a design with all factors at 2 levels, called 2^n designs. Like Latin Squares, they only require a fraction of the conditions to be tested, and similarly, they do not allow all interactions to be calculated. They are typically advocated in studying the response of a multivariate industrial process to a selection of the postulated variables in a very economical manner so that the most important variables can be determined efficiently. So far they have not been widely used in E/HF, although a few examples exist (e.g. Bishu et al., 1992; Lin and Radwin, 1998; Naugraiya and Drury, 2009). Rather than 2^n designs, these designs are called 2^{n-k} designs, where n factors are tested within the resources of an (n − k) design. The assumption behind fractional factorial designs of all types (see, e.g. Taguchi, 1965) is that higher-order interactions are inherently unlikely. Thus, we can confound these interactions with each other and not calculate them. Each fractional factorial design has a design operator, which is an identity equation showing which effects are confounded with which other effects. Typically, we use a design operator that confounds the main effects and lower-order interactions of interest with (unlikely) higher-order interactions.

The Naugraiya and Drury (2009) experiment was intended as a screening experiment to examine the significance of a large number of factors and interactions for a simulated process control task. It used a 2^{6-1} fractional factorial and examined six factors, each having two levels, with 32 cells rather than the 64 cells required for a full 2^6 factorial. Participants were assigned randomly with one to each of the 32 unique conditions tested, a most unusual procedure in E/HF experimental design, but one quite common in industrial experimentation. Each participant performed the simulation four times, ‘producing’ and ‘shipping’ 200 industrial parts under different quality challenges.

The six factors were

Operator expertise (E)
Operator training (T)
Process capability (Cp)
Challenge direction (D)
Challenge amount (A)
Cost criterion (C)

The design operator for the experiment was 1 = ETCpDAC. This ensures that main effects and low-order interactions are only confounded with the more unlikely higher-order interactions:

Main effects confounded with five-way interactions
Two-way interactions confounded with four-way interactions
Three-way interactions confounded with three-way interactions

Thus, we assume that main effects and two-way interactions are in practice unconfounded. We could try to estimate three-way interactions, but each is confounded with the interaction between the remaining factors (e.g. E × T × Cp is confounded with D × A × C) so that we cannot disentangle their separate effects.

With six main effects and 15 two-way interactions, much was learned about the task. Note however that the unusual experimental design had only a single degree of freedom for each effect tested, with 103 degrees of freedom for the error variance. Note also that there could be no 'between participants' effect calculated with a single participant per condition.

SEQUENTIAL EXPERIMENTATION

There is an alternative to fractional factorial experiments that can be useful in E/HF where a sequence of studies is performed instead of a single study. It has the potential to measure which interactions are important rather than merely assuming them away. Also it can deal with factors at more than the two levels assumed in most DoE texts for factorial experiments. With only two levels tested for each factor, we cannot find out much about the underlying response surface (i.e. the shape of the relationships between the independent variables and the effects). The strategy is to perform a 2^n complete factorial, measure which interactions are important, then perform a set of experiments with the desired levels of each factor, but only for factor combinations that have measurable interactions. As an example, suppose the design we would like to run has five factors at the following levels:

A at 5 levels
B at 2 levels
C at 3 levels
D at 3 levels
Replications at 2 levels to provide an error estimate

The full factorial will need 5 × 2 × 3 × 3 × 2 = 180 trials. It will allow calculation of

4 main effects: A, B, C, D
6 two-way interactions: AB, AC, AD, BC, BD, CD
4 three-way interactions: ABC, ABD, ACD, BCD
1 four-way interaction: ABCD
1 error term

Perhaps, we do not need all of these from a single grand design. We can use a 2^4 complete factorial with 2 replications as a screening experiment specifically to test for interactions. This will require 2 × 2 × 2 × 2 × 2 = 32 trials. Then run experiments on the significant interactions. For example if only A × B and C × D are significant, then we can run two additional smaller experiments:

A × B with 5 levels of A, 2 levels of B and 2 replications, requiring 5 × 2 × 2 = 20 trails
C × D with 3 levels of C, 3 levels of D and 2 replications, requiring 3 × 3 × 2 = 12 trails

We can thus measure all of the effects we were initially interested in with 32 + 20 + 12 = 70 trails instead of the original 180 trials. Of course, not all screening experiments produce the same interaction structure, so we could go all the way from every interaction being significant (requiring 180 trials) to no interactions being significant (requiring only (5 + 2 + 3 + 3) × 2 replications = 26 trails). We can always re-use data between experiments if that is logically possible, for example the 10 combinations in the A × B experiment include 2 × 2 = 4 conditions that have already been studied, leaving only 6 additional conditions with 2 replications each.

ANALYSIS ALTERNATIVES

Throughout this chapter, we have assumed that ANOVA will be the analysis method, primarily because it is well suited to multifactorial experimentation, and almost any E/HF study will have to be multifactorial. ANOVA uses the additive property of variances to decompose a total experimental variance into components associated with each variable and interaction. In simple cases, such as complete factorial designs with replications, we can have independent tests of each factor and interaction. With more complex, but less complete designs, such as fractional factorials or Latin Squares, we forgo some independence for experimental convenience or even for study feasibility. The familiar ANOVA table can be found in any DoE text and will not be repeated here. There are strict rules for how to compute the significance of effects based on their F statistic value, depending upon the structural model of the experiment. This tells the components of variance and thus what denominator to use in the F test.

Most E/HF professionals will not calculate variance components and F-values by hand, relying on statistical packages to perform the computations. These, such as SPSS (Statistical Package for the Social Sciences) or MINITAB, will require the user to input a structural model to control the computations. That is why understanding ANOVA and models remains important in times of automated computation. Failing to treat repeated measures correctly, confusing fixed and random effects or failing to check the ANOVA assumptions are all analysis errors that can be committed by experienced E/HF professionals. The use of packages where ease of use is a design feature should not reduce the diligence of the experimenter at the analysis phase. In this section, we enumerate various issues with analysis, examining what to do when there are multiple dependent variables, how to test assumptions (and what to do if they are not met), and sources for the many different statistical analysis packages available today.

DEALING WITH DATA

We have discussed experimental design and ANOVA as if there were a single number or measure in each cell of the design, but we now need to expand beyond this. Each cell is the data from a single replication of each combination of factor levels in the design. Assuming that issues of reliability, validity, etc. have been addressed, the first thing to note is that we do *not* put the data into a spreadsheet and calculate means. Each data point must be kept separate: ANOVA procedures will provide table of means, variances and confidence intervals *ad nauseam*. Keeping data separate provides the full degrees of freedom for error variance, helping to ensure that the power of the tests is maintained. Second, many measures come in the form of a continuous variable recorded over a time interval, such as the record of car position in a lane while driving, or the continuous movement of the centre of gravity of a standing operator. In these cases, measures must be derived from the continuous records, for example root-mean-square error or average position in each dimension. It is this number which becomes the data for the ANOVA. These are quite simple matters to deal with and have usually been addressed during the initial design of the experimental study.

Quite often, however, there will be multiple measures per cell of different aspects of performance or well-being or both (i.e. multiple effects). In the security screening study, we measured four performance variables (hits, false alarms, time per image when a threat was found, time per image where no threat was found) and four well-being variables (two measures of body part discomfort, NASA TLX, number of non-work-related movements), giving 8 dependent variables. The obvious way to proceed is to perform an ANOVA on each dependent variable, but this raises two problems. First, because we are performing 8 tests for each factor or interaction, the likelihood of concluding that a statistically significant relationship between one or more independent variables and an effect variable exists when in fact it does not exist (i.e. the likelihood of Type 1 error) increases. Another way to think about this is that the significance levels become inflated. If we chose a 1 in 20 chance of false alarm for any single variable, then the likelihood of having at least one false alarm would be $(1 - (1 - 0.05)^8) = 0.33$, which is not what we had planned. Second, it might be that the dependent variables show similar patterns, so that even if one variable is not significant, the same pattern across several variables may be significant. The standard way to proceed is to perform a multivariate analysis of variance, or MANOVA, across the complete suite of dependent variables. Then, if any factor or interaction proves significant, univariate ANOVAs can be run to determine which variables were responsible. This procedure does not inflate the significance levels, and provides an orderly exploration of the data.

Another technique that can be a powerful tool for reducing a large suite of data to a more manageable number of orthogonal and hence independent tests is factor analysis. Factor analysis (nothing to do with factors in DoE) groups together dependent variables with high inter-correlations. With a modification called Varimax rotation, it will produce a small number of new dependent variables called, confusingly, factors that are orthogonal to each other and summarise a large fraction of the total variability in the data set. It has a long history in the social sciences, and has been used many times in E/HF, from early papers (e.g. Drury and Daniels, 1980) to more recent studies (Ryan et al., 2009). In the security screening study, it was used to explore the inter-correlation matrix of all 8 dependent variables. We found three factors that met the usual criterion for significance: performance (4 variables), posture (3 variables) and workload (TLX only). Each could be analysed in the confidence that only three independent tests were being carried out, and that all of these new independent variables were orthogonal. To simplify the interpretation of the findings, the ANOVA results associated with each of the individual measures that contributed to each of the orthogonal factors were reported in the paper.

DEALING WITH ASSUMPTIONS

Various assumptions have been made in this treatment of experimental design, assumptions that can often be tested directly from the data collected. We have already considered the additivity assumption and shown how intersection terms can extend the simple additive model of main effects to the generally more interesting interaction effects. There are other ways to deal with non-additivity in special cases. The combined gas law used as an example earlier has a multiplicative effect of P and V on T. Many human functions are multiplicative, particular sensory functions. In any E/HF study where effects look multiplicative, logarithms are a useful transformation tool to allow ANOVA while still preserving the model structure. It is simple to transform the common gas law into an additive function by taking logarithms of the equation:

$$\mathrm{Ln}(T) = \mathrm{Ln}(p) + \mathrm{Ln}(V) - \mathrm{Ln}(k)$$

Additivity is now satisfied. Transforms are an integral part of many science and engineering formulations, and are frequently used in measuring human performance. Examples are the use of the decibel scale in auditory perception and the same scale to study human tracking behaviour.

Transforms can also help with the ANOVA normality assumption. We can test this assumption by having the analysis package plot residuals (the difference between a data point and its expected value from the ANOVA model) as a cumulative normal distribution. Either visually or statistically, we can determine whether the normal distribution is a good fit to the data. If it is, then the ANOVA is valid for the normality assumption: if not, the pattern of deviations from the normal distribution function provides clues to suitable transforms.

Some measures of human performance and well-being are quite normally distributed, but task completion times are often not. Most performance time data will have a lower bound beyond which the human cannot react or move any more rapidly. But the upper bound is often unlimited, leading to time distributions that are positively skewed, with a longer ‘tail’ to the right. A lognormal distribution is often a good fit, so that transforming the raw time data to Ln(time) will produce normally distributed data suitable for use in ANOVA. In odd cases, even more skewness is expected: Search times in extended search tasks (such as security screening) are expected theoretically to follow a negative exponential distribution (Morawski et al., 1980). A ln(time) transform is usually sufficient to normalise the data, however. Most statistical texts include at least something on transformation to improve normality and homogeneity of variance or to meet additivity assumptions, removing interactions (e.g. Winer et al., 1991, pp. 354–358).

Probability or frequency data generated from repeated measures made on nominal or categorical scales, particularly the (0,1) form of data, are inherently non-normal. There is much discussion in the statistical and social science literature of the legitimacy of using ANOVA for categorical data, with some claiming it can be used with minimum danger, while others recommend arc-sine or logistic transforms before using ANOVA. Rather than take sides in this, articles such as Jeager (2008) should be consulted for the most recent findings.

This introduction of nominal data brings in the whole question of alternative forms of data analysis to ANOVA itself. Nominal data can often be best analysed using contingency tables and the chi-square or Fisher's exact tests. As texts on categorical data analysis (e.g. Agresti, 1996) explain, contingency tables can go far beyond the 2 × 2 example given in most statistics texts. One-way, two-way and even three-way tables can be analysed rather simply to give the equivalent of ANOVA for nominal data. One can even calculate standardised residuals to provide post hoc comparisons and determine which cells in the design have significantly high contributions to the overall chi-square statistic. The use of chi-square for contingency table analysis is particularly suitable for the relative small counts of rare events during experiments, such as errors. For many years (e.g. Drury and Daniels, 1975), one author has used ANOVA to analyse task completion times while using chi-square to analyse error frequencies.

The final assumption in ANOVA is homogeneity of variance. The homogeneity of variance assumption requires the variability of the data to be the same across the conditions tested with ANOVA. This is treated in most texts (e.g. Winer, 2012), and tests such as Bartlett's test and Scheffe's test are recommended to be run on the data to check that the variance is not different across cells in the design. If the test is not satisfied, the analysis can sometimes proceed with a transformation. For example if the variance in each cell is related to the cell mean, then a logarithmic or square-root transform will help homogeneity.

DEALING WITH ANALYSIS PACKAGES

One recent survey found over 40 available software packages for DoE and statistical analysis of the data. Online comparisons between features and costs are easily located using Internet search engines. As already noted, people do not perform ANOVAs manually any longer, although most E/HF practitioners learn this skill in elementary statistics courses. With the advent of data-based quality philosophies in industry (e.g. Six-Sigma) has come the wide demand for statistical analysis tools. These quality philosophies rarely take into account E/HF topics, so it is unsurprising that some statistical packages also omit the unique issues of experimenting with people, for example repeated measures designs. The range of software is huge, as is the range of costs. Some excellent software is free, such as 'R' which is integrated in the Internet-based *e-Handbook of Statistical Methods* available at the *National Institute of Standards and Technology* (http://www.itl.nist.gov/div898/handbook/index.htm) in the United States. Others cost many thousands of pounds/dollars, although academic licences are usually available to ease the cost to students. Some have free trial introductions, enabling users to test their suitability using their own data. Most packages work on the usual computer platforms of Windows, Mac-OS and Linux, although MINITAB and SAS/Stat appear (2014) to have stopped Mac-OS support. The authors have used MINITAB, SPSS, SAS, Number Cruncher Statistical System (NCSS) and PASS at different times, and so these are the basis for the following comments. Note that statistical software is updated frequently, so that the lack of a particular feature in one package may not be forever.

Three general observations are in order when considering statistical packages. First, the experimenter should be extremely cautious when attempting to use general-purpose spreadsheets (e.g. Excel) for statistical testing of experimental data. While such programs have statistical analysis routines, they are very simple and may not have much utility in E/HF experimentation. They may also make assumptions (e.g. on F-tests) that the experimenter may not have intended, and rarely have the ability to perform ANOVA with the appropriate main effect and interaction structure discussed previously in this chapter. Second, when using a statistical analysis software package, the experimenter should always perform a single analysis on the complete design (i.e. the full model), rather than multiple one-way ANOVAs on each factor. While statistics texts never advocate the use of multiple one-way ANOVAs instead of using the full model, unfortunately, students or industrial users who have only had brief introductions to statistics seem more comfortable using such an approach. The complete ANOVA will calculate all available effects, and thus remove their variance contribution from the residual error, leading to much more powerful F-tests. Lastly, it is recommended that the experimenter, after designing the experiment, generate random data to populate the data array *before* running the experiment. In this way, the experimenter can see what the output will be like, which tests are possible and whether any factors are confounded. Running the analysis on random data first is a good final check on the design before data collection.

All of the packages mentioned perform ANOVAs to considerable levels of complexity, except for PASS which is a dedicated power and sample size calculation tool. All have moved many years ago from a command line interface to a graphical user interface. Data can be entered directly into a spreadsheet-style tableau, or pasted in from a file collected as part of the experiment. Import and export of data files between statistical packages is an important asset, allowing the strengths of different packages to be complementary. Most packages will accept standard spreadsheet files, such as .xls files from Excel, so that this intermediary can often be used if no automatic data transfer routine is available.

MINITAB grew from a very simple and small data analysis system, as its name implies. It is now a very complete package aimed at industrial users, with much support for process analysis and statistical process control. It works well for fractional factorial designs, both generating them and analysing the results. At times, the format of the structural model can be confusing in MINITAB, but it generally will cover your experiment if the model is specified correctly. NCSS has also grown considerably since it was first introduced as software for the computer industry. It includes complex ANOVAs, with options such as repeated measures designs useful to E/HF professionals. SAS, specifically SAS/STAT, is the statistical analysis component of a much broader range of business analytics. SPSS (Statistical Package for the Social Sciences), now part of a suite of statistical software from IBM, is more oriented to the social sciences as its name implies, with good support for ANOVA models where participants are a factor. It, like MINITAB, SAS/STAT and NCSS, will perform ANOVAs and MANOVAs, calculate components of variance and effect sizes, check for normality and homogeneity of variance and provide a variety of post-hoc comparisons between specific levels of a factor or interaction. In addition, all provide multivariate analysis such as factor analysis, non-parametric statistics and analysis of contingency tables.

This chapter has framed the traditional benefits and challenges of statistical DoE in terms of the work of the E/HF professional. Whether the issue is our basic science or more applied knowledge, the same principles apply: choice of dependent and independent variables, choice of sampling procedures and sample size, detailed design of the experiment to maximise effectiveness without excessive time and cost and the ethics of experimentation on humans. Without an experiment, it is difficult to impute causality to results. Without a designed experiment, it is difficult to explore the many causal factors known (or suspected) to influence human well-being and system performance. While excellent statistical texts provide many details of how to design and analyse experiments, they must use very general examples to make them applicable to a broad audience. In our discipline, we can use our specialist knowledge and insights to make more informed choices of what to vary, what to measure and what to control so as to remove sources of contamination or uncertainty. Experimental design is not performed in isolation: it is one more powerful means for us to make decisions about future actions.

REFERENCES

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*, Wiley, New York.

Bishu, R.R., Wei Wang, Hallbeck, M.S. and Cochran, D.J. (1992). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 36, 816–820. Santa Monica, CA: Human Factors and Ergonomics Society.

Chang, S.-K. and Drury, C.G. (2007). Task demands and human capabilities in door use. *Applied Ergonomics*, 38(3), 325–335.

Drury, C.G. (1994a). Function allocation in manufacturing. In: S.A. Robertson (Ed.), *Contemporary Ergonomics 1994*, Keynote address to the Ergonomics Society Meeting, Taylor & Francis Group, London, U.K., pp. 2–16.

Drury, C.G. (1994b). The speed-accuracy trade-off in industry. *Ergonomics*, 37, 747–763.

Drury, C.G., Cardwell, M.C. and Easterby, R.S. (1974). Effects of depth perception on performance of simulated materials handling task. *Ergonomics*, 17, 677–690.

Drury, C.G. and Daniels, E.B. (1975). Performance limitations in laterally constrained movements. *Ergonomics*, 18, 389–395.

Drury, C.G. and Daniels, E.B. (1980). Predicting bicycle riding performance under controlled conditions. *Journal of Safety Research*, 12(2), 86–95.

Drury, C.G. and Dawson, P. (1974). Human factors limitations in fork-lift truck performance. *Ergonomics*, 17, 447–456.

Drury, C.G., Deeb, J.M., Hartman, B., Woolley, S., Drury, C.E. and Gallagher, S. (1989a). Symmetric and asymmetric manual materials handling. Part 1: Physiology & psychophysics. *Ergonomics*, 32(5), 467–489.

Drury, C.G., Deeb, J.M., Hartman, B., Woolley, S., Drury, C.E. and Gallagher, S. (1989b). Symmetric and asymmetric manual materials handling. Part 2: Biomechanics. *Ergonomics*, 32(6), 565–583.

Drury, C.G. and Forsman, D.R. (1996). Measurement of the speed accuracy operating characteristic for visual search. *Ergonomics*, 39, 41–45.

Drury, C. G. (1997). Ergonomics Society Lecture 1996: Ergonomics and the Quality Movement. *Ergonomics*, 40(3), 249–264.

Drury, C.G., Holness, K., Ghylin, K.M. and Green, B.D. (September 2009). Using individual differences to build a common core dataset for aviation security studies. *Theoretical Issues in Ergonomics Science*, 10(5), 459–479.

Drury, C.G., Hsiao, Y.L., Joseph, C., Joshi, S., Lapp, J. and Pennathur, P.R. (2008). Posture and performance: Sitting vs. standing for security screening. *Ergonomics*, 51(3), 290–307.

Drury, C.G., Mirka, G.A. and Marras, W.S. (2007). *The Publishable Practicum*, *Contemporary Ergonomics 2007*, Taylor & Francis Group, London, U.K., pp. 247–252.

Drury, C.G. and Sinclair, M.A. (1983). Human and machine performance in an inspection task. *Human Factors*, 25, 391–399.

Fisher, D.L., Schweickert, R. and Drury, C.G. (2012). Mathematical models in engineering psychology: Optimizing performance. In: G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics*, John Wiley & Sons, Inc., Hoboken, NJ, pp. 962–989.

Hoffmann, E.R. (1992). Fitts' law with transmission delay. *Ergonomics*, 35(1), 37–48.

International Ergonomics Association, http://www.iea.cc/.

Jackson, D. (1956). A short form of the Witkin's embedded-figures test. *Journal of Abnormal and Social Psychology*, 53(2), 254–255.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.

Java applets for power and sample size, http://homepage.stat.uiowa.edu/~rlenth/Power/.

Laughery, K.R. and Drury, C.G. (l979). Human Performance as Strategy is a Two–Variable Optimization Task. *Ergonomics*, 22(l2), l325–l336.

Latorella, K.A. and Chamberlain, J.P. (2001). Decision-making in flight with different convective weather information sources: Preliminary results. *Focusing Attention on Aviation Safety. Proceedings of the 11th International Symposium on Aviation Psychology*, Columbus, Ohio.

Lin, M.L. and Radwin, R.G. (1998). Agreement between a frequency–Weighted filter for continuous biomechanical measurements of repetitive wrist flexion against a load and published psychophysical data. *Ergonomics*, 41(4), 459–475.

Lenth, R.V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, 55, 187–193.

Morawski, T., Drury, C.G. and Karwan, M.H. (1980). Predicting search performance for multiple targets. *Human Factors*, 22(6), 707–718.

Mozrall, J. and Drury, C.G. (1996). Effects of physical exertion on task performance in modern manufacturing: A taxonomy, a review and a model. *Ergonomics*, 39, 1179–1213.

Naugraiya, M. and Drury, C.G. (2009). A fractional factorial screening experiment to determine factors affecting discrete part process control. *Theoretical Issues in Ergonomics Science*, 10(1), 1–17.

NIST Statistics Handbook, http://www.itl.nist.gov/div898/handbook/.

Oxford English Dictionary, www.oxforddictionaries.com.

Pew, R.W. (1969). The speed-accuracy operating characteristic. *Acta Psychologica*, 30, 16–26.

Poulton, E.C. (1974). *Tracking Skill and Manual Control*, Academic Press, New York.

Ryan, B., Wilson, J.R., Sharples, S., Morrisroe, G. and Clarke, T. (2009). Developing a Rail Ergonomics Questionnaire (REQUEST). *Applied Ergonomics*, 40, 216–229.

Siegel, S. and Castellan, N.J. (1988). *Non-Parametric Statistics for the Behavioural Sciences*, 2nd edn., McGraw-Hill, New York.

Taguchi, G. (1986). *Introduction to quality engineering: Designing quality into products and processes*. Tokyo, Japan: Asian Productivity Organization.

Treisman, A. (1986). Properties, parts and objects. In: K.R. Boff, L. Kaufman and J.P. Thomas (Eds.), *Handbook of Perception and Human Performance*, Wiley, New York.

Wickens, C.D. and Carswell, C.M. (2012). Information processing. In: G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics*, John Wiley & Sons, Inc., Hoboken, NJ, pp. 117–161.

Winer, B.J. (2012). *Statistical Principles in Experimental Design*, Literary Licensing, LLC, New York.

Winer, B.J., Brown, D.R. and Michels, K.M. (1991). *Statistical Principles in Experimental Design*, McGraw-Hill Inc., New York.

Witkin, H. (1950). Individual perceptions of ease of perception. *Journal of Personality*, 19(1), 1–15.

3 Study and Analysis of Complex Cognitive Work

Ann Bisantz, Emilie M. Roth and Jennifer Watts-Englert

CONTENTS

INTRODUCTION AND SCOPE

Work analysis in human factors and ergonomics has shifted from focusing primarily on the physical demands of work to include both cognitive and socio-technical components of human-technology systems. This shift has been motivated by fundamental changes in work due to the advent of information technology and automation as well as high-profile system failures (e.g. Three Mile Island, numerous aviation accidents, and military incidents such as the USS *Vincennes* incident), which drew attention to the need to understand and support cognitive activities associated with complex system control. With this shift has come an evolution of work analysis methods from the time and motion study methods developed as part of the scientific management movement at the beginning of the twentieth century (Gilbreth and Gilbreth, 1917; Taylor, 1911) to task analytic techniques (e.g. Hierarchical Task Analysis; Annett and Duncan, 1967). These methods allowed the physical, perceptual, and cognitive demands of task components to be compared against human capabilities, compared with those methods which support analysis of complex cognitive and collaborative work (Bisantz and Roth, 2008; Crandall et al., 2006; Rasmussen et al., 1994; Vicente, 1999). In addition, there has been cross-fertilization of methods from the social sciences with the more traditional work analysis techniques and research and design questions addressed in human factors. For example, ethnography is now relatively commonplace in human factors (Blomberg et al., 1993; Hammersley and Atkinson, 1983). This intersection has informed both research and design, particularly regarding studies on the impact and design of modern computer and information technologies (Beyer and Holtzblatt, 1998; Carroll, 1995, 2000; Nardi, 1997; Sachs, 1995; Suchman and Trigg, 1991; Zuboff, 1987).

This chapter focuses on methodologies intended to explicitly identify the requirements of cognitive work in order to anticipate contributors to performance problems (e.g. sources of high workload, contributors to error) and specify ways to improve individual and team performance through, for example, new forms of interfaces, automation schemes, communication and coordination methods, or training approaches. Specifically, the chapter covers the various foci for analysing complex, cognitive work; provides a broad overview of the types of data collection methods typically employed; describes the theoretical perspectives and analytic frameworks that inform the analysis and outputs of the collected data; presents various outputs and applications of the methods; and describes techniques that have been used to ensure that the analyses and applications have an impact on system design and deployment.

Bisantz and Roth (2008) provide a more comprehensive literature review of cognitive analysis methods and applications. Crandall et al. (2006) and Hoffman and Militello (2008) provide in-depth coverage and ‘how-to’ guidance. Annett (2000) provides an excellent discussion of historical roots and precursors of cognitive analysis methods.

PERSPECTIVES FOR ANALYSING COMPLEX COGNITIVE WORK

Methods for analysing complex cognitive work typically provide descriptions which are informed by two mutually reinforcing perspectives: a focus on the work and domain-driven factors which shape, support and constrain performance, and a focus on the knowledge, strategies and skills held by domain practitioners (particularly, expert practitioners) that allow them to operate successfully in the domains (Bisantz and Roth, 2008).

The analysis of domain characteristics provides a framework for understanding the goals and constraints present in the domain; the resources available to achieve system goals or purposes; the information necessary to successfully oversee, control and intervene in the work domain; complexities such as interactions, uncertainty and time pressure that make system control challenging; and the likely cognitive demands imposed by these factors. For instance, analyses can identify what information is available to practitioners and whether key resources necessary for control are present, interacting goals or resource limitations that make it difficult to achieve success, information that needs to be sensed to allow operator control and contexts in which automation or other aids could be effectively deployed.

The second, complementary perspective examines the goals, motivations, knowledge, skills and strategies of domain practitioners that enable domain practitioners to operate at an expert level as well as the cognitive factors that limit the performance of less experienced individuals (e.g. incomplete or inaccurate mental models). The results can be used to identify opportunities to improve performance either through training or through the introduction of systems that more effectively support cognitive performance.

Across these perspectives, there is a focus on communication, collaboration and task coordination among multiple human and technological agents. Aspects of both the work domain (e.g. information displays, shared planning systems) and expert performance (e.g. coordination strategies, implicit and explicit communication methods) are important contributors to successful communication and coordination. Often, there is particular interest in understanding the role that existing tools or 'cognitive artefacts' play in supporting work, including the support of communication and task coordination. Such tools can range from low-tech (and often flexible) artefacts such as whiteboards and post-it notes; to specially designed, task-relevant objects (e.g. a military patrol reporting form; a pre-filled and labelled drug syringe; a light meter designed for photographers); to high-tech digital technologies such as information displays and messaging systems. For instance, investigations of cognitive performance, expertise and in situ use of technology and artefacts have been used to identify 'workaround' strategies that signal the need for more effective cognitive support (e.g. Mumaw et al., 2000; Vicente et al., 2001), as well as effective strategies or artefact properties that should be preserved or reproduced as new technology is introduced (e.g. Bisantz et al., 2010; Roth and Patterson, 2005; Roth et al., 2006; Xiao, 2005).

Across the many different data collection methods and analytic traditions associated with the analysis of complex, cognitive work, researchers and practitioners may choose to place the emphasis of their analysis on either the work domain, or practitioner expertise in response to the goals of the analysis. However, the two perspectives are clearly mutually informing. Task demands interact with practitioner expertise, strategies and practices, and work supports to make aspects of system control more or less challenging. To effectively support system design and performance-aiding efforts, analyses must reveal these complex interdependencies by taking into account both perspectives (Hoffman and Lintern, 2006; Potter et al., 2000).

OVERVIEW OF DATA COLLECTION METHODS

Numerous data collection techniques have been employed to document and describe these various foci. Due to the combined requirements to understand complexities of the work domain, along with strategies, skills and knowledge deployed by experts, a common thread among these methods is that they rely heavily on the participation of domain experts: as interview respondents, as the focus of observations or as participating members of the design teams. An analysis of complex, cognitive work simply cannot be completed without the involvement, at some level, of those individuals who perform the work. Additionally, the observations and data produced by these methods can be driven by both a top-down, and bottom-up approach. In the former case, theoretical perspectives or specific research questions (e.g. developing a model of naturalistic decision-making in a particular work setting, understanding the role of a specific artefact in communication) will guide what observations are collected. In the latter case, and often in order to generate novel hypotheses or models, observations will be influenced by the situation, opportunistically. Such data can be analysed using a grounded theory approach (Glaser and Strauss, 1967) to develop themes or hypotheses, which can in turn influence more focused data collection.

Commonly used techniques include observations, interviews and focus groups, process-tracing methods, generative methods like diaries and collages, artefact and document analysis, and analysis of critical incidents.

OBSERVATIONAL METHODS

An important method for gaining insights into complex, cognitive work is the observation of work activities. Observational methods can vary along a number of key dimensions (e.g. see Bisantz and Drury, 2005), many of which are relevant to the use of these methods for understanding complex, cognitive work. These choices include the setting for observations, degree of interaction among the research and those being observed, whether observations are drawn from real-life or through audio or video-taping sessions and the degree of structure in collecting and reporting observational data.

Setting

While typical applications of observational methods occur in naturalistic settings as work activities are unfolding (e.g. Roth et al., 2004), it is important to recognize the difference between studies in real world settings, often considered field or case studies, and the use of observational, or non-intrusive data-gathering techniques. It is possible to conduct studies in field settings which do not rely only on observation as a measurement technique (e.g. eye gaze patterns; Montague et al., 2011). Likewise, it is possible to make observations in settings which are pseudo-realistic, such as task simulators (Bowers et al., 1998) or training exercises (Artman, 2000), or even in laboratory settings, where observations can be made of behaviour or activities during the performance of more controlled tasks. In some cases, individuals may be instructed to think aloud as they perform the task, to provide an ongoing verbal protocol of the task (Bainbridge and Sanderson, 1995; Ericsson and Simon, 1993). For instance, Gorman and Militello (2004) observed functionally blind users performing specified Internet search tasks in conjunction with a screen reader. Users were asked to think aloud during the task in a laboratory environment, and decision models were developed to describe their activities.

Observational studies in field settings are particularly useful for identifying mismatches between how work is depicted in formal processes and procedures and how it is actually performed, often revealing 'home-grown' tools and workarounds that domain practitioners generate to cope with aspects of task complexity that are not well supported (e.g. Roth et al., 2006). Differences between 'canonical' descriptions of work and actual work practice can reveal opportunities to improve performance through more effective support.

Degree of Interaction

At one extreme, the researcher or analyst acts as an observer, or the instrument of measurement, with minimal interaction with the people and situations being observed. This type of observation may be appropriate in cases where research questions are well defined, where the researcher has a strong understanding of domain activities and where participant actions are easily observed. More typically, however, real-time observations in actual work settings are often combined with informal interviews conducted as the task progresses. In some cases, participant observation methods are employed in which analysts participate in the work performance (often in an 'apprenticeship' capacity). For instance, Burns and Vicente (2000) acted as human–machine design consultants, and thus as participant-observers, during their observations of the interaction between human factors and other design constraints.

Parallel Sources of Data

In most studies that utilize observational techniques, additional measurement techniques are used. Objective records of unfolding events such as events and process variable states may be collected and combined with the observations that are made (either in real time or from recordings) to create a rich protocol or 'process trace' that captures the unfolding events and task activities, thus allowing the activities of operators to be understood within the context of the task itself (Woods, 1993). For instance, Cook and Woods (1996) utilized process-tracing methods, collecting data regarding patient states (e.g. physiological parameters) and conducting concurrent interviews in addition to making observations of practitioners using medical devices. Similarly, Seagull and Sanderson (2001) constructed process-tracing logs which included surgical events and equipment states, as well as observed activities. Observations of artefact use, teamwork and collaboration, and information use may be supplemented with interviews (either concurrently, as the work activity progresses, or in a more formal interview setting), focus groups or through the analysis of archival data sources. In another example, Degani and Wiener (1993, 1997) supplemented observations of pilot checklist use with interviews with pilots and analysis of archival data (aviation accident and incident records).

Collecting Observational Data

Observations can vary from unstructured, opportunistic field notes to more structured observations based on predetermined categories. Regardless of the format, observations are guided or shaped by research questions and/or analysts' expertise, goals and theoretical perspectives. Data collection methods can range from decidedly low-tech (pen and paper) to audio and video recordings, to computer-supported data collection. For instance, Patterson and Woods (2001) conducted observations that focused on space shuttle mission control shift change and handovers during an actual space shuttle mission. They combined observations with handwritten logs and spontaneous verbalizations of the controllers (captured via audiotape), along with flight plans, to identify handover activities that were related to fault management, re-planning and maintaining common communicational ground. Handwritten notes can be free form (in the case of unstructured data) or rely on scales and forms developed based on structured categories of data to be collected (Sharples et al., 2011). Specifically developed software programs combined with portable or handheld computers can be used to allow coding of real-time activities. There are also hardware and software systems which facilitate analysis of real-time or videotaped data (e.g. Noldus™, NVivo™, Morae™, also see Mackenzie and Xiao, 2011). These systems provide functionality such as time-based assignment of codes and descriptions to video segments and integrated control of video.

Video and audio recordings can be used to capture observational data for later analysis. Advantages of collecting video data, and making observations from the recordings, are numerous. Unlike in-person observation, situations captured in tapes can be reviewed multiple times, in order to fully assess aspects of complex or fast-paced activities. For example, Kirschenbaum (2004) observed groups of weather forecasters in either their everyday work setting, performing normal forecasting duties, or in a simulated shipboard forecasting centre, working on a provided scenario. Team activities, along with think-aloud protocols, were captured via videotape to allow for detailed qualitative data analysis of cognitive activities related to weather forecasting. Video records also make it possible to collect cued retrospective explanations of task performance by the individuals who participated in the task (Hoc and Leplat, 1983). Seagull and Xiao (2001) used video recordings on which eye-tracking data had been superimposed to study a surgical procedure. The recordings were made from the perspective of the physician performing the procedure (wearing mobile recording and eye-tracking equipment). The eye-tracking data indicated where (in the operating room) the physician looked throughout the procedure. The tapes were reviewed by the physician and other subject matter experts (SMEs) to determine what the physician had to look at to accomplish the task, what that information would indicate and why it was sought by the physician at that point in the task – in essence, to identify information cues and their purpose during the task. They can also be leveraged to elicit additional knowledge from other SMEs. Miller et al. (2006) described a critiquing process for performing a cognitive task analysis (CTA) that relies on video- and audio-recorded data of a novice performing a task. The results are used to create a 'script' of the novice's performance that can then be critiqued by SMEs. They recorded a novice completing a complex (military intelligence analysis) task, during which the novice was asked to think aloud. Six expert intelligence analysts were read a transcript of the novice's verbalisations while being shown additional material (e.g. screen shots captured, documents accessed and handwritten notes generated by the novice during the task). Experts were asked to comment on the novices' performance as the script was presented. Audio and video recordings, along with handwritten notes of the critiquing process, were used to generate a protocol, which was then analysed to provide insight into how experts approach this task.

However, there are also disadvantages to the use of video. There are some situations, for legal or practical reasons, where the use of video cameras is not possible. For instance, Seagull and Sanderson (2001) were unable to secure permission to videotape surgeries. For studies involving large areas, a single (or even multiple) fixed cameras may be impractical. Setting up, capturing, storing and converting video data between the variety of available digital and analogue formats (if desired) alone can be costly and time consuming, without even considering the time required for analysis. Sanderson and Fisher (1994) suggest that the time for analysis of sequential observational data, such as that stemming from recorded observations, can reach levels as high as 100 to 1, depending on factors such as the granularity of analysis being conducted (i.e. the length or frequency of activities being recorded, such as gestures or motions vs. movement across a room, or topics raised at a meeting), the specificity of research questions being asked and the complexity of inferences required from the data. Camera angles and distances from the subject can lead to distorted or obstructed images, making it difficult to identify all activities, or detailed interactions with workplace artefacts, reliably. For instance, it may be possible to determine that a person referred to a piece of paper, but not identify what the paper or its content was. Events which occur out of camera coverage cannot be analysed.

These latter cases, in particular, could be alleviated with the supplement of in-person observation or retrospective verbal protocols. While a human observer might not be able to capture all elements of sequential activities in a situation, it is possible for documents to be examined, questions to be asked and activities and equipment to be scrutinized in a more detailed fashion.

INTERVIEW AND FOCUS GROUP METHODS

Interviews and focus groups are among the most commonly applied knowledge acquisition methods. Unstructured interviews tend to be free form, in which neither the content nor the sequence of the interview topics is predetermined (Cooke, 1994), and are most appropriate early in the knowledge acquisition process, when the analyst is attempting to gain a broad overview of the domain. More typically, analysts will use a semi-structured interview approach in which a list of topics and candidate questions is generated ahead of time, but the specific topics and the order in which they are covered are guided by the responses obtained (e.g. Mumaw et al., 2000; Roth et al., 1999). Structured interview techniques utilize a specific set of questions in a specific order.

A number of semi-structured interview techniques have been developed to uncover the demands of complex cognitive work and the strategies that domain practitioners have developed to meet those demands. A powerful technique is to ground the interviews in analysis of actual past critical incidents so as to understand what made them challenging and why the individuals who confronted the situation succeeded or failed (Flanagan, 1954). The critical decision method (CDM) is a structured interview technique that uses the critical incident technique (Hoffman et al., 1998; Klein and Armstrong, 2005; Klein et al., 1989). CDM focuses on the analysis of challenging situations that the domain practitioner has experienced, and includes four interview phases, or ‘sweeps’, that examine a past incident in successively greater detail. These phases provide a description of the subtle cues, knowledge, goals, expectancies and expert strategies that domain experts use to handle cognitively challenging situations. CDM has been successfully employed to analyse the basis of experts in a variety of domains, such as fire fighting, neonatal caregiving, and intelligence analysis (Baxter et al., 2005; Hutchins et al., 2003; Klein, 1998).

Concept mapping is another interview technique that is widely used to uncover and document the knowledge and strategies that underlie expertise (Crandall et al., 2006). In concept mapping, the analyst helps domain practitioners build up a representation of their domain knowledge using concept maps, which are directed graphs made up of concept nodes connected by labelled links. They are used to capture the content and structure of domain knowledge that experts employ in problem-solving and decision-making. Concept mapping is typically conducted in group sessions that include multiple domain practitioners (e.g. three to five) and two facilitators. One facilitator provides support in the form of suggestions and probe questions, and the second facilitator creates the concept map based on the participants’ comments for all to review and modify. The output is a graphic representation of expert domain knowledge that can be used as input to design of training or decision aids.

Focus groups are interviews conducted with multiple participants. Some special considerations with respect to focus groups include both the number and choice of participants. General recommendations are to limit the size of focus groups to 5–8 participants (Krueger and Casey, 2009), so that everyone has the opportunity to contribute. Some care should be taken, particularly in work settings where participants are likely to know and work with one another, to avoid groups which mix participants of differing status (e.g. employees and managers, physicians and technicians), so that all participants may feel free to contribute openly. Focus groups are best suited for unstructured and semi-structured questioning, which allows participants to add to other participants' comments.

GENERATIVE METHODS

Generative methods are analysis techniques that invite participants to explore and contribute insight about their lives and work practices to the design process. These methods can be used to better understand user needs and the context in which a product might be used (Stappers et al., 2003). Findings can contribute to the development of new product concepts and the generation of user scenarios and design requirements. With generative methods, participants create, or generate, new material, or artefacts (Sanders, 2000) in response to specific questions posed by researchers. These artefacts can take many forms, including collages, where participants create collections of images and/or words in response to research questions; diaries, where participants independently record information about their activities and practices or Velcro models, where participants create a design for a product or system using elements provided by the researcher (Chung, 2004; Stappers and Sanders, 2003, 2005).

Generative methods are usually guided by an open-ended question motivated by the research topic. For example, Stappers et al. (2003) describe a generative study where participants created a collage showing their experience of home in the past, present and future, and Watts-Perotti et al. (2011) asked mobile workers to create collages showing their current work environment and their ideal work environment. Watts-Perotti et al. comment that

> The collage activity allowed participants to examine and describe their work environment in nonverbal ways, and led to a richer description of what they liked about their current environment, and what they would like to change. This part of the study provided details about how work and personal life intermingled, and how participants struggled to maintain a balance in their lives.

After creating the collages, participants described the elements of the collage and explained what these elements meant to them. This discussion was richer than a typical interview, because the collages helped participants become more aware of their work environments, and how they felt about them.

Participants can use personal materials, or those provided by researchers, to create generative artefacts. For example, they could create collages using personal images and objects from home, or using words, images and other materials provided by the researchers. In the case of Velcro modelling, participants typically use material provided by the researchers. Chung (2004) describes a Velcro modelling study that explored ideas for the controls and interface design for microwave ovens. In this study, participants used felt shapes provided by researchers to create an arrangement of controls and displays for a microwave. Velcro modelling is especially good for helping participants create ideas in three-dimensional form (Stappers and Sanders, 2005). For example, Sanders (1993) describes a study where Velcro modelling was used to understand user needs related to the layout of controls for the cab of a military loader vehicle.

Once participants have created the generative materials, they usually verbally present, or explain, the materials to the researchers (Stappers and Sanders, 2005). In this process, participants access and describe parts of their work or experiences that would not typically arise from a natural conversation. Also, as with the Watts-Perotti et al.'s study, the discussion with researchers is grounded in participants' experiences, because they have spent some time exploring their experience through the creation of the generative materials. It is through this explanation that researchers can better understand participants' personal meaning behind the elements of their materials (i.e. diary entries, collage or mind map elements, etc.).

Analysis of the materials can include qualitative analysis of the explanation of the materials, together with qualitative or quantitative analysis of the created materials. For example, researchers can trace visual or conceptual themes conveyed across materials created by different participants, and/or they can perform statistical analysis on the content of the materials like words or specific images or themes (Stappers and Sanders, 2003).

Generative analysis methods have specific advantages, and are often used to complement other methods of cognitive inquiry. These methods, similar to psychological projective techniques, 'are based on the belief that all people can project and express their needs, wants and aspirations through the use and interpretation of ambiguous visual stimuli' (Stappers and Sanders, 2005). They have the potential to reveal latent unmet needs and aspirations (Chung, 2004), and can provide a direct language for designers, who can in turn incorporate the visual characteristics of the materials into design concepts (Stappers and Sanders, 2005). See Watts-Perotti et al. (2011) for more details about a study that combined generative techniques (diaries and collages) with more traditional cognitive analysis techniques such as observations and interviews to understand the work practices of mobile workers.

THEORETICAL PERSPECTIVES AND ANALYSIS FRAMEWORKS

Cognitive Psychology and Cognitive Science

One prominent approach to cognitive analysis draws on laboratory study methods from the cognitive science and cognitive psychology literature (Ericsson and Simon, 1980; Hoffman, 1987, 2008). This approach has contributed a variety of specific cognitive analysis methods for eliciting and representing the knowledge and strategies that underlie performance at different levels of expertise. This includes the use of 'think-aloud' protocols and the application of psychological scaling methods such as multidimensional scaling of pairwise ratings that provide a means to uncover how experts organize their knowledge (Cooke, 1994; Ericsson and Simon, 1993). It also led to methods for modelling the mental operations that underlie cognitive performance (Gray, 2007).

Cognitive Systems Engineering

Another approach to cognitive analysis has its roots in the cognitive systems engineering (CSE) tradition (Hollnagel and Woods, 1983; Rasmussen, 1986). Analyses of complex accidents, such as the Three Mile Island nuclear power plant incident, revealed that crisis events often involve a confluence of events that have not been previously anticipated. As a consequence, the training, procedures and support systems in place often prove to be inadequate (Sanderson, 2003). One of the major innovations of cognitive systems engineering was the development of functional analysis methods that define the goals, constraints and affordances in a domain that constitute the cognitive problem-space that practitioners need to cope with. The results of the functional analyses are used to develop systems that enable domain practitioners to directly 'perceive' and reason about system goals, constraints and affordances – so as to be able to perform effectively under unanticipated conditions (Hollnagel and Woods, 2005; Rasmussen, 1986; Roth and Woods, 1988; Woods and Hollnagel, 1987, 2006; Woods and Roth, 1988).

Cognitive work analysis (CWA) is the most fully developed analysis framework that grew out of this tradition (Rasmussen, 1986; Vicente, 1999). CWA includes five interlinked analyses that focus successively on different layers of constraints ranging from characteristics of the work domain to cognitive competencies of the individuals engaged in the work. The grounding layer of a CWA is