Forming connections between human performance and design, this new edition of *Engineering Psychology and Human Performance* examines human–machine interaction. The book is organized directly from a psychological perspective of human information processing, and chapters correspond to the flow of information as it is processed by a human being—from the senses, through the brain, to action—rather than from the perspective of system components or engineering design concepts. Upon completing this book, readers will be able to identify how human ability contributes to the design of technology; understand the connections within human information processing and human performance; challenge the way they think about technology's influence on human performance; and show how theoretical advances have been, or might be, applied to improving human–machine interactions.

This new edition includes the following key features:

- A new chapter on research methods
- Sections on interruption management and distracted driving as cogent examples of applications of engineering psychology theory to societal problems
- A greatly increased number of references to pandemics, technostress, and misinformation
- New applications
- Amplified emphasis on readability and commonsense examples
- Updated and new references throughout the text.

This book is ideal for psychology and engineering students, as well as practitioners in engineering psychology, human performance, and human factors. **The text is also supplemented by online resources for students and instructors.**

Christopher D. Wickens is Professor Emeritus from the University of Illinois, Department of Psychology; Adjunct Professor, Colorado State University, Department of Psychology; and Senior Scientist, Alion Science and Technology, Boulder, Colorado. He has won numerous teaching awards, including the Psi-Chi award and the Paul M. Fitts award from the Human Factors and Ergonomics Society.

William "Deak" S. Helton is Professor and Director of Human Factors and Applied Cognition at George Mason University and an Adjunct Professor at the University of Canterbury (New Zealand). He was awarded the Earl Alluisi Award for Early Career Achievement by the American Psychological Association and the Griffith Prize by the Southern Society for Philosophy and Psychology.

Justin G. Hollands is a Defense Scientist and Senior Advisor with the Human Systems Integration Section at Defence Research and Development Canada—Toronto. He is also an Adjunct Professor of Mechanical and Industrial Engineering at the University of Toronto.

Simon Banbury is co-founder and President of C3 Human Factors Consulting Inc., an independent Canadian-based human factors consultancy specializing in optimizing how people interact with technology.

Cover image: © Getty Images

ENGINEERING PSYCHOLOGY / HUMAN FACTORS

Routledge
Taylor & Francis Group
www.routledge.com

COMPANION WEBSITE
www.routledge.com/9781032011738

an informa business
ISBN 978-1-03-201173-8
9 781032 011738

Routledge titles are available as eBook editions in a range of digital formats

ENGINEERING PSYCHOLOGY AND HUMAN PERFORMANCE

FIFTH EDITION

Christopher D. Wickens, William S. Helton, Justin G. Hollands, and Simon Banbury

ROUTLEDGE

ENGINEERING PSYCHOLOGY AND HUMAN PERFORMANCE

FIFTH EDITION

Christopher D. Wickens, William S. Helton, Justin G. Hollands, and Simon Banbury

ROUTLEDGE

Engineering Psychology and Human Performance

Forming connections between human performance and design, this new edition of *Engineering Psychology and Human Performance* examines human–machine interaction. The book is organized directly from a psychological perspective of human information processing, and chapters correspond to the flow of information as it is processed by a human being—from the senses, through the brain, to action—rather than from the perspective of system components or engineering design concepts. Upon completing this book, readers will be able to identify how human ability contributes to the design of technology; understand the connections within human information processing and human performance; challenge the way they think about technology's influence on human performance; and show how theoretical advances have been, or might be, applied to improving human–machine interactions.

This new edition includes the following key features:

- A new chapter on research methods
- Sections on interruption management and distracted driving as cogent examples of applications of engineering psychology theory to societal problems
- A greatly increased number of references to pandemics, technostress, and misinformation
- New applications
- Amplified emphasis on readability and commonsense examples
- Updated and new references throughout the text.

This book is ideal for psychology and engineering students, as well as practitioners in engineering psychology, human performance, and human factors. **The text is also supplemented by online resources for students and instructors.**

Christopher D. Wickens is Professor Emeritus from the University of Illinois, Department of Psychology; Adjunct Professor, Colorado State University, Department of Psychology; and Senior Scientist, Alion Science and Technology, Boulder, Colorado. Wickens taught engineering and experimental psychology, human factors engineering, and aviation psychology for 30 years at the University of Illinois, where he supervised 38 PhD students. For 20 years he was also head of the Aviation Human Factors program at the University of Illinois. He has won numerous teaching awards, including the Psi Chi award for excellence in undergraduate teaching and the Paul M. Fitts award for outstanding contributions to the education and training of human factors specialists from the Human Factors and Ergonomics Society. He has also received the annual FAA Excellence in Aviation Award. He is a Fellow of the Human Factors Society. His main research interests are in applied attention theory, aviation psychology, and human performance modeling. He is an avid mountain climber.

William "Deak" S. Helton is Professor and Director of the Human Factors and Applied Cognition program at George Mason University and an Adjunct Professor at the University of Canterbury (New Zealand). He has supervised 13 PhD students to graduation. His research interests are in attention, applying ergonomic approaches to workers other than humans (dogs in particular), wearable interfaces, disaster response, and human performance. He has published more than 100 peer-reviewed journal articles and edited the book *Canine Ergonomics: The Science of Working Dogs*. He was awarded the Earl Alluisi Award for Early Career Achievement by the American Psychological Association and the Griffith Prize by the Southern Society for Philosophy and

Psychology. He received his PhD in Human Factors Psychology from the University of Cincinnati in 2002 under the mentorship of the late Joel Warm.

Justin G. Hollands is a Defense Scientist and Senior Advisor with the Human Systems Integration Section at Defence Research and Development Canada—Toronto. He is also an Adjunct Professor of Mechanical and Industrial Engineering at the University of Toronto. From 1994 to 1999, Dr. Hollands was an Assistant Professor of Psychology at the University of Idaho. He received his PhD in Cognitive Psychology from the University of Toronto in 1993 and an MA in Human Factors Psychology from the University of Guelph in 1989. He received a BA in Psychology (honors cooperative program) from the University of Waterloo in 1986. His cooperative program experience at the Ontario Ministry of Transportation and Communications, Bell-Northern Research, and IBM Canada sparked his interest in human factors and engineering psychology. Dr. Hollands has authored or coauthored more than 50 peer-reviewed articles. He is interested in display and interface design, human reliance on automated systems, visual momentum, visual attention, and human perception and psychophysical scaling.

Simon Banbury is co-founder and President of C3 Human Factors Consulting Inc., an independent Canadian-based human factors consultancy specializing in optimizing how people interact with technology. He received his PhD in Cognitive Psychology from the University of Reading in 1996 and has over 25 years of human factors consultancy and applied research experience in defense, industrial, and academic domains. Over this time he has worked as a human factors consultant in the defense and industrial sectors, a Lecturer in Psychology at Cardiff University, and a defense scientist for the United Kingdom's Defense Evaluation and Research Agency. He has published widely on the applied aspects of attention and memory, including the effects of extraneous noise on performance in the office and the flight deck.

Fifth Edition

Engineering Psychology and Human Performance

Christopher D. Wickens, William S. Helton,
Justin G. Hollands, and Simon Banbury

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Fifth edition published 2022
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2022 Taylor & Francis

The right of Christopher D. Wickens, William S. Helton, Justin G. Hollands, and Simon Banbury to be identified as authors of this work has been asserted by them in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

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

First edition published by Douglas Books 1984
Fourth edition published by Pearson 2012

Library of Congress Cataloging-in-Publication Data
Names: Wickens, Christopher D., author. | Helton, William S., author. | Hollands, Justin G., author. | Banbury, Simon, author.
Title: Engineering psychology and human performance / Christopher D. Wickens, William S. Helton, Justin G. Hollands and Simon Banbury.
Description: 5th edition. | New York, NY : Routledge, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020057306 (print) | LCCN 2020057307 (ebook) | ISBN 9781032011738 (hardback) | ISBN 9781003177616 (ebook)
Subjects: LCSH: Human engineering. | Human-machine systems. | Psychology, Industrial.
Classification: LCC TA166 .W53 2021 (print) | LCC TA166 (ebook) | DDC 620.8/2—dc23
LC record available at https://lccn.loc.gov/2020057306
LC ebook record available at https://lccn.loc.gov/2020057307

ISBN: 978-1-032-01173-8 (hbk)
ISBN: 978-1-003-17761-6 (ebk)

DOI: 10.4324/9781003177616

Typeset in Minion Pro
by Apex CoVantage, LLC

Access the companion website: routledge.com/9781032011738

Dedicated to Raja Parasuraman, Neville Moray, and Joel Warm: pioneers and leaders in engineering psychology

BRIEF CONTENTS

CONTENTS

PREFACE

Each edition of this book was written to address the gap between the problems of system design and much of the excellent theoretical research in experimental psychology and human performance. Many human–machine systems do not work as well as they could because they impose requirements on the human user that are incompatible with the way people attend, perceive, think, remember, decide, and act; that is, the way in which people perform or *process information*. Over the past seven decades, tremendous gains have been made in understanding and modeling human information processing and human performance. Our goal is to show how these theoretical advances have been, or might be, applied to improving human–machine interaction.

Although engineers encountering system design problems may find some answers or guidelines either implicitly or explicitly stated in this book, it is not intended to be a handbook of human factors engineering. Many of the references in the text provide a more comprehensive tabulation of such guidelines as well as practical guidelines on how to apply them. Instead, we have organized the book directly from the psychological perspective of human information processing. The chapters generally correspond to the flow of information as it is processed by a human being—from the senses, through the brain, to action—rather than from the perspective of system components or engineering design concepts, such as displays, illumination, controls, computers, and keyboards. Furthermore, although the following pages contain recommendations for certain system design principles, many of these are based only on laboratory research and theory; they have not been tested in real-world systems.

A solid grasp of theory provides a strong base from which the specific principles of good human factors can be more readily derived. Our intended audience, therefore, is: (1) the student in psychology, who recognizes the real-world relevance of the theoretical principles of psychology that he or she may have encountered in other courses; (2) the engineering student, who, while learning to design and build systems with which humans interact, appreciates not only the nature of human limitations—the essence of human factors—but also the theoretical principles of human performance and information processing underlying them; and (3) the actual practitioner in engineering psychology, human performance, and human factors engineering, who understands the close cooperation that should exist between principles and theories of psychology and issues in system design.

The 13 chapters of the book span a wide range of human performance components. The introduction in Chapter 1 places engineering psychology into the broader framework of human factors and system design. Chapter 2, new to this edition, presents information on research methods. Chapters 3 through 9 deal with perception, attention, cognition (both spatial and verbal), memory, learning, and decision making, emphasizing the potential applications of these areas of cognitive psychology. Chapters 10 and 11 cover the selection and execution of control actions, error, and time-sharing. Chapter 12 covers two more integrated concepts: workload and stress. Chapter 13 addresses topics of human–automation interaction. Finally, an Epilogue is provided that highlights certain critical issues that transcend many of the prior chapters.

Although the 13 chapters are interrelated (just as are the components of human information processing), we have constructed them in such a way that any chapter may be deleted from a course syllabus and still leave a coherent body. Thus, for example, a course on applied cognitive psychology might include Chapters 1 through 9 and Chapter 11, and another emphasizing engineering applications might include Chapters 1, 2, 3, 5, 6, 10, 11, 12, 13, and the Epilogue.

NEW TO THIS EDITION

Changes since the Fourth Edition that appear throughout the text:

- *A new coauthor: William S. Helton provides expertise in research methods, sustained attention, workload, and stress.*
- *A new chapter on research methods (Chapter 2).*
- *Revision of all chapters to improve clarity and simplify understanding.*
- *New sections on interruption management and distracted driving as cogent examples of applications of engineering psychology theory to societal problems.*
- *Greatly increased number of references to pandemics, technostress, misinformation, etc.*
- *New applications.*
- *Amplified emphasis on readability and commonsense examples.*
- *Updated as well as new references throughout the text.*

Chapter-by-Chapter Changes

CHAPTER 2

- *New chapter on research methods.*

CHAPTER 3

- *Revised section on signal detection theory.*
- *New section on vigilance and added material on alternative theories, rest breaks, memory load, etc.*

CHAPTER 4

- *New treatment of selective visual attention with substantial reorganization and new examples.*
- *Revised section on change blindness, with the addition of material on change deafness and change numbness.*
- *New material on distractions.*

CHAPTER 5

- *New material on video and 3D audio displays.*
- *Reorganized and simplified graph guidelines.*
- *New examples of technologies using spatial displays.*

CHAPTER 6

- *New section on Global Positioning System (GPS) navigation.*
- *New section on individual differences in spatial abilities.*

- *New material of distortions in physical properties and visualization of uncertainty.*
- *Extension of tracking examples.*
- *New material on augmented reality and head-mounted displays.*

CHAPTER 7

- *New material on the word-superiority effect.*
- *New examples for medical and health applications.*
- *New and updated material on instructions and warnings.*
- *New sections on the communication of health risks and communicating misinformation.*

CHAPTER 8

- *Updated examples of expertise and chunking.*
- *New section on system-level situational awareness.*
- *Updated section on planning and problem solving.*
- *New section on adaptive training.*
- *New and updated material on knowledge representation.*

CHAPTER 9

- *New section on complementary approaches of naturalistic and dynamic decision making.*
- *Updated material and examples of diagnosis and situation assessment.*
- *Updated framing effects material and the role of expertise in decision making.*
- *New sections on risks and nudges.*
- *New material on debiasing and improving decision making.*

CHAPTER 10

- *Revised section on the selection of action.*
- *New practical examples of the role of the selection of action in accidental shootings.*
- *New information on tactile warnings.*
- *New sections on operator variables, the speed–accuracy trade-off, and response types, with practical examples.*
- *New and updated examples of human error in the real world.*

CHAPTER 11

- *Revised material on computational models of resources.*
- *New sections on interruption and task management and individual differences in multitasking.*
- *Updated materials on distracted driving.*
- *New section on individual differences in multitasking, with a focus on differences related to abilities, expertise, and aging.*

CHAPTER 12

- *Revised to remove materials on individual differences and to downplay neuro-ergonomics to focus more on workload and stress.*
- *Updated material on the measurement of workload and stress.*
- *New section on technostress.*

CHAPTER 13

- *Updated material on automation trust and dependence.*
- *New and expanded sections on mitigating human–automation problems.*
- *Expanded and revised section on automation transparency.*
- *New section on training and individual differences in automation use.*

EPILOGUE

- *The epilogue integrates several of the central and recurring themes of the book.*

SUPPLEMENTS

Please visit the companion website at [INSERT URL].

ACKNOWLEDGMENTS

In any project of this kind, one is indebted to numerous people for assistance. For all of us, the list includes several colleagues who have read and commented on various chapters, have provided feedback on the prior editions, or have stimulated our thinking. In addition to all acknowledgments in the first two editions (the text of which, of course, remains very much at the core of the current book), the first author would like to acknowledge the contributions of faculty colleagues and countless students who, in one form or another, have offered feedback regarding either bad or good sections of prior editions. In particular, we would like to acknowledge the contribution of the late Raja Parasuraman (one of the authors of the Fourth Edition); he was a giant in the field of engineering psychology and is missed by everyone.

Christopher D. Wickens would like to acknowledge the contributions of four specific individuals who have contributed to the development of his interest in engineering psychology: his father, Delos Wickens, stimulated his early interest in experimental psychology; Dick Pew introduced him to academic research in engineering psychology and human performance; the late Stanley Roscoe pointed out the importance of good research applications to system design; and Emanuel Donchin continues to emphasize the importance of solid theoretical and empirical research.

William S. Helton would like to acknowledge his early mentors who introduced him to engineering psychology: Joel Warm, William Dember, and Gerald Matthews. He would also like to acknowledge the many colleagues and students at George Mason University with whom he discussed topics of this book. Finally, he would like to thank his wife and children for their patience and support while working on this book.

Justin G. Hollands would like to acknowledge his many colleagues in the Human Systems Integration Section at Defence Research and Development Canada—Toronto, with whom he discussed various topics in the book at different times. He would especially like to thank Linda Bossi for seeing the "big picture" and supporting his efforts. He also thanks Stewart Harrison for tracking down many of the references. Finally, he thanks his family, and especially his wife, Cindy, for their patience while this book was being written.

Simon Banbury would like to thank Cindy Chamberland and Alexandre Marois for their help conducting the literature searches and Tess Banbury for her help compiling the Instructor's Manual. Thanks also to his wife, Jennifer, and daughters, Tess and Charlotte, for their support and encouragement.

Christopher D. Wickens
William S. Helton
Justin G. Hollands
Simon Banbury

1 INTRODUCTION

The field of human factors engineering (Salvendy & Karwalski, 2022), along with the closely related disciplines of human–systems integration (Durso & Boehm-Davis, 2014), human–computer interaction (Jacko, 2012), cognitive engineering (Lee & Kirlik, 2013), **Human Factors Engineering** (Lee, Wickens, Ng-Boyle, & Liu, 2017) and user-interface design (Buxton, 2007), addresses issues of how humans interact with technology. The field has developed rapidly since its origin after World War II. During World War II, experimental psychologists were called in to help understand why pilots were crashing perfectly good aircraft (Fitts & Jones, 1947), why vigilance for enemies was sometimes wanting (Mackworth, 1948), or how learning theory could be harnessed to better train military personnel. Since that time, over the past 80 or so years, the field has seen growth and expansion into areas such as consumer products, business, highway safety, telecommunications, health care, and, most recently, cybersecurity.

1.1 DEFINITIONS

1.1.1 Engineering Psychology

Within the broader field of human factors lies the discipline of **engineering psychology**, the focus of this book. Engineering psychology focuses on "human factors from the neck up," in contrast to many applications of human factors to issues "below the neck," such as lower back injuries, fatigue, work physiology, and so forth. Much of this latter focus is covered in the general discipline of **ergonomics**, the study of work, although classic ergonomics has itself spawned the study of **cognitive ergonomics**, and/or **cognitive engineering**, both of which naturally focus on human mental work, "above the neck" (Vicente, 1999; Jenkins Stanton et al., 2009). An additional contrast with the broader field of human factors engineering (Lee et al., 2017) is that human factors focuses much more heavily upon *design* (of products, workstations, etc.) and the evaluation of those designs, than does engineering psychology. Engineering psychology is, after all, a subdiscipline of psycholog*y*, and not engineering.

Engineering psychology can also be described within the broader discipline of psychology, and within this, the somewhat narrower discipline of **applied psychology**. In the latter, the study of behavior and cognition is focused on the applications of those principles and theories of behavior and cognition to areas beyond the laboratory, such as industry, schools, counseling, mental illness, and sports. Within this broader set of applications then, the focus of engineering psychology tends to be on performance *in the workplace* (expanded to include transportation and some aspects of the home), hence characterizing its close linkage back to ergonomics, the study of work, and particularly cognitive ergonomics.

DOI: 10.4324/9781003177616-1

But to highlight the uniqueness of engineering psychology again, what distinguishes it from cognitive ergonomics is that the former has the heavy, and some would say necessary, basis in *theory*: the theories of brain, behavior, and cognition that are applicable to the workplace. Cognitive ergonomics is certainly not devoid of theory, but it also broadens its focus to consider issues of task description and analysis, design, and principles of design that may not directly translate to or arise from theory.

In distinguishing engineering psychology from many aspects of basic psychology (and usually experimental psychology), engineering psychology must be concerned with the eventual applications of its theories and principles, whereas experimental psychology need not be. This has three implications for research in the two related disciplines. First, experimental psychology is quite concerned with the issues of **experimental control**. All variables should be held constant except those manipulated in the experiment. Second, the concern for statistical significance often dominates that of practical significance. A statistically significant effect measured in the laboratory of only 10 milliseconds (msec) can signal an exciting discovery, but such an effect may be of limited usefulness in the workplace beyond the laboratory. Third, the task of the participant in basic laboratory research is typically that designed by the experimenter for theoretical reasons. The task for the engineering psychologist typically has a mapping to some real-world activity.

In engineering psychology, although there is still concern for control in its experimental research, too much control may produce effects that, like the 10-msec effect above, would simply "wash out" when the person performs in the workplace, with its many other competing (and hence "noisy") influences on human behavior. The second difference is related to the first. Although engineering psychologists do pay a lot of attention to statistics and statistical significance (see Chapter 2), they also realize that without considering practical significance, a particular finding or principle will simply not scale up to the workplace, where it may be "handed off" to the human factors engineer, with the commitment to design. Third, in designing a task for experimental participants, the engineering psychologist must always consider its relevance to tasks beyond the laboratory. The engineering psychologist should know and understand the relevant real-world context and tasks, and this knowledge should inspire the design of the experimental task.

Of course, in practice, such distinctions are fuzzy rather than crisp. We have noted the fuzziness of defining what is and is not the "workplace." For example, highway safety is very much within the domain of engineering psychology, but it does not matter whether the person is driving a truck for work or a car for pleasure. As another example of this fuzziness, sometimes issues below the neck influence those above, as when we are *distracted* by the discomfort resulting from a poorly designed physical workplace. Furthermore, many issues of design addressed by human factors depend on engineering psychology principles (Peacock, 2011), and when designs are evaluated outside the laboratory, their results may lead to further controlled experiments to refine the principles upon which those designs were based. And in this same way, lessons learned and challenges felt by the engineering psychologist should always feed back to the basic psychologist to inform where new theory is needed or old theory may be wanting. Experimental psychologists often are interested in knowing the limitations of their models and principles in real-world settings, and by providing such feedback, engineering psychologists help to ensure that application is considered even when more basic research is conducted.

1.1.2 Human Performance

The second part of the title of the book, **human performance**, also deserves some explanation. Here, our emphasis is on the *quality* of performance (e.g., better or worse), and here we typically think of measures of "the big three":

Speed (faster is better),
Accuracy (higher is better), and
Attention demands (less is generally better).

Thus, we might judge that the perfect principle in engineering psychology is one that, if applied to design, will allow the user to perform more rapidly, more accurately, and with reduced attentional demand (so that other tasks can be done concurrently).

Of course, as we will see, many times these measures may trade off in practice. And furthermore, engineering psychologists are quite interested in many cognitive phenomena that are not *directly* reflected in performance, such as the degree of learning or memory of a concept, the quality of a mental model about a piece of equipment, the level of situation awareness about a process, the level of overconfidence in a decision, or the strategy of information processing that is invoked to obtain a given level of performance (e.g., serial versus parallel processing, speed stress versus accuracy stress, intuition versus analysis). Still, all of these cognitive phenomena may ultimately be expressed in some measure of performance in the workplace, and so long as they are, such **intervening variables** lie very much at the heart of human performance theory.

1.2 RESEARCH METHODS

Many different research methods can be employed to help discover, formulate, and refine theory-based principles regarding "what works" to support human performance. These can be roughly laid out on a continuum, from *laboratory experiments*, to *human-in-the-loop simulations*, to *field studies*, to actual *real-world observations*. The latter may come from *surveys* of users, *observational studies*, *case studies (analyses)* of **major accidents** and serious incidents. In some professions, such as health care and aviation, a corpus of minor incidents is available to create a large *database* of human performance issues, such as errors, that occur in the workplace. Each method has strengths and weaknesses. There is no "best" technique, because attributes, such as cost, fidelity to the workplace, and so forth, trade off along the continuum, and an effective engineering psychologist needs to be aware of the different methods, the various studies that have been conducted in a particular domain, and be able to interpret their results appropriately. We describe these research methods in much more detail in the next chapter.

1.3 A MODEL OF HUMAN INFORMATION PROCESSING

Knowing the different dimensions of performance (e.g., speed and accuracy) that can be measured in different research environments (e.g., lab, field studies) can assist the human factors engineer in understanding *how* performance is changed by system design or environmental differences. But such knowledge is not always sufficient for the engineering psychologist, who is interested in *why* performance might be changed. For example, a new interface for a car radio control might invite errors for the following reasons:

- The control cannot be touched without bumping another one.
- The control is too sensitive.

- The driver is confused about which way to adjust the control to increase frequency.
- The driver cannot understand the icon on the control.

The distinctions between the different psychological and motor processes affected by design are of critical importance because, on the one hand, they link to basic psychological theory, and on the other hand, they can help identify different sorts of design solutions.

A model of the stages of **human information processing** (Wickens & Carswell, 2022), shown in Figure 1.1, provides a useful framework for analyzing the different psychological processes used in interacting with systems, and for carrying out a task analysis, as well as a framework for the organization of the chapters in this book. The model depicts a series of **processing stages** or mental operations that typically (but not always) characterize the flow of information as a human performs tasks. We use as an example the task of driving toward an intersection. On the left of the figure, events in the environment are first processed by our **senses**—eyes, ears, touch, etc.—and may be briefly held in **short-term sensory store (STSS)** for no more than a second. Thus, the driver approaching the intersection will see the traffic light, the flow of the environment past the vehicle, other cars, and may be hearing the radio and the conversation of a passenger.

But sensation is not perception, and of this large array of sensory information, only a smaller amount may be actually *perceived*, for example perceiving that the light has turned yellow. **Perception** involves determining the *meaning* of the sensory signal or event, and such meaning is, in turn, derived from past experience (a yellow light means caution). As we see below, this past experience is stored in our **long-term memory** of facts, word meanings, images, and understanding of how the world works.

After perception, our information processing typically follows either (or both) of two paths. At the bottom, perceiving (understanding) a situation will often trigger an immediate response, chosen or selected from a broader array of possible responses. Here the driver may choose to depress the accelerator *or* apply the brake, a decision based on a variety of factors, but one that must be made rapidly. Then, following **response selection**, the response is *executed* in stage 4 of our sequence in a manner that not only involves the muscles, but also the brain control of those muscles.

But perception and situation understanding do not always trigger an immediate response. Following the upper path from perception, the driver may use **working**

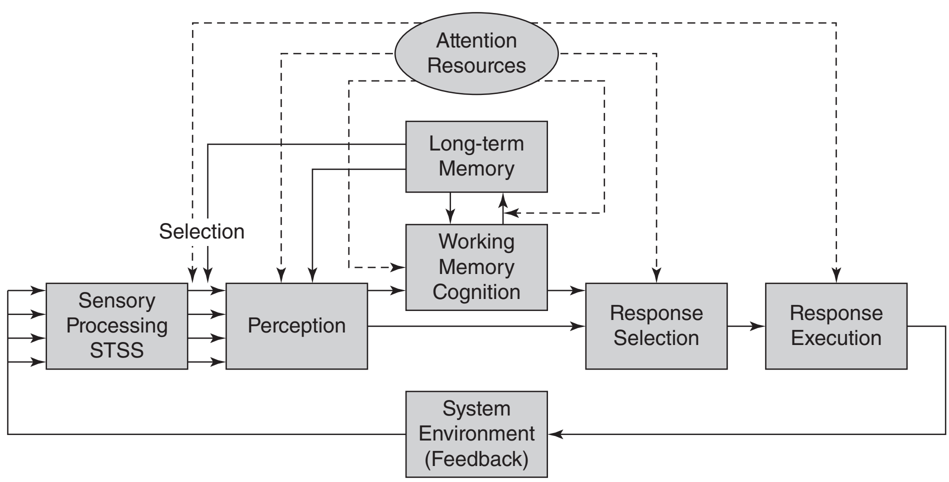

FIGURE 1.1 A model of human information processing stages.

memory, to temporarily retain the state of the light (yellow) while scanning the highway and the intersecting road ahead for additional information (e.g., an approaching vehicle, a possible police car). In fact, in many cases an overt action does not follow perception at all. As you sit in lecture you may hear an interesting fact from the lecturer, but choose not to take notes on it (no response selection and execution), but rather to ponder it, rehearse it, and **learn** it. That is, to use working memory to commit the information to **long-term memory**, for future use on an exam, or in applications outside of the classroom. Thus, the function of working memory is not just to store information, but also to think about it: the process of **cognition**.

At this point we note that the processes of perception and working memory are not as distinct from each other as the separate boxes would suggest. There is a fuzzy boundary between them, and hence this second stage, after sensation, but before response selection, can often be described as "cognition," generically describing the *interpretation of sensed material*, sometimes rapidly as the traffic light, and sometimes more slowly, as the idea presented by the lecturer.

To this four-stage + memory model, we add two vital elements, **feedback** and **attention**. First, in many (but not all) information processing tasks, an executed response changes the environment, and hence creates a new and different pattern of information to be sensed, as shown by the feedback loop at the bottom. Thus, if the driver applies the accelerator, this will not only increase the perceived speed of the car, but also may reveal new sensory information (the police car is suddenly revealed waiting behind a sign), which, in turn, may require a revision of the stop–go response choice.

Second, attention is a vital tool for much of information processing, and here it plays two qualitatively different roles (Wickens & McCarley, 2008; Wickens, 2021). First, as a *filter* of information that is sensed and perceived, it selects certain elements for further processing, but blocks others, as represented in Figure 1.1 by the smaller output (fewer arrows) coming from perception than input to it. Thus, the driver may focus attention fully on the traffic light, but "tune out" the conversation of the passenger, or fail to see the police car. Second, as a *fuel*, attention provides *mental resources* or energy to the various stages of information processing, as indicated by the dashed lines flowing from the supply of **resource**s at the top. Some stages demand more resources in some tasks than others. For example, peering at the traffic light through mist will require more effort for perception than seeing it on a dark night. But our supply of attentional resources is limited. Hence, the collective resources required for one task may not allow enough resources to be supplied to a concurrent one, creating a decrement in **multitasking**.

While Figure 1.1 provides a useful framework for conceptualizing information processing (and the organization of this book), it should not be taken too literally. Thus, although the primary operations associated with the different stages are *somewhat* associated with different brain structures (see Chapters 11 and 12), the association is not crisp; nor must the stages operate in strict sequence. Thus, the student in lecture may, in parallel, rehearse the lecturer's words and write them down. And, of course the major feedback loop at the bottom means that there is no fixed "start" and "end" to the information processing sequence. After all, a task might be initiated by an inspiration, thought, or intention to do something, originating from long-term memory, flowing to working memory and then to response, with no perceptual input whatsoever. Nevertheless, as we will see, the stage distinction is quite useful in analyzing tasks, describing principles and recommending solutions, and, in many cases, in developing the theories upon which engineering psychology is based.

The model shown in Figure 1.1 also provides a framework for organizing many of the chapters in this book. After describing research methods in Chapter 2, in Chapter 3 we discuss the more basic aspects of perception and the distinctions between detection and pattern recognition. In Chapter 4, we consider the attention filter, the selective aspects of attention. Chapters 5, 6, and 7 address the more complex aspects of perception and cognition that are relevant to the design of displays for space and spatial operations, including manual control (Chapters 5 and 6), and for language in Chapter 7. Chapter 8 addresses the role of cognition and both working memory and long-term memory and their relevance to learning and training. Chapters 9 and 10 address the selection and execution of action on the right side of Figure 1.1. In Chapter 9, this selection is the deliberative process of decision making, which also heavily involves memory. In Chapter 10, the selection represents more rapid actions such as those taken at the traffic light. Chapter 11 addresses the issues of multitasking as various combinations of stages and multiple tasks need to compete with each other for the limited fuel of attention resources. In Chapter 12, we address issues of mental workload and stress, as all stages work in concert to carry out tasks. In Chapter 13, we consider issues of human–automation interaction, as artificial intelligence tools are designed to assist or replace the human with information processing and cognition. A final short Epilogue summarizes some key themes.

1.4 PEDAGOGY OF THE BOOK

There are a few critical features that we would like to highlight to our readers before they jump into the chapters that follow.

First, we have tried to cite a large amount of literature to indicate the wealth of research that lies behind the concepts, principles, and findings that we present. In doing so, we have tried to emphasize "take-home messages" from the collective body of research, more so than the specific methods and findings from a single study. As a consequence, we may have glossed over details of particular studies, but we think we have been true to the studies' main conclusions. Our extensive reference list will allow the curious reader to delve in greater detail for any specific topic he or she desires. Many former students using previous editions of this text are now engineering psychologists or human factors practitioners themselves; a common remark is that the book remains a useful reference for their professional career, long after they have taken the course.

Second, the reader will detect a rich network of cross-references between chapters. We hope that any distraction this may cause will be offset by a realization of the complexity of human performance, and how interwoven the performance components are in their application to the workplace. As just one example, we find that the cognitive phenomenon of **overconfidence**, keeps reappearing in different guises, across different stages and types of human performance and cognition (and therefore different chapters).

Third, the reader will note the distinction between our use of *italics* and **boldface**. Boldface is meant to highlight new key terms or concepts, which can form the basis of a glossary, whereas italics are simply used to *emphasize* a word or phrase that should already be familiar to the reader, either in common language usage, or from boldfacing in a prior chapter.

Finally, as befits the distinction between engineering psychology and human factors, we do tend to emphasize more the general principles that support effective human performance (Peacock, 2009), rather than the specific design examples (although we do not entirely neglect the latter). It is hoped that the material in this book provides an effective "hand-off" to those truly interested in design applications, who can then follow these up in more applied human factors treatments (e.g., Salvendy & Karwalski, 2022; Lee et al., 2017; Peacock, 2009). We would be delighted if students grabbed a particular principle, looked around their environment and explored their experiences to see how the principle might have been violated in "the real world".

In summary, we hope that our approach provides a distinctive counterpoint to the existing literature. The audience is intended to be upper-division undergraduates or graduate students, with a background in human science (e.g., psychology, cognitive science, kinesiology) or applied science (engineering, computer science). The science student may be more interested in how that which is known about information processing and human performance can be applied in real-world situations. The engineering student will likely be more interested in knowing more about psychology and its theory and why it matters to the design of engineered products and systems. We hope that both classes of students find the book has an appropriate balance of these qualities.

Key Terms

applied psychology *1*
attention *5*
cognition *5*
cognitive engineering *1*
cognitive ergonomics *1*
computational models *3*
engineering psychology *1*
ergonomics *1*
experimental control *2*
feedback *5*
filter *5*
fuel *5*
human information processing *4*
human performance *2*
intervening variables *3*
learn *5*
long term memory *4*
major accidents *3*
mental resources *5*
meta-analyses *3*
overconfidence *6*
perception *4*
response selection *4*
senses *4*
short term sensory store (STSS) *4*
working memory *5*

Bibliography

Buxton, W. (2007). *Sketching user experience*. San Francisco: Morgan Kaufmann.

Durso, F., & Boehm-Davis, D. (2014). *APA handbook of human-systems integration*. Washington, DC: American Psychological Association.

Fitts, P. M., & Jones, R. E. (1947). Analysis of factors contributing to 460 "pilot-error" experiences in operating aircraft controls (pp. 332–358). Wright-Patterson Air Force Base, OH: Aero Medical Laboratory.

Jacko, J, (2012). *Handbook of human-computer interaction* (3rd Ed.). Orlando: CRC Press.

Jenkins, D., Stanton, N., Salmon, P., & Walker, G. (2009). *Cognitive work analysis: Coping with complexity*. Burlington, VT: Ashgate.

Lee, J., & Kirlik, A. (2013). *Handbook of cognitive engineering*. Oxford, UK: Oxford University Press.

Lee, J., Wickens, C., Ng-Boyle, L., & Liu, Y. (2017). *Designing for People*.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology, 1,* 5–61.

Peacock, B. (2009). *The laws and rules of ergnomics in design*. Santa Monica, CA: Human Factors Society.

Salvendy, G., & Karwalski, W. (2022). *Handbook of human factors*. Orlando: CRC Press.

Vicente, K. J. (1999). *Cognitive work analysis*. Mahwah, NJ: Erlbaum.

Wickens, C. D. (2021). Attention: Theory, principles, models and applications. *International Journal of Human Computer Interaction, 37*(5), 403–417. https://doi.org/10.1080/10447318.2021.1874741

Wickens, C., & Carswell, C. M. (2022). Information processing. In G. Salvendy & W. Karwalski, (Eds.), *Handbook of human factors*. Orlando, FL: CRC Press.

Wickens, C. D. & McCarley, J. M. (2008). *Applied attention theory*. Boca Raton, FL: CRC Press.

2 RESEARCH METHODS

2.1 OVERVIEW OF THE ENGINEERING PSYCHOLOGY RESEARCH PROCESS

In the profession of human factors, we ask how researchers go about making an inference about "what works and how well it works," in design, training, procedures, and selection for real-world tasks (Lee Wickens, Boyle, & Liu, 2018). Figure 2.1 shows a data source at the bottom (human factors is an empirical science); we ask questions of those data, often questions about cause and effect, driven by applications, and we assess the reliability of those answers, and then use them to improve human–system interaction.

In engineering psychology, we modify and elaborate on this general model somewhat. Our focus is now more tied to the following question: How do we derive theory-based psychological principles that we are confident will make a difference in real-world applications? That is, what principles will generalize to real-world tasks, and, ideally, across different applications? For example, what principle regarding the optimum level of human-automation authority will apply equally to self-driving cars and to the flight deck of the modern aircraft? To accomplish these goals, we refer to Figures 2.2 and 2.3, which provide the framework for the material in this chapter.

Across the top of Figure 2.2 are the three most important elements of engineering psychology. Highlighted in the largest font are the PRINCIPLES of design and training that should be applicable in the world outside the laboratory and should be grounded in the fundamental theories of psychology. These theories, in turn, capture the essence

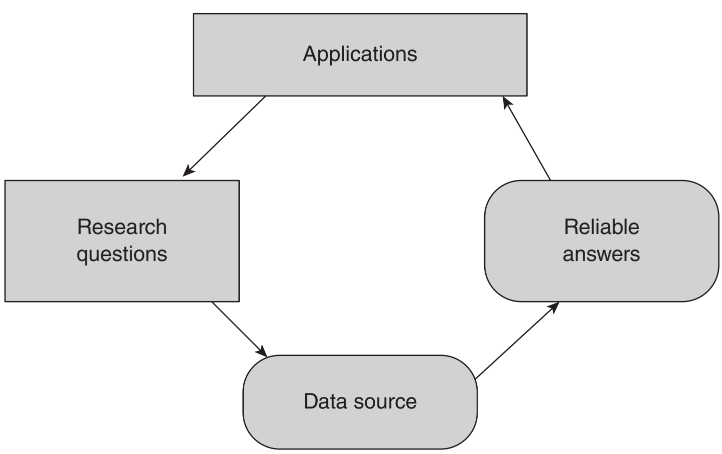

FIGURE 2.1 Flow of human factors.

DOI: 10.4324/9781003177616-2

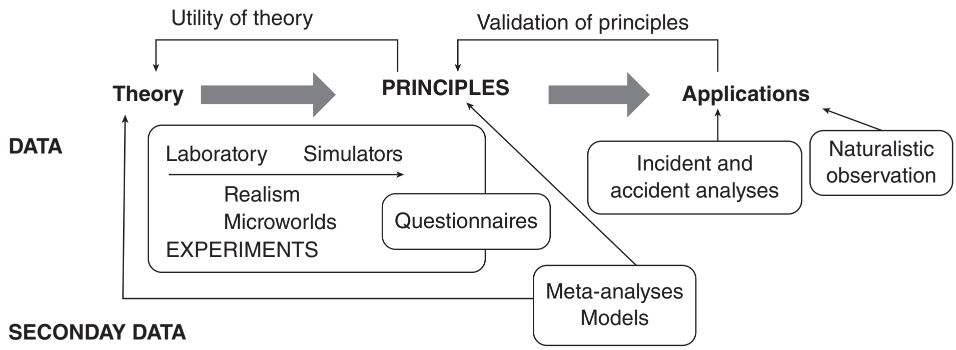

FIGURE 2.2 The human factors framework.

of human performance and cognition. The feedback loops at the top indicate successful application. Successful applications, in turn, provide support for the theory that spawned those applications.

The next row of the figure is the vital path to validation. We need *data*, and arrayed across the row are the different sources of data. On the left are experiments, most often conducted in the laboratory, and well known to any of us who have taken or taught courses in experimental design. Their data serve both theory and principles. Thus, an experiment may be used to help validate a theory of task switching (Trafton & Monk, 2007), but also to validate a principle of how to train people to more safely multitask in the vehicle or airplane (Chapter 11; Loukopoulos, Dismukes, & Barshi, 2009). An experiment may help to validate a theory of object perception, but also to validate a principle of designing a display of two variables to configure as a rectangle, rather than two bar graphs (Chapter 5). An experiment may help validate a theory of memory interference, but also a principle of spacing practice in order to maximize learning (Chapter 8).

A major challenge to the engineering psychology researcher is to achieve the necessary realism of experimental laboratory data sources relative to the ultimate applications that they are destined to serve. In order to seek such realism, we move to the right along the DATA row, from low-fidelity experiments to what we loosely label as simulators. Because these simulators are also hosts for experiments, they require the same attention to experimental control (see below), but also contain more features that match the real-world application to which their results are supposed to generalize. A low-fidelity study of cell phone distraction in the vehicle can use a basic tracking task to simulate driving (Drews, Johnston, & Strayer, 2003), but its results will likely be more valid if carried out in a driving simulator (Caird et al., 2018); and even more so if the simulator has the realism of a full simulator with 180-degree vision, as compared to a desktop computer.

A close cousin to simulator realism (or **fidelity**) is *environmental fidelity*, or the degree of realism of the experimental environment to the application. Thus, for example, data to support a principle of learning may come from a laboratory experiment, or from a controlled intervention in the actual classroom, in which half of the students are given one kind of study strategy or quiz schedule, and the other half a different type. A final element in the EXPERIMENTS box in Figure 2.2 is the **microworld**. This is typically a small-scale simulation, usually on a desktop or laptop computer, designed to emulate a fairly complex dynamic process, like distilling a chemical or managing a production (Gonzales, Fakhari, & Busemeyer, 2017). The emphasis in designing microworlds is to accurately capture the perceptual and cognitive *processes* involved, more so than to create perceptual and *physical realism* of the real system, which is often the goal of simulators.

Questionnaires are another source of data that may be administered in experiments or simulators, and also, in human factors, to real-world users of a system. This dual source of questionnaire data is reflected in the figure by being both inside and outside the EXPERIMENTS box. As distinct from most data sources, these questionnaire data are not measures of performance speed or accuracy, but are typically rated opinions or subjective scales of such variables as situation awareness, mental workload, or trust. These questionnaires could also include open ended questions or other qualitative methods (Hoffman, 2008).

Naturally, designing experiments creates a trade-off with either simulator or environmental realism. Greater realism imposes greater logistics complexity, often greater expense, and sometimes leads to a loss of experimental control, as we discuss below.

In our journey across data sources, we then make a discrete jump in realism as we move to the far right of the DATA row. In the box to the right, two further sources of data can be provided from the real-world applications domain itself. The first of these are reports of accidents and incidents experienced by operators. The difference between accidents and incidents, sometimes formally defined in safety-oriented professions like aviation, is whether there is injury or fatalities to personnel involved or damage to equipment (accidents) or not (incidents). The latter case involves an operator report of something that was "done wrong" but did not produce an accident. Accident reports, such as those conducted and written up by the National Transportation and Safety Board, are formal and involuntary. In contrast, incident reports, such as those within the Aviation Safety Reporting System (ASRS), are generally voluntary and anonymous, a feature incorporated in order to encourage operator submission and therefore populate the database, without the operator's fear of recrimination.

Finally, although accident and incident reports provide data to the researcher regarding what went wrong, and implicitly or explicitly how to fix it, naturalistic observations of the professionals at work (including their descriptions of what they are doing) are more often indicators of "what went right." Such descriptions, if provided by experts, can form an important part of the signatures for expertise in an area.

There is clearly no single source of data that is best for forming conclusions about applied research on what principles are most effective. Ideally, researchers should try to tap more than one point along the continuum, and look for common trends across them, as, for example, has been done in understanding the appropriate level of automation to keep the human sufficiently in the loop, so that she may respond appropriately if there is a failure (Onnasch, Wickens, & Manzey, 2014; Kaber & Endsley, 2004).

There are also two important trade-offs between different points along the continuum. First, the more complex simulations to the right involve more realism (good), but also more logistics and expense (bad). Second, the greater realism and complexity that mimics the real-world system often involves fewer participants because of their limited availability. Think of the scarcity of the well-trained transport pilot flying an advanced experimental simulator, compared to the undergraduate participant flying a desktop simulation. Fewer participants thereby diminishes the statistical power, an issue we discuss in detail in Section 2.5.

In the row below the DATA row, we identify two further sources of input to our researcher, which we label "Secondary data." They are secondary, because both rely upon raw data, typically from experiments or simulations, but are not themselves the direct source of data. The first of these is the **meta-analysis**; quite literally, this is an analysis of analyses, each of the latter typically the output from a single experiment. For example, this might be as straightforward as presenting the mean difference, from several experiments, each of which has established a mean difference between an auditory and a visual presentation of instructional information. More typically, the meta-analysis is accomplished by aggregating (usually averaging) the *statistical effect size*, the proportion of variance accounted for by a particular experimental manipulation (Rosenthal, 1991; Borenstein, Hedges, Higgins, & Rothstein, 2009), a concept that we will elaborate below. Meta-analyses have the advantage of pooling a lot of data, and hence reaching more stable conclusions about the magnitude or direction of an effect than any single experiment. But the validity of the meta-analysis depends critically on the quality of the studies that went into it.

The second source of secondary data is the **computational model**, which uses the output of a formula or a computer simulation to provide the researcher with an estimate of what the data would say (Pew & Mavor, 1998; Byrne, 2014). For example, the researcher could use a computational model called Fitts' law to provide an estimate for the time for an operator to reach and depress a key on a keyboard (Fitts & Posner, 1967; Card, Newell, & Moran, 1986), and this model-based answer would be much more efficient for the researcher than having to do an experiment for such a determination. Like the meta-analysis, the model is only as good as the data upon which it is based. The model itself needs an empirical foundation, and will benefit from several empirical validations. We discuss models further in Section 2.6.

Models and meta-analyses can often work hand-in-hand in applied psychology. Because the meta-analyses are typically quite reliable and stable in the quantitative estimates of performance that they offer, those estimates can be imposed in models to predict performance that are then well validated. As an example, the conclusions of a meta-analysis of the effects of sleep deprivation on performance have been incorporated into a computational model of that process (Wickens, Hutchins, Laux, & Sebok, 2015).

Figure 2.3 simplifies and aggregates all of the sources of data in Figure 2.2, but now highlights four elements that are vitally important to the ability of one or more of the data sources above to contribute to the engineering psychologist researcher's task of integrating theory, principle, and application. We describe each briefly here, but elaborate upon them in separate sections below.

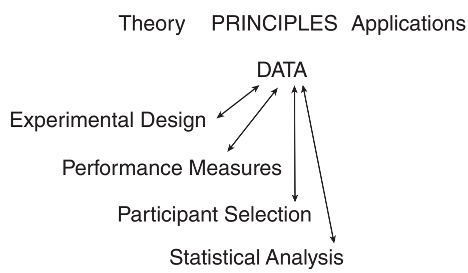

FIGURE 2.3 Critical elements of data in engineering psychology research.

For experiments, **experimental design** describes the conditions used to establish an effect or a causal relation between an independent and a dependent variable. Closely related, *performance measurement* describes the dependent variable itself. For example, do we measure the speed or accuracy of performance, or both, when we try to draw inferences about human information processing in an applications-relevant task? Any engineering psychology experiment must involve *human participants* (or at least workers, see Helton, 2009). Who are they? College sophomores fulfilling psychology class requirements? Paid volunteers on a signup sheet or on the Internet? Professionals trained in the task that may be simulated for the experiment? And how many participants? This last question leads to the final element: *statistical analysis*. How do we establish the validity of the experimental data to generalize (1) to people as a whole (the larger population outside of those sampled for the experiment) and (2) to the population of professionals to whom we wish our results to generalize (e.g., healthcare professionals, licensed drivers, intelligence analysts). We now elaborate on each of these elements.

2.2 EXPERIMENTAL DESIGN

2.2.1 Two-Condition Designs

Typically, engineering psychology experiments are designed to establish "if there is a difference" (in a dependent variable) due to some critical independent variable of importance in the real world, for example, sleep deprivation or display modality. In the simplest case, this difference is just between two conditions. For example, a "treatment" condition, such as an automation support for a task like steering a car, is compared to a "control" condition, such as manual steering. Often, two treatment conditions are compared with each other, such as an auditory (voice) versus a visual display of navigational information to the driver. The typical statistical test for the two-level comparison is the **t-test** (although some advocates prefer a confidence interval of a mean difference or a Bayes factor with some kind of cutoff, to be discussed below).

Such a two-condition design requires at least one further decision by the researcher: Should different people receive the two conditions (*between-subjects design*) or should each participant in the experiment receive both conditions (*repeated-measures design*). Each type has different costs and benefits. The primary benefit of the repeated-measures design is that there is less variability in the effect of the treatment because the same person is receiving both conditions. Less variability means more *statistical power* and a greater likelihood of finding a significant difference (see Section 2.5). However, the repeated-measures design can have an inherent problem. If the same person receives both conditions, then if he receives A before B, the participant may be better practiced at B than A, and hence any difference favoring B could be due to practice, rather than to its inherent superiority to A. We call this a **confound**. Alternatively, if the experiment is long, subjects could be more fatigued when doing B, and hence do worse as a result. Both fatigue and practice confounds can exist in the same experiment, and both will introduce an unwanted bias in favor of one condition or the other.

To address the confound problems of a repeated-measures design, it is possible to alternate conditions A and B frequently, so that any learning or fatigue effects would be washed out. Alternatively, a *counterbalanced design* will divide the subject population in half, giving one group the order AB and the other group the order BA. This also averages out the sequential biases when the two groups' performances are averaged. When participants are scarce, the repeated-measures design is preferable because one gets more statistical power (higher *N*) in each condition. Although counterbalancing is useful, it does not remove the possibility of asymmetrical carryover effects. The engineering psychologist should give their experimental design extensive consideration *prior* to deploying an experiment.

2.2.2 Details and Qualifiers of the Effect: More Than Two Conditions and Factorial Designs

In order to generalize to the real world beyond the experimental environment, a simple comparison between two conditions rarely makes sense because whether a difference is found or not may be based on contextual factors or other factors, likely to influence performance beyond the laboratory or simulator. Consider two different cases of an experiment on vehicle automation. First, it may be the case that the extent to which automation improves performance above pure manual operation (our control condition) depends on *how much* automation there is. It may provide only a steering assist, or, alternatively, both a steering and a speed control. This produces a *three-level design*: control, steering only, steering and speed. The statistical t-test is no longer appropriate, because we are no longer looking at a single difference, but rather we are looking at the net effect of three differences (the pairwise comparison of each of the three conditions). This then is the *multilevel design*.

Second, we might ask whether automation is more, or less, beneficial when a driver is just driving, or is driving and multitasking. Here we could cross the manipulation of automation (present or absent), with the manipulation of dual-task loading (absent or present), creating a 2 × 2, four-condition **factorial design**. The typical statistical test for the multilevel or multifactor design is the **analysis of variance**, or **ANOVA**.

Factorial designs provide two great advantages. First, they are efficient. One can examine the effects of two independent variables (automation and dual-task loading) at once within a single experiment. Second, and more important, they enable the investigator to examine the presence of ***statistical interactions*** (Figure 2.4), to ask questions such as the following: Is the advantage of automation enhanced, or perhaps diminished, under dual-task loading (i.e., when multitasking)? As a result, one can determine the generalizability of an effect of one independent variable across levels of another. If the effect is neither enhanced nor diminished, but is constant across levels of the other independent variable, we say the two variables are *additive* in their effect. This would be the case if the benefits of automation were the same whether multitasking or not. The general questions asked of, and answered by, interactions are of the form: To what extent does the effect of variable A depend upon the level of variable B? In Figure 2.4, any of four different statements might characterize this "effect on an effect" manner of describing an interaction (depending on more specific contrasts between pairs of points):

"Dual task loading harms driver performance only in manual but not automated mode"
"Multitasking harms manual more than automated driving"
"Imposing automation only helps in dual task conditions"
"Imposing automation helps more in dual than single task conditions"

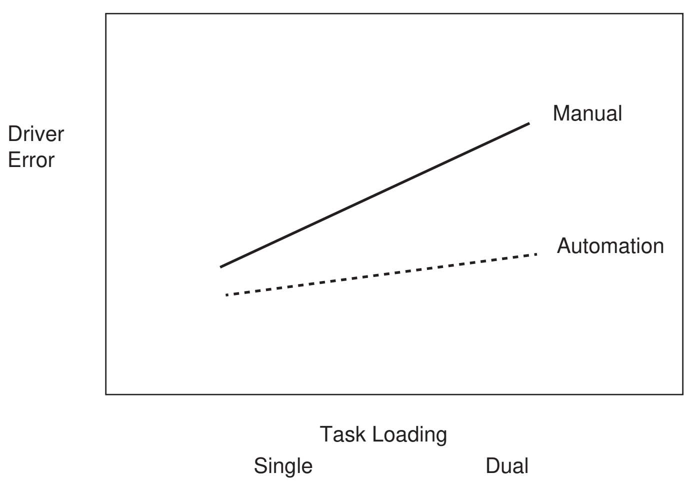

FIGURE 2.4 Data from a factorial design showing an interaction.

The 2 × 2 design is the simplest factorial design. To increase complexity one might, for example, create a 3 × 2 design by crossing the three levels of automation with dual-task loading. Furthermore, one could create any number of levels across each of the two factors. We may also propose a *three-factor design* if, for example, we ran our original 2 × 2 design with both fatigued and well-rested subjects, creating a 2 × 2 × 2 design. And, then we could extend this to a 2 × 2 × 3 design, and so forth. The possibilities are unlimited!

As we discussed with the simple two-level design, we can again here distinguish between a repeated-measures design and a within-subjects design. However, there is now a third, compromise alternative called a *mixed design* or *split-plot design* in which one, or some of, the factors are repeated measures and other(s) are between subjects. As an example, suppose, because of a desire for higher statistical power, we wanted to vary automation level within subjects, but we also wanted to examine the effects of age on automation benefits. Older versus younger subjects must, by definition, become a between-subjects variable.

2.2.3 The Continuous Independent Variable

Whereas ANOVAs are applied to designs with fixed, categorical levels of the independent variable (e.g., display modality), researchers will often design experiments with interval- or ratio-scale levels of the independent variable. For example, this might be an experiment examining the effect of degrees of visual separation from the fovea on response time to warning lights. Here, the data will be subjected to **regression analysis** rather than to an ANOVA to establish the strength and significance of effect. Finally, in some cases we cannot control or select the values of either of the variables involved in a two-way relationship, even if that relationship is part of our experimental question. An example might be: How do differences in situation awareness relate to differences in workload across our participants? Here, our design uses the *product moment* ***correlation*** as the analysis tool to determine, for example, the extent to which people with higher situation awareness also experience higher workload. As the above cases make clear, experimental design and statistical analysis are intimately linked.

2.3 PERFORMANCE MEASUREMENT

The three most important measures of performance in engineering psychology are considered to be speed (often measured by response time, or RT), accuracy (percent correct or its converse, error rate), and, to a slightly lesser extent, mental workload or attention demand (see Chapter 12). Each of these has an obvious polarity in terms of which is better for overall performance.

As seen in Figure 2.2, performance measures are often augmented by questionnaires or subjective ratings of key cognitive variables such as workload, situation awareness or trust.

Although all variables need to be addressed statistically (see Section 2.5), it is important to keep in mind the *practical* significance of results. In applied research, an effect of perhaps 50 msec in response time may not be of great practical interest, even if it is statistically significant: but a 5% increase in error rate (say from 1% to 6%) could be of critical applied importance. Practical significance is not the same as **statistical significance**.

It is also important to keep in mind the possible trade-offs that result as an independent variable is manipulated, particularly the tradeoff between speed and accuracy (see Chapter 10). For example, changing modalities of a navigation display from visual to auditory may shorten RT but also increase error rate because of the forgetting of auditory information. Another example is an *effort–performance trade-off*, where participants in one condition expend greater effort to increase performance than in another condition. The increased subjective effort results in increased strain, or stress responses, which may have consequences if the experimental results are generalized to actual workers. The manipulation may have increased participant motivation, not performance itself per se, or, in other words, the performance improvement is not without an otherwise hidden cost. When such trade-offs appear, they should be clearly articulated. When possible, multiple dependent measures are useful, for example, speed, accuracy, and subjective "feelings."

In addition to these three general measures, other performance measures are sometimes of value, often derived by combining variables. As we will see in the next chapter on signal detection theory, a very important measure is the "bias" or tendency to report all signals, which is quite different from the accuracy measure of signal detection which reflects the ability to discriminate those signals from background noise.

2.4 PARTICIPANT SELECTION

For simple laboratory experiments, to the left of the top row of Figure 2.2, the prototypical "college sophomore" is often adequate as a participant to generate data. But as described in Section 2.1, the tasks here are often somewhat simple compared to those of the microworld or simulator, and therefore their generalizability to a real-world task may be questioned.

As we move to the right along this DATA row of Figure 2.2, particularly to the simulator, two approaches are typically required. First, we can train participants to acquire more of the specialized skills of the professional that must be harnessed to perform the simulation. Such a requirement often exceeds the number of volunteer hours available for academic course credit in a typical university "subject pool." It hence may require paying volunteers. But such payment has the added advantage of allowing some pay-for-performance bonuses in order to motivate high attention to the task.

Second, we can solicit actual professionals in the task that we have simulated; for example, licensed pilots for the flight simulator experiments. These people already possess some of the requisite skills, and need less training prior to collection of the experimental data. Such personnel need not be true experts at the task. Often students within the professional training program for the skill in question will be adequate.

One challenge in both of these approaches relates to limited participant availability. Applying the first approach, people simply may not be able to commit to the several hours of experiment time that the extra training will impose. For the second approach, those professionals may simply not be accessible, and even if they are, may be reluctant to allocate that time away from their professional careers. Thus, in both cases there is the danger of reducing the sample size, or N, in the experiment, in a way that reduces the statistical power to find the real effects that may contribute knowledge to improving performance in applications. We address this issue further in the next section. However, one solution to these challenges is to augment the professional sample size with a larger sample of non-professionals, or student trainees. If the pattern of results is roughly equivalent between the two groups, they can be combined in the larger sample statistical analysis.

2.5 STATISTICAL ANALYSIS

As Figure 2.3 makes clear, statistical analysis is a nearly essential stage in establishing the extent to which experimental results can generalize to and predict performance effects in the world beyond the laboratory. We assume that our reader is familiar with traditional statistics as it is typically taught in undergraduate psychology and engineering programs; that is, statistics based on what is called **null hypothesis significance testing (NHST)**. This typically yields the F-test or t-test to establish the statistical significance of a difference or effect or trend. This approach has a lot to offer; however, recently behavioral scientists have identified some important flaws (Cumming, 2012; Dienes & McLatchie, 2018), and these flaws gain in their importance when the statistics are applied to human factors science and engineering psychology. In the following section, we lay out these problems and propose solutions to overcome them, including a Bayesian approach (Lee & Wagenmakers, 2013). This material is based heavily on the chapter "Commonsense statistics in aviation safety research" (Wickens & McCarley, 2017).

What NHST provides. The statistics of significance provided by NHST depend critically upon the *statistical power* of an experiment, which, in turn, depends on three factors:

1. The **effect size**, or the difference between means that are being compared.
2. The **variability of data**, typically assessed by the standard deviation of the observations around the mean.
3. The **sample size**, or N, of the experiment.

To the extent that variability is low and sample size is large, statistical power is high, and to the extent that the effect size is also large, then the statistical significance of the NHST is more likely to emerge. This significance in a t-test or F-test has often been operationally defined as a ***p*-value** (e.g., $p < 0.05$). While still conventionally used, there are four important problems with this reasoning, and reliance on the p-value to signal the importance of a finding, which we describe as follows:

2.5.1 Problem 1: The All-or-None Interpretation of .05

Researchers often report that if p is less than .05 ($p < .05$) they have found an effect, and if $p > .05$ "there is nothing there." In fact, the p-value was never intended by Fisher (1935), its developer, to convey such all-or-none, black-and-white thinking, but rather to provide a probabilistic indicator of the *continuously distributed degree of evidence* for an effect. Thus, for example, the gain in evidence from .08 to .06 is just as great as from .06 to .04, even as conventional thinking is that the latter is much greater (or much more important), because it defines greater or less than the "magic .05." This faulty thinking flies in the face of evidence-based science.

There is of course some rationale for establishing a standard, as .05 has been, but the idea that one should treat .06 as "nothing there" can do a major disservice to the accumulation of scientific wisdom across multiple experiments. We must also keep in mind that the high N that is often necessary to achieve sufficient statistical power to find the .05 effect is often difficult to obtain, as we discussed above in the context of the valuable experiments conducted with highly trained professionals. A .06 effect observed in establishing the better performance with a new warning system using 10 highly skilled professional pilots is of great benefit for the advancement of aviation safety. This .06 effect should not be discussed as "nothing there."

Besides leading to dismissal of effects that don't meet the .05 criterion, this black-and-white thinking also allows a spurious effect that sneaks under the $p = 0.05$ cutoff to live on, in the literature and in application, indefinitely. Having achieved statistical significance once, the "finding" is often deemed real, and failures to replicate it are often blamed on poor method or low statistical power. In truth, a p-value in the range of 0.05 is at best tentative evidence of a replicable effect (Cumming, 2012, 2014). An effect of

$p = 0.06$ or so should not be dismissed, nor should an effect just under $p = 0.05$ be treated as conclusive. Several meta-analyses have revealed the extent to which effects of $p < .05$ fail to consistently replicate (Wetzels et al., 2011; Ioannidis, 2005, 2008, Open Science Collaboration, 2015).

2.5.2 Problem 2: NHST Is Biased Toward the Status Quo

Table 2.1 presents the standard decision matrix underlying NHST. Across the top is the "ground truth" state of the world that the researcher wishes to discover. For example, this may be the truth of whether the augmented auditory navigational display in the car improves driving safety. We run an experiment to test that possibility, compute statistics, and then derive a conclusion based on whether our p-value falls below alpha (α), which is typically set at .05. Our potential conclusions are represented in the two rows of the table: accept the null hypothesis of "no difference" or reject the null hypothesis and assume that there is a difference. That is, the effect in our experiment is of sufficiently large N and the statistical power is sufficiently high that what we found in the laboratory is likely to sustain in the world beyond the laboratory. Factorially combining the two states of the world by the two potential conclusions, two forms of statistical error are possible. A *type I statistical* error, in the top right cell of the matrix, occurs when we erroneously conclude there is an effect where in fact there is none. A *type II* error, in the bottom left cell, occurs when we fail to detect an effect that does, in fact, exist.

TABLE 2.1 A conventional table of statistical decisions within NHST

		State of the world	
		New auditory display improves safety	**New display does not improve safety**
Experimental results	Disconfirm H_0 ($p < .05$)		Type I error. Strongly discouraged.
	Fail to disconfirm H_0 ($p > .05$)	Type II error. Considered more tolerable than a type I error.	

Convention has established that we keep the type I error rate no higher than 5% ($p < .05$), which is why we are so reluctant to say that a $p = .06$ effect is a "real effect." But in doing so, we completely ignore the probability of the type II error, which, unless we have high statistical power (resulting from high N and/or low variability) is usually a lot higher than .05 (e.g., around 0.20).

This state of asymmetric concern for type I over type II errors is a natural outgrowth of concern in basic sciences that false-positive discoveries are more costly than false negatives: and obviously, it is counterproductive for researchers in any domain to assert effects that turn out to be untrue. As the recent "crisis of replication" in psychology and other sciences has shown (Pashler & Harris, 2012), non-replicable effects undermine confidence in research, and ultimately make it difficult to convince the government, industry, and the public at large that they should support our studies and trust our claims. But should the imbalance of concern for type I versus type 2 errors be the same in applied, safety-related research as in basic science? As argued below, it should not be, and this is the basis of the third problem with NHST.

2.5.3 Problem 3: Conventional NHST Practice Considers Values in Decision Making Bluntly and Inflexibly

As noted above, Fisher argued that alpha should not be inflexible. Fisher (1935) noted that, "It is usual and convenient for experimenters to take 5 per cent as a standard level of significance," but acknowledged that the choice is subjective and sensitive to context. Neyman and Pearson (1933), the founders of NHST, wrote of type I and type II error: "in some cases it will be more important to avoid the first, in others [to avoid] the second . . . just how the balance should be struck, must be left to the experimenter." Modern NHST practice, unfortunately, generally ignores this advice to be flexible, and assumes an α of .05 almost universally. This practice does not support adjusting the criterion when type II errors (rejecting a true safety enhancement for example) may be very costly.

To illustrate the problem with this approach, Table 2.2 presents a classic decision table from expected value theory (see Chapter 9), populated by the specific characteristics of our display design example. It is similar in some respects to Table 2.1, but distinct in others. The two possible ground-truth states of the world are again shown in the two columns, and the two rows again represent potential decisions. Here, though, these are not the researchers' decisions to reject or accept the null hypothesis, but the research customers', or *consumers'*, decisions to either implement the new display or reject it. The consumers' decision is very different from the researchers'. Most importantly, the consumer's decision considers the context-specific costs and benefits of different outcomes, particularly for the two types of decision errors. Rather than simply assuming that type I errors are worse than type II errors, as conventional NHST practice does, it attaches precise payoffs to various decision outcomes, including the costs of developing and marketing and the costs of accidents that might be avoided with the new display. With these explicit payoffs in mind, accompanied by some estimate of the effect size under study, the decision-makers can select an α-level appropriate to the context given the statistical power of the experiment, trading off the costs and benefits of type I and type II errors in order to maximize the expected value of their decision.

TABLE 2.2 The classic expected value decision matrix

		State of the world	
		New display improves safety [*p(H)*]	New display does not improve safety [1 – *p(H)*]
Consumer's decision	Adopt the display	Value of collisions avoided minus cost of adoption	Cost of adoption
	Do not adopt the display	Cost of avoidable collision	No cost

2.5.4 Problem 4: NHST Does Not Consider the Prior Probabilities of the Null and Alternative Hypotheses in Decision Making

In the real world of research, there may be occasions when one has a strong *a priori* belief that a particular effect will come out in one direction rather than the other. Indeed, this is what is typically expressed in the hypotheses of an experimental write up, placed just before the methods are introduced. Such an *a priori* belief is often based on prior experiments that have demonstrated the effect in question; for example, that an auditory navigation display will be safer, in the visually dominant driving environment, than a visual display. The NHST approach does not allow for such an *a priori* biasing to permit less evidence to confirm the predicted direction (auditory better than visual) than the unpredicted one.

This incorporation of prior probabilities or beliefs in interpreting effects is a hallmark of what is called *Bayesian reasoning*, to be described in more detail in Chapter 9 (Berger, 1985). Bayesian reasoning has found a home in **Bayesian statistics** (Lee & Wagenmakers, 2013; Dienes, 2011, 2016), and in particular what is called the **Bayes factor**, as an alternative to the more traditional *p*-value (Wetzels et al., 2011). This approach essentially allows a symmetric test of evidence for the null versus the alternative hypothesis, not the asymmetry bias of the NHST, and then establishes, along a continuum defining the Bayes factor, the *degree of evidence* for one, the other, *or neither*. This latter alternative allows the investigator to say: "we just don't have enough evidence yet to either accept or reject the null hypothesis: more data are needed." The one key element required by Bayesian statistics is an explicit statement of the size of the alternative hypothesis effect, and the distribution of that effect. We will see in Chapter 3 the distinct similarity of this approach to that in signal detection theory.

In summary, there are two general points to be made here. First, the consumer of the research, who ultimately decides whether to implement potentially safety-critical procedures, needs more information from the researcher than simply the "reject/don't reject the null hypothesis" output of a statistical decision rule, an output that implicitly removes this responsibility from the consumer of the research making policy or design decisions. Second, the application of a binary decision rule without adequate statistical power or consideration of payoffs and prior probabilities produces an inherent bias *against* adopting procedures or equipment that might improve safety.

2.5.4.1 WHAT IS TO BE DONE? Below, we outline two general categories of remedies for this state of affairs: (1) changes to how the researcher should approach experimental design and analysis and (2) changes to the way data are presented in written reports and articles. These are elaborated by Wickens and McCarley (2017; Wickens, 1998).

2.5.4.2 DESIGN AND ANALYSIS

Increase statistical power. As noted above, by running more subjects or eliminating sources of unwanted variance, we reduce statistical noise and increase statistical power. The higher power will allow us to reduce the probability of a type II error without a corresponding increase in type I error (which, by definition, is what happens when we simply raise the α-level for significance; e.g., to *0.10*). Of course, as we have noted, in some high-fidelity simulations increasing power by increasing *N* is simply impossible.

Careful framing of experimental questions. When statistical power cannot be increased by increasing *N*, the experimenter may benefit in power by carefully framing the most important research question before an experiment is conducted. In particular, a *one-tailed t-test* can be employed in comparing two conditions; for example, a currently used display and one designed to improve safety or otherwise enhance performance. Your prior hypothesis is that you will find better performance (shorter RT and/or higher accuracy) with the new display. You do not care if it shows equivalent or worse performance. This allows you to use a one-tailed test for significance, which greatly increases statistical power. With ANOVAs, if you plan your comparisons in advance to determine whether a particular effect is significant in a particular direction it allows for more statistical power than a post hoc test to see whether any differences are significant. However, you are limited in the number of independent planned comparisons you can propose (Hayes, *1994*).

Formulate an alternative hypothesis. A third approach, inherent in the Bayes factor discussed above, is to formulate a specific **alternative hypothesis** (to the null hypothesis of no effect) that specifies the size of the effect that you would consider important, or that you would allow you to conclude that your innovation (e.g., the new display) would actually improve performance beyond the laboratory (e.g., by decreasing the vehicle navigation error rate by at least *10*% or improving lane-keeping performance by at least *20*%). This would allow you to conclude, after your results are analyzed, the extent to which they support the null hypothesis or the specific alternative hypothesis, and not just whether to support or reject the null hypothesis. It also gives you an option of deciding, formally, that the results are inconclusive and that perhaps more data need to be collected.

Replace null-hypothesis tests with parameter estimates, effect size estimates, or model comparisons. A more ambitious solution to the problems inherent in p-values and alphas is to forego NHST in favor of alternative analytic techniques. The *New Statistics* movement (Cumming, *2012*, *2014*) recommends that scientists abandon hypothesis tests and replace them with confidence interval and effect size estimates. Other reformers (e.g., Kruschke, *2010*) advocate use of Bayesian parameter estimates and *credible intervals* in place of hypothesis tests. Parameter and effect size estimates shift the focus of analysis from the question, "Are the means different?" to the question, "How different are the means?" A confidence or credible interval can tell the researcher whether an effect size is plausibly different from zero, but just as important, whether it is plausibly big enough to be of practical importance. Effect size and parameter estimates also allow easier accumulation of information for a meta-analysis (see below) than do dichotomous "significant/n.s." decisions.

2.5.4.3 PRESENTATION OF EXPERIMENTAL RESULTS

Show the data. It is worth highlighting the importance of presenting more, rather than less, raw data to research customers. By raw data, we do not mean the data points from individual participants (though those may sometimes be appropriate), but rather graphs, confidence intervals, effect size measures, and statistical test outputs other than those of the magical "$p < .05$" type). The added importance of this last bit of guidance to meta-analyses will be described below.

Choose language carefully. We should be very careful that the language we use does *not* convey the impression that effects that might be important for safety improvement but fail to reach the magic *.05* levels should be disregarded. Potential offenses here, ranked from bad to worse, might be to describe an effect of, say, $p = .07$, with the phrases "not significantly different," "not different", or "equivalent." Even if we report the p-values for such effects, readers who have the time and attention span only for our Abstract, Discussion, or "key points" may overlook them. More advisable phrasing would be to label such an effect as "approaching conventional levels of statistical significance" or as a "nonsignificant trend." Equally important when such effects are in evidence is to describe in the text (not in just tables and graphs) their raw magnitudes, in terms such as "a *4*-second savings in response time" or "a *30*% gain in accuracy." This allows the human factors readers and research customers to assess the practical importance of the effect, and not just its degree of statistical significance.

Accumulate evidence over experiments. Earlier, we referred to "prior probabilities" for assuming that an effect might actually exist in the world, before we have seen the data from our current experiment. And previous research is the best source of such prior beliefs. Literature reviews can qualitatively summarize that research, but the ideal tool for accumulating evidence over studies is the **meta-analysis** (Borenstein et al., *2009*; Rosenthal, *1991*; Cumming, *2014*). Meta-analytic approaches provide quantitative estimates of the "collective wisdom" of that prior research, which may enable us to not only know that an effect is likely to be there (or not), but also to provide a point estimate of how large it is likely to be; that is, an explicit alternative hypothesis as discussed above. Recognizing the importance of meta-analysis has two implications for us. First, in our own literature reviews, we can use the meta-analysis to estimate effect sizes. Second, in reports of our data, we can include the statistical details of our effects including both significant and, importantly, non-significant effects, with effect sizes given for both. This will help other researchers produce unbiased effect size estimates in the meta-analyses that they may wish to conduct.

2.6 COMPUTATIONAL MODELING

A *computational model* of human performance will compute, via computer simulation, some key performance outcome, such as the time or accuracy to perform a particular task (Byrne, 2014; Pew & Mavor, 1998). Such models are generally of two forms: analytic equations and discrete event simulations.

2.6.1 Analytic Equations

Analytic equation models are relatively simple to develop and understand, often involving linear algebra. The terms of such models usually come directly from regression weights derived from empirical data. A simple example is the serial *self-terminating visual search model* (Sternberg, 1966). This model will predict how long it will take to find a target amongst a cluttered set of non-targets, such as a finding a name in a non-alphabetized list. Each item must be examined in turn before determining whether or not it is the desired target; and once the target is found, the search is stopped. If several trials are conducted, with the target name sometimes present anywhere in the list and sometimes absent, and the time to find the target (when present) is recorded, then a regression analysis of these times against the total number of items in the list (N) will be well fit with a linear equation indicating: Search time (ST) = $a + NT/2$ when the target is present.

In this equation a is a constant representing the time to decide that the single target is present, and T is the average time required to examine each non-target item and establish that it is not a target before moving on to the next item. The division by 2 results because the target is randomly placed in the list, sometimes early, sometimes late, but, on the average, halfway through. We discuss this model in more detail in Chapter 4.

Similar equation models have been derived to predict the human detection of automation errors as a function of automation reliability (Wickens & Dixon, 2007) or to predict task performance as a function of sleep deprivation and circadian cycle (Wickens et al., 2015). Useful models need not capture only a monotonic linear relationship. In fact, one of the most enduring models of human performance, predicting the time required to move a cursor to a target known as Fitts' law (Fitts & Posner, 1967), is based on a logarithmic relation between the distance or amplitude of the movement, and the target width: MT= $a + \log[2A/W]$.

Stevens's law of psychophysics describes the relation between the subjective intensity of a stimulus and its physical intensity as SI = PI(*n*th) where the exponent, *n*, can be greater than 1 or less than 1, depending on the sensory magnitude. Simple models of memory postulate an exponential decay of the quality of that memory as a function of time of the form: memory strength = *T*(*n*th): $N < 1$.

2.6.2 Discrete Event Simulation Models

The **discrete event simulation (DES) model** runs in real time to simulate a process inferred to operate within the brain. An example is a visual scanning or eye movement model called salience effort expectancy value (SEEV; Wickens, 2015), to be discussed in Chapter 4. SEEV consists of four terms that determine at any moment in time the attractiveness of any area of interest (e.g., a display) in the visual workspace, an attractiveness that determines the relative likelihood of moving the eye there. The eye then moves to the next location, and the relative attractiveness of all areas is computed again and the process repeats.

The main advantage of the discrete event simulation model is that it can impose the variability on the process that is an inherent feature of human performance. Thus, the SEEV model predicts that the eye will not always rigidly follow the same scan pattern (Wickens, Sebok, Gacy, & Li, 2015). When the model user specifies a variability term for the model, this will then produce a distribution of outputs, with a standard deviation just like those of the actual human performance data. While the data of the DES model may be like that of human performance, the DES model may require less than a second of computer time to generate the full distribution. Human performance data, in contrast, may take hours to obtain with human-in-the-loop experiments.

One particular advantage of the distributions provided by DES models is that they allow determination of worst-case predictions. For example, if a model is developed to predict the response time of a driver to react to an unexpected stop of the vehicle in front and apply the brake and stop, the distribution may predict the minimum headway separation that will allow 95% of rear-end collisions to be avoided. Such information can be quite valuable as a basis for safety recommendations.

A further advantage of computational models of human performance is that they can be coupled with models of non-human system performance to predict performance of the aggregate human–system. In the above example, our braking response model will be coupled with a model of vehicle dynamics and road-surface conditions to predict the actual collision likelihood.

Like the meta-analysis, a model is only as good as the data that went into it, and models must be validated to be useful (Wickens & Sebok, 2014; Wickens, Vincow, Shopper, & Lincoln, 1997). A typical model validation will "run" the model through several different conditions to predict performance, and then "run" human participants through the same conditions, with a sufficiently high sample size to ensure reliable data. A relatively high product moment correlation between predictions and data (e.g., above 0.60 to 0.70) will then validate the model and ensure the user that, for future applications, model outputs can substitute for more expensive and time-consuming human-in-the-loop simulations.

2.7 CONCLUSION

In conclusion, a variety of research methods and factors are available to the engineering psychologist to ensure that the conclusions they draw in formulating principles of human performance and human interaction with systems are valid and useful. Such validity can grow as more different methods converge on the same answer. For example, an experiment provides results to validate a principle, which then, in application, reduces the accident rate. Although no single researcher is likely to produce all such data, ample familiarity with the published literature can easily substitute. We trust that the chapters that follow will provide access to much of that literature.

Key Terms

alternative hypothesis *20*
analysis of variance (ANOVA) *13*
Bayes factor *19*
Bayesian statistics *19*
computational model *11*
Confound *12*
correlation *14*
discrete event simulation (DES) model *22*
effect size *16*
experimental design *11*
factorial design *13*
fidelity *10*
meta-analysis *11*
null hypothesis *16*
null hypothesis significance testing (NHST) *16*
p-value *16*
regression analysis *14*
sample size *16*
statistical interaction *13*
statistical significance *14*
t-test *12*
variability of data *16*

Bibliography

Berger, J. (1985). *Statistical decision theory and Bayesian analysis* (2nd Ed.). New York: Springer-Verlag.
Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester: John Wiley & Sons, Ltd.
Botzer, A., Meyer, J., Bak, P., & Parmet, Y. (2010). User settings of cue thresholds for binary categorization decisions. *Journal of Experimental Psychology: Applied, 16,* 1–15.
Byrne, M. D. (2014). Human performance modeling. In D. Boehm-Davis, F. Durso, & J. Lee (Eds.), *APA handbook of human-system integration* (Chapter 22). Washington, DC: American Psychological Association.
Caird, J., Simmons, S. M., Wiley, K., Johnston, K. A., & Horrey, W. (2018, February). Does talking on a cell phone, with a passenger, or dialing affect driving performance? An updated systematic review and meta-analysis of experimental studies. *Human Factors*, *60*(1), 101–133.
Card, S. K., Newell, A., & Moran, T. P. (1986). The model human processor. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and performance* (Vol. II, Ch. 45). New York: Wiley.
Cumming, G. (2012). *The new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.
Cumming, G. (2014) The new statistics: Why and how. *Psychological Science*, *12*, 7–29.
Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*, 274–290.
Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology, 72*, 78–89.
Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin Review*, *25*, 207–218.
Drews, F. A., Johnston, W. A., & Strayer, D. L. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology, Applied, 9*, 23–32.
Erlick, D. E. (1964). Absolute judgments of discrete quantities randomly distributed over time. *Journal of Experimental Psychology, 67,* 475–482.

Fisher, D. L., Schweickert, R., & Drury, C. G. (2006). Mathematical models in engineering psychology: Optimizing performance. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (3rd Ed., pp. 997–1024). New York: Wiley.

Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.

Fitts, P. M., & Posner, M. A. (1967). *Human performance*. Pacific Palisades, CA: Brooks Cole.

Getty, D. J., Pickett, R. M., D'Orsi, C. J., & Swets, J. A. (1988). Enhanced interpretation of diagnostic images. *Investigative Radiology*, *23*, 240–252.

Gonzales, C., Fakhari, P., & Busemeyer, J. (2017). Dynamic decision making: Learning processes and new research directions. *Human Factors, 59*, 713–731.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley. (Reprinted 1988, Los Altos, CA: Peninsula).

Hayes, W. (1994). *Statistics*. Orlando, FL.: Harcourt Brace.

Helton, W. S. (Ed.). (2009). *Canine ergonomics: The science of working dogs*. Boca Raton, FL: CRC Press.

Hoffman, R. R. (2008). Human factors contributions to knowledge elicitation. *Human Factors*, *50*(3), 481–488.

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*(5), 640–648.

Ioannidis, John P. A. (2005). *Why most published research findings are false. PLOS Medicine*, *2*(8), e124. doi: 10.1371/journal.pmed.0020124. ISSN 1549-1277. PMC 1182327

Kaber, D. B., & Endsley, M. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*, 113–153.

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences, 14*, 293–300.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science, 6*, 299–312.

Lee, J., Wickens, C., Liu, Y., & Ng Boyle, L. (2018). *Designing for people: An introduction to human factors engineering* (3rd Ed.). Charleston, SC: CreaeSpace.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.

Loukopoulos, L., Dismukes, R. K., & Barshi, E. (2009). *The multi-tasking myth*. Burlington, VT: Ashgate.

Lusted, L. B. (1976). Clinical decision making. In D. Dombal & J. Grevy (eds.), *Decision making and medical care*. Amsterdam: North Holland.

Macmillan, N. A., & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107,* 401–413.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd Ed.). Mahwah, NJ: Erlbaum.

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50, 215–241*.

Metzger, U., & Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: An empirical study of active control versus passive monitoring. *Human Factors, 43,* 519–528.

Neyman, J., & Pearson, E. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. A, 231*, 289–337.

Onnasch, L., Wickens, C., Li, H., & Manzey, D. (2014, May). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, *56*(3), 476–488.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 943–951.

Parasuraman, R. (1985). Detection and identification of abnormalities in chest x-rays: Effects of reader skill, disease prevalence, and reporting standards. In R. E. Eberts & C. G. Eberts (eds.), *Trends in ergonomics/human factors II* (pp. 59–66). Amsterdam: North-Holland.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*, 531–536. doi:10.1177/1745691612463401