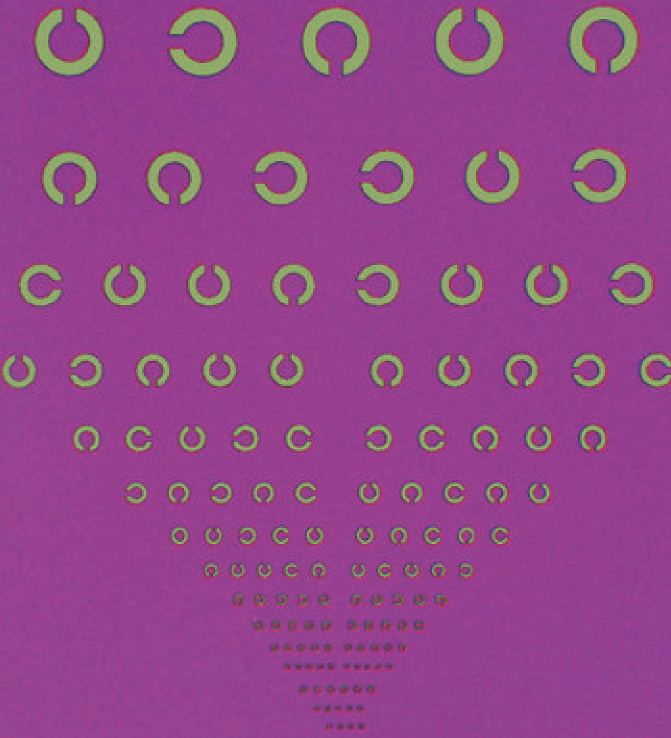
THE CAMBRIDGE HANDBOOK OF APPLIED PERCEPTION RESEARCH

Edited by Robert R. Hoffman, Peter A. Hancock, Mark W. Scerbo, Raja Parasuraman, James L. Szalma



The Cambridge Handbook of Applied Perception Research

The Cambridge Handbook of Applied Perception Research covers core areas of research in perception with an emphasis on its application to real-world environments. Topics include multisensory processing of information, time perception, sustained attention, and signal detection, as well as pedagogical issues surrounding the training of applied perception researchers. In addition to familiar topics, such as perceptual learning, the *Handbook* focuses on emerging areas of importance, such as human-robot coordination, haptic interfaces, and issues facing societies in the twenty-first century (e.g., terrorism and threat detection, medical errors, the broader implications of automation). The volume is organized into sections representing major areas of theoretical and practical importance for the application of perception psychology to human performance and the design and operation of human-technology interdependence. It also addresses the challenges to basic research, including the problem of quantifying information, defining cognitive resources, and theoretical advances in the nature of attention and perceptual processes.

ADVANCE PRAISE FOR *THE CAMBRIDGE HANDBOOK OF APPLIED PERCEPTION RESEARCH*

"A comprehensive review of current research by the top-tier authors in the field, with wide application to human-system integration."

– Thomas Sheridan, Professor Emeritus of Applied Psychology, Massachusetts Institute of Technology

"A handbook of applied perception research is a challenging undertaking. This one meets the challenge with 54 chapters authored by an impressive array of experts, spanning multiple sensory modalities, diverse methodologies, processes ranging from sensory to emotional, individual variation along with norms – and all of these from a basic and applied perspective. Among the welcome expansions on traditional topics like ergonomics and attention are chapters on the modalities of touch and olfaction, human-robot interaction, effects of video-game play, ecological approaches, and development across the life span. Each chapter offers a concise introduction that will send the interested reader further, and the 1100+ pages as a whole provide an exciting and comprehensive portrait of this rapidly evolving field."

– Roberta Klatzky, Professor of Psychology, Carnegie Mellon University

"*The Cambridge Handbook of Applied Perception Research* is not just for perception researchers – it is the comprehensive resource on perception that all practitioners and researchers who hope to make an applied contribution have been waiting for."

– Frank Durso, Professor of Psychology, Georgia Tech

"This stimulating collection dramatically illustrates the breadth of applied perception research: from the effects of video-game play on visual attention to the possibilities of olfactory interfaces. The book is also a testimony to the enduring impact of Joel Warm on the study of vigilance in particular and applied perception research in general."

– Jeremy Wolfe, Brigham and Women's Hospital, Harvard Medical School

"*The Cambridge Handbook of Applied Perception Research* weaves stories of the challenges faced by application-inspired researchers into the fabric of today's core disciplinary ideas. Contributors note that the story of applied perception research is the story of experimental psychology more broadly, and many of the chapters in this volume provide evidence that this claim has merit. Forward-looking chapters also show how questions posed in the context of emerging applications, such as human-robot coordination, virtual environments, and security management, might provide direction for both experimental psychology and cognitive science in the years to come."

– C. Melody Carswell, Professor of Psychology, Associate Director, Center for Visualization and Virtual Environments, University of Kentucky

The Cambridge Handbook of Applied Perception Research

Edited by

Robert R. Hoffman
Institute for Human and Machine Cognition, Pensacola, FL

Peter A. Hancock
University of Central Florida

Mark W. Scerbo
Old Dominion University

Raja Parasuraman
George Mason University

James L. Szalma
University of Central Florida

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107096400

First published 2015

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication data
The Cambridge handbook of applied perception research / [edited by] Robert R. Hoffman, Institute for Human and Machine Cognition, Pensacola, FL, Peter A. Hancock, University of Central Florida, Mark Scerbo, Old Dominion University, Raja Parasuraman, George Mason University, James L. Szalma, University of Central Florida.
pages cm. – (Cambridge handbooks in psychology)
Includes bibliographical references and index.
ISBN 978-1-107-09640-0 (hardback)
1. Perception – Research. I. Hoffman, Robert R.
BF311.C184 2014
153.7–dc23 2014006187

ISBN 978-1-107-07290-9 Volume 1 Hardback
ISBN 978-1-107-07293-0 Volume 2 Hardback
ISBN 978-1-107-09640-0 Two-Volume Hardback Set

Contents

Volume 2

Part V Perception and Design

Part VI Perception and Domains of Work and Professional Practice

Color plates follow pages 356 and 800.

Contributors

BRITTANY L. ANDERSON-MONTOYA, Old Dominion University

HEATHER R. BAILEY, Washington University in St. Louis

CARRYL L. BALDWIN, George Mason University

DAPHNE BAVELIER, University of Rochester

JAMESON D. BEACH, Institute for Human and Machine Cognition

JEFFREY S. BEDWELL, University of Central Florida

KEVIN B. BENNETT, Wright State University

RICHARD A. BLOCK, Montana State University

DEBORAH A. BOEHM-DAVIS, George Mason University

COREY J. BOHIL, University of Central Florida

DAVID B. BOLES, University of Alabama

AVINOAM BOROWSKY, Ben Gurion University of the Negev

JESSICA BRAMLETT, Georgia State University

ALLISON A. BRENNAN, University of British Columbia

J. CHRISTOPHER BRILL, Old Dominion University

MATTHEW S. CAIN, Duke University

MEREDITH CARROLL, Design Interactive, Inc.

ROBERTO CHAMPNEY, Design Interactive, Inc.

KAIT CLARK, Duke University

NANCY J. COOKE, Arizona State University Polytechnic

LORI M. CURTINDALE, East Carolina University

CLARE DAVIES, The Open University, U.K.

PATRICIA R. DELUCIA, Texas Tech University

ANDREW E. DEPTULA, University of Central Florida

MICHAEL B. DILLARD, University of Alabama

COLIN D. DRURY, State University of New York–Buffalo

CHRISTOPHER EDMAN, Wright State University

JAMES T. ENNS, University of British Columbia

SARA IRINA FABRIKANT, University of Zürich

VICTOR S. FINOMORE, Air Force Research Laboratory

ARTHUR D. FISK, Georgia Institute of Technology

JOHN M. FLACH, Wright State University

MATTHEW E. FUNKE, Naval Medical Research Unit

ANDRE GARCIA, George Mason University

ADAM GAZZALEY, University of California San Francisco

DOUGLAS J. GILLAN, North Carolina State University

REBECCA A. GRIER, Institute for Defense Analyses

SIMEN HAGEN, University of Victoria

KELLY HALE, Design Interactive, Inc.

DIANE F. HALPERN, Minerva Schools at Keck Graduate Institute

PETER A. HANCOCK, University of Central Florida

DEBORAH L. HARM, NASA Johnson Space Center

MARY HEGARTY, University of California

LAURIE M. HELLER, Carnegie Mellon University

NICOLE D. HELTON, University of Canterbury

WILLIAM S. HELTON, University of Canterbury

ROBERT R. HOFFMAN, Institute for Human and Machine Cognition

JERRED HOLT, Wright State University

XIAOGANG HU, Rehabilitation Institute of Chicago

RICHARD J. JAGACINSKI, The Ohio State University

KEITH S. JONES, Texas Tech University

ASTRID M. L. KAPPERS, Vrije Universiteit Amsterdam

SIMON KEMP, University of Canterbury

ROBERT C. KENNEDY, RSK Assessments, Inc.

ROBERT S. KENNEDY, RSK Assessments, Inc.

ALAN KINGSTONE, University of British Columbia

IOANA KOGLBAUER, Independent Human Factors Consultant

NORMAN E. LANE, RSK Assessments, Inc.

ROBERT D. LATZMAN, Georgia State University

CYNTHIA LAURIE-ROSE, Otterbein University

PATRICIA LEE, Wright State University

RICHARD LOWE, Curtin University

VALERIE LUGO, Design Interactive, Inc.

POORNIMA MADHAVAN, Old Dominion University

LEONARD S. MARK, Miami University

GERALD MATTHEWS, University of Central Florida

JYOTI MISHRA, University of California, San Francisco

STEPHEN R. MITROFF, Duke University

TRACY L. MITZNER, Georgia Institute of Technology

ALEXANDER M. MORISON, The Ohio State University

TAYLOR MURPHY, The Ohio State University

TAKAMICHI NAKAMOTO, Tokyo Institute of Technology

JOHN G. NEUHOFF, The College of Wooster

KARL M. NEWELL, Pennsylvania State University

TAL ORON-GILAD, Ben Gurion University of the Negev

RAJA PARASURAMAN, George Mason University

TIFFANY A. PEMPEK, Hollins College

ROBERT W. PROCTOR, Purdue University

KATIE A. RAGSDALE, University of Central Florida

ANIL K. RAJ, Institute for Human and Machine Cognition

MILLARD F. RESCHKE, NASA Johnson Space Center

EVAN F. RISKO, University of Waterloo

MATTHEW RIZZO, University of Iowa

WENDY A. ROGERS, Georgia Institute of Technology

JESSE Q. SARGENT, Frances Marion University

MARK W. SCERBO, Old Dominion University

NATASHA B. SCHWARTZ, Georgia State University

F. JACOB SEAGULL, University of Michigan Medical School

CORY-ANN SMARR, Georgia Institute of Technology

L. JAMES SMART, Miami University

KAY STANNEY, Design Interactive, Inc.

JAMES STASZEWSKI, Carnegie Mellon University

CLAYTON L. STEPHENSON, University of Southern California

MARY E. STUART, Institute for Human and Machine Cognition

BREANNA E. STUDENKA, Utah State University

JOEL SUSS, Michigan Technological University

LEEDJIA SVEC, United States Navy

JAMES L. SZALMA, University of Central Florida

JAMES TANAKA, University of Victoria

JAMES THOMPSON, George Mason University

WOUTER M. BERGMANN TIEST, Vrije Universiteit Amsterdam

LAUREN A. VASSILIADES, Institute for Human and Machine Cognition

MICHAEL A. VIDULICH, Air Force Research Laboratory

PAUL WARD, University of Greenwich

JOEL S. WARM, Air Force Research Laboratory

DAVID A. WASHBURN, Georgia State University

CHRISTOPHER D. WICKENS, Alion Science and Technology

SCOTT J. WOOD, NASA Johnson Space Center

DAVID D. WOODS, The Ohio State University

MOTONORI YAMAGUCHI, Edge Hill University

LIN YE, Innovation Design

JEFFREY M. ZACKS, Washington University in St. Louis

About the Editors

ROBERT R. HOFFMAN, PH.D., specializes in cognitive systems engineering and human-centered computing. He is a Fellow of the Association for Psychological Science, a Fellow of the Human Factors and Ergonomics Society, and a Fulbright Scholar. Hoffman has been recognized internationally for his research and work on human factors in remote sensing, on the psychology of expertise and the methodology of cognitive task analysis, and on human-centered computing issues and intelligent systems technology, as well as the design of macrocognitive work systems. Hoffman is a coeditor of the Department on Human-Centered Computing in *IEEE: Intelligent Systems* and editor of the book series Expertise: Research and Applications. He was a cofounder of two journals: *Metaphor and Symbol* and the *Journal of Cognitive Engineering and Decision Making*. His current research focuses on the psychology of intelligence analysis, methodological issues in the analysis of complex systems, and performance measurement for macrocognitive work systems. His most recent books are *Collected Essays on Human-Centered Computing* (2012) and *Accelerated Expertise: Training for High Proficiency in a Complex World* (2014). A full vita and all of his publications are available at www.ihmc.us/users/rhoffman/main.

PETER A. HANCOCK, D.SC., PH.D., is Provost Distinguished Research Professor, Pegasus Professor, and trustee chair in the department of psychology and the Institute for Simulation and Training, as well as in the department of civil and environmental engineering and the department of industrial engineering and management systems at the University of Central Florida. He also directs the MIT^2 (Minds in Technology, Machines in Thought) Research Laboratories. He is the author of more than seven hundred refereed scientific articles and publications and author or editor of fifteen books, including *Human Performance and Ergonomics* (1999); *Stress, Workload, and Fatigue* (2000); and *Performance under Stress* (2008). He has been continuously funded by extramural sources for every one of the thirty years of his professional career. He is a Fellow of numerous scientific societies and was the president of the Human Factors and Ergonomics Society in 2000. His work revolves primarily around the reciprocal influence between human beings and technology. He also works on the theoretical and empirical exploration of time and is an award-winning historian. Details of his work can be found at www.peterhancock.ucf.edu.

MARK W. SCERBO, PH.D., is professor of human factors psychology at Old Dominion University. He began his career as a research assistant at AT&T Bell Laboratories in 1980, received his Ph.D. in experimental psychology from the University of Cincinnati in 1987, and then returned to AT&T, where he managed the Systems Evaluation Center in New Jersey from 1987 to 1990, introducing usability engineering to the Network Operations Division. He is a Fellow of the Human Factors and Ergonomics Society, and he received his Modeling and Simulation Professional Certification in 2002. He has published more than 160 scientific publications and currently serves as an associate editor for the journals *Simulation in Healthcare* and *Human Factors*. He has more than thirty years of experience researching and designing systems and displays that improve user performance in academic, military, and industrial work environments. His current research interests are focused on user interaction with medical simulation technology. In addition, he has studied human factors issues related to the behavioral and physiological factors that affect human interaction with virtual environments, automated systems, and adaptive interfaces.

RAJA PARASURAMAN, PH.D., is University Professor of Psychology at George Mason University, Virginia. He is also director of the Graduate Program in Human Factors and Applied Cognition and of the Center of Excellence in Neuroergonomics, Technology, and Cognition. His primary research is in human factors and cognitive neuroscience, as well as in molecular genetics of cognition and in neuroergonomics, which he defines as the study of brain and behavior at work. His books include *The Attentive Brain* (1998), *Neuroergonomics: The Brain at Work* (2007), and *Nurturing the Older Brain and Mind* (2012). Parasuraman is a Fellow of numerous organizations, including the American Psychological Association (1991) and the International Ergonomics Association (2006), and a national associate of the National Academy of Sciences. He is the recipient of numerous awards and honors, including the Jerome H. Ely Award for best paper in the journal *Human Factors* (1997, 2001), the Franklin V. Taylor Award for Lifetime Achievement in Applied Experimental and Engineering Psychology from the American Psychological Association (2004), the inaugural Raymond S. Nickerson Award for Best Paper in the *Journal of Experimental Psychology: Applied* (2010), and the Admiral Kollmorgen Spirit of Innovation Award for Contributions to Neuroergonomics (2010). His recent awards include the Triennial Outstanding Educators Award from the International Ergonomics Association and the Celebration of Scholarship Award from the College of Humanities and Social Sciences at George Mason University (both in 2012).

JAMES L. SZALMA, PH.D., is an associate professor in the psychology department at the University of Central Florida. He received a B.S. in chemistry from the University of Michigan and an M.A. and Ph.D. in applied experimental/human factors psychology from the University of Cincinnati. His lab, the Performance Research Laboratory, studies how variations in task characteristics interact with the characteristics of the person (i.e., cognitive abilities, personality, emotion, motivation) to influence performance, workload, and stress of cognitively demanding tasks. His primary research interests include signal/threat detection (e.g., friend/foe identification), training for threat detection, and the influence of the characteristics of tasks and operators on performance at tasks that require sustained attention or that include human-automation interaction. He is currently conducting research on the application of video game–based tasks to train sustained attention, and on the utility of fuzzy signal detection theory for performance evaluation in threat detection.

Dedication

The editors and contributing authors dedicate this volume to Emeritus Professor Dr. Joel Warm.

Professor Warm joined the faculty of the University of Cincinnati shortly after receiving his doctorate from the University of Alabama in 1966 and completing postdoctoral research at Fort Knox, Kentucky. Currently he is Senior Scientist at the Warfighter Interface Division, Human Effectiveness Directorate, Air Force Research Laboratory, Wright-Patterson AFB, Ohio.

He is a Fellow of the American Association for the Advancement of Science, the American Psychological Association, the Association for Psychological Science, and the Human Factors and Ergonomics Society. He has served on two National Research Council committees and is a member of the editorial boards of *Human Factors* and *Theoretical Issues in Ergonomic Science*. Professor Warm was honored by the UC Graduate School for his outstanding contribution to the training of graduate students. He received the Paul Fitts Award for Outstanding Contributions to the Education and Training of Human Factors Specialists by the Human Factors and Ergonomics Society.

Professor Warm has coauthored or edited four books and dozens of book chapters. He has presented well over two hundred papers at professional meetings. He has authored or coauthored more than ninety articles in refereed journals, including the most prestigious experimental psychology journals: *Psychological Bulletin,* the *Journal of Experimental Psychology: General,* the *Journal of Experimental Psychology: Human Perception and*

Performance, Contemporary Psychology, the *British Journal of Psychology,* the *American Journal of Psychology, Psychonomic Science, Perception & Psychophysics,* the *Bulletin of the Psychonomic Society, Psychophysiology, Motivation and Emotion,* the *Journal of General Psychology, Human Factors,* and the *International Journal of Aviation Psychology.*

Professor Warm taught dozens of people how to conduct research. He always held that a four-hour exam during which students lose several quarts of blood is a pop quiz. He chaired thirty-eight dissertations and fifty master's theses and served as a committee member on numerous others. Many of his students have gone on to serve with distinction as teachers and researchers at universities; many have taken influential positions in government and the private sector, conducting applications of experimental psychology.

Professor Warm has been a major force in perception research for nearly four decades. He has accomplished far more than cumulative science: His contribution to the field has laid the theoretical and methodological foundation for expanding the horizons of research and its applications. He continues to break new ground and open new avenues for better understanding human perception by establishing linkages among attention, human performance, and the cognitive and physiological processes that underlie them. His contributions to the field of perception are literally multimodal, including the visual, auditory, tactile, olfactory, and time perception modes, and relations among modalities. His recent work has entered the realm of psychophysiology, pioneering the study of the relations between performance and physiological response of cerebral blood flow. This ongoing programmatic work will shape further developments in the cognitive resource theory of attention and vigilance by articulating more precisely how energetic resources are manifested.

His major substantive contributions include:

- Establishing a link between performance in sustained attention and the workload associated with such tasks.
- Contributing to the theoretical understanding of sustained attention through both his empirical research and his synergistic efforts.
- Publishing important research on basic perceptual processes associated with cross-modality perception and perceptual illusions.
- Conducting important research on the practical application of tactile perception to detection.
- Performing seminal research on the influence of olfactory stimulation on sustained attention.

Professor Joel Warm: Our Appreciations

From David B. Boles, Ph.D., Professor of Psychology, University of Alabama:

> In 2004 Jeff Phillips and I had just finished delivering two papers at the Human Factors and Ergonomics Society meeting, when a gentleman approached and asked if I'd have a cup of coffee with him. Not quite

knowing what to make of it but flattered by the attention, I agreed. That is how, in my fifties, I met Joel Warm, the mentor I never had. Joel introduced himself and said he'd never thought it possible that mental resources could be measured, as we did, through a user questionnaire. He wanted to hear more. Apparently he liked what he heard, because the Multiple Resources Questionnaire has become a staple in the vigilance research of his ever-widening group. Somehow, after that first conference meeting, Jeff and I found ourselves included in the group photo of his people. In 2007, at Joel's urging, we published papers on the validity of the questionnaire in predicting dual-task interference. In 2008, I happily found myself an attendee at his Festschrift at the University of Cincinnati. He subsequently guided one of my recent students, Michael Dillard, through two Repperger Internships at the Air Force Research Laboratory at Wright-Patterson Air Force Base. In 2012, Michael held a National Research Council Research Associateship Award there. Joel is that type of mentor. Incredibly generous with his time and attention, he gathers productive people around him and gently pushes them toward excellence. To be sure, I had caring, generous mentors in graduate school who set me firmly on my career path, but Joel is of a different caliber altogether. I am delighted to be a contributor to this volume in his honor.

From Ken Boff, Former Chief Scientist of the Human Effectiveness Directorate, Air Force Research Laboratory:

I believe that my earliest interactions with Joel Warm date to the auspicious year of 1984. That, of course was the title of Orwell's famous novel, the year that the Apple Macintosh computer was introduced at the Super Bowl, and the year that we conspired through the offices of the University of Dayton Research Institute to put on a Human Perception and Performance Workshop and Short Course in Dayton, Ohio. I recall that Joel gave a particularly outstanding lecture on monitoring, vigilance, and search and the practical implications of the research for the design of displays and controls. I believe it was most highly rated and widely praised by the students taking the course, who for the most part were practicing engineers and human factors professionals. Over the years since, Joel and I have maintained an active collaboration among his lab and students and the various research organizations I've led. I consider myself particularly fortunate as the principal beneficiary of this relationship. Joel's students who came to work on projects at the Air Force Lab were very well prepared to engage in high-level research and were exemplars of Joel's enthusiastic and careful approach to laboratory science. Some of them stayed after graduation and are on the way to becoming key leaders in the lab. Joel's many contributions as a teacher, mentor, and scientist are a cornerstone of the human factors profession.

From Traci Galinsky, National Institute for Occupational Safety and Health:

I think that over the years many friends and family members of Joel's students often wondered why his students worked so incredibly hard for him. They'd see the huge workload and the long hours and the working on weekends and sometimes even on holidays. They'd hear the grumbling from the students about how long they had to wait for Dr. Warm while he was meeting with other students – and the never-ending data analyses and rewrites – the unrelenting pursuit of flawlessness every step of the way. They'd see all this and wonder why these students put up with it. For me the answer was simple. Dr. Warm never asked us to work harder than he himself worked. He never asked us to care more about the work and our accomplishments than he himself cared. Making someone like him – who is so smart and so accomplished – proud of me made all of the hard work so very, very worth it.

From Diane F. Halpern, Claremont McKenna College:

I owe a great personal and professional debt to Joel, my academic father. I recall many "Joel stories." For example, there was the time when he went out in a wicked snowstorm to retrieve something from his office for a class demonstration in perception. He returned looking like a snowman. One of my first conference presentations was at Psychonomic Society, where only members (and never graduate students) were permitted to present their research. Joel had agreed to present our paper, and then feigned a sore throat so that I did it. I remember how worried I was about starting a doctoral program with two small children and how supportive he was, assuring me that I could do it. Unlike some faculty who talked about being supportive of women students, Joel actually made sure that all of the students who worked with him would succeed. Although he often appeared gruff, those of us who were lucky enough to know him well saw a caring professor who loved his family, his students, and experimental psychology, in that order. I did not realize how well he prepared me for academic life until I graduated and began my first academic job. Several newly minted Ph.D.s started with me on my first job – all were from prestigious universities. I was surprised to find that I was as well prepared as they were (maybe even better) for the rigors of academic life – designing research, asking good questions, and turning my work into publications. As I worked on my first research projects, I heard Joel's voice in the background, reminding me how to control for extraneous variables and how to argue effectively when challenged. Joel is a modest man, so I expect that all of this praise is making him uncomfortable, so I will end with a simple thank you. Please know how much you have affected the lives of so many people.

From Peter Hancock, University of Central Florida:

Each individual's future turns on the smallest of things. My world turned at the 1982 Psychonomic Society Meeting where, as a midlevel doctoral student, I first met Joel Warm. Our meeting was happenstance. I was standing alongside my advisor Karl Newell and the late Dennis Holding, when an elegant but harassed-looking individual hove into sight. Who else but Joel! He was editing a text in Holding's book series. "I've lost one of my contributors and have no one to do environmental stressors!" It was at that moment that the tumblers of the universe rotated and Holding, suffering from what Churchill called "acute terminological inexactitude," replied, "Well here's a young chap, Hancock, knows all about environmental stressors." How Holding's pants failed to burst into flames I have yet to discover! Joel eyed me up and down and I felt very much that I knew what a drowning man clutching at a straw looks like.

"Well if you say so, but we have to have it soon." The deal was sealed before I'd even opened my mouth and the rest, as they say – From that day on, Joel Warm has been part of my life and has changed that life. He has acted as exemplar, mentor, and friend, for which I am, and remain, most sincerely grateful to both the universe and Dennis Holding. It has been an honor and a privilege to have worked and collaborated with Joel, and to have learned from him. As you can see from this volume, in this I am very, very far from being alone.

From Robert Hoffman, Institute for Human and Machine Cognition:

I was fortunate to have learned mnemonic memory techniques from one of my University of Cincinnati professors, Joe Senter, prior to taking Joel Warm's experimental psychology course. I will confess to having done pretty well in the laboratory part of the course. But like all undergrad psychology wannabes, I dreaded The Course That Decided One's Fate. But I used the mnemonics to memorize all the junkier bits that we had to learn for that course (i.e., schedules of reinforcement, "verbal behavior," and other oddities in the history of the science of mind). In his fashion, Joel announced midterm grades in class, and by name. When I was called first I was shocked. "First" might have meant 65 percent. Fortunately, it did not mean a score nearly that low, and the remainder of my undergraduate career was little more than a halo effect. At the time when I advanced to the graduate program, human factors interested me not. I went the psycholinguistics route with another of my beloved UC professors, Dick Honeck. But Joel was on my committees. His tough questions were punctuated by his smiles at department TGIFs and at ball games; Joel does not hold back smiles. His graduate course in perception was astounding for its coverage and the vigor and depth of the discussions. I went through something like twenty-five yellow note pads. Still have 'em. Who would have known that knobs and dials would morph into cognitive systems engineering? But this is where I find myself. Able to see as far as I can because I stand on the shoulders of my mentors. And on this occasion, compelled by my heart to look at those shoulders upon which I stand, with affection and gratitude.

From Mary Ann McCarty, Dallas:

I was one of many, many students in Professor Warm's research methods course, the much-feared Psychology 381. He stopped me on my way out of one of his lectures and stated he was concerned. I was going through a particularly rough time, and it finally came out that I did not even have the textbook for the course, I was so strapped. Making small talk and convivial conversation along the way to keep me calm, we walked to the bookstore, where he quietly had a word with some staff and a manager, bought a copy of the textbook, and gave it to me. I still have that book. I still tear up remembering him doing this for me, a lowly undergrad in the back of the room whose name he didn't even know. I'm crying right now. He helped me. I am grateful for his absorbing lectures, his eleven horsemen, his vocabulary choice, but most of all, for this kind thing he did for me, his affirmation of me as a student and as a human being. I went on to graduate from the University of Cincinnati, and without him, that might not have happened.

From Raja Parasuraman, George Mason University:

I had the privilege and honor first to meet Joel Warm in 1976 at an international conference in Italy. I was a raw graduate student in the United Kingdom and he was already a leading figure in research on sustained attention and perception. He immediately made a strong impression on me by his encouragement of my early research efforts in these areas. A few years later when I went to do postdoctoral work in the United States, Joel invited me to stay at his home in Cincinnati. His kindness and his supportive encouragement of my career aspirations deeply touched me. I had been turned down by no fewer than twenty universities in seeking an academic position, but Joel urged me to persevere, and I was finally successful in 1982, with the help of his recommendation letter. The rest, as they say, is history. The many fruitful research collaborations with Joel over the past three decades have brought me great pleasure. But those early exchanges with Joel at his home and at conferences are what are seared in my memory and bring me much joy. His is the definition of true friendship.

From Robert Proctor, Purdue University:

I was not a student of Joel Warm's, but I feel a very close affinity to him as a researcher. Joel takes a systematic, empirical approach to basic and applied research that is among the best that experimental psychology has to offer. Joel's contributions to the understanding of vigilance alone are sufficient to ensure a lasting contribution to applied perception. His work in that area provides an excellent case study of how one can go about investigating a topic of both basic and applied importance in a systematic manner. My first interactions with Joel were with regard to the chapter "Things That Go Together: A Review of Stimulus-Response Compatibility and Related Effects" that he and Earl Alluisi wrote for the book *Stimulus-Response Compatibility: An Integrated Perspective* (1990), which I edited with Gil Reeve. This chapter, which provides a thorough and insightful review of research on stimulus-response compatibility through 1990, set the tone we were wanting for the volume, and thus we placed it as the first chapter. Since that time, I have come to know Joel well, and he has always had kind words and encouragement for me about the various projects in which I was engaged. The word that comes to mind when I think of Joel is "gracious." He is a very gracious person, of highest integrity, whom it has been my good fortune to know.

From David Washburn, Georgia State University:

In 1993 I presented the results of a study on the mechanisms of the Stroop effect at the meeting of the Southern Society for Philosophy and Psychology. After the presentation, Professor Joel Warm spoke with me at length about the data and their implications. Years later, contemplating the topic of this volume and its origins in a Festschrift for Professor Warm brought freshly to mind that conversation and the encouragement it provided to a young behavioral scientist. The anecdote illustrates an important lesson about the wide-ranging influence that an established and respected scholar can have, even beyond his own advisees, even in such short informal interactions.

Preface

Perception research and theory have always had clear links to applications. For instance, a student of Wilhelm Wundt who studied binaural sound location went from his dissertation to develop a mechanism for the German Air Force to determine the range of aircraft, flying at a great distance, only on the basis of the sound of their engines received at two large horn-type receivers. Another student of Wundt went from his dissertation on motion perception to develop one of the very first simulators, to study the job of railway motormen. As is well known, research on vigilance was stimulated by the need to understand and redesign the tasks of radar operators during World War II. With regard to sustained attention in particular, the scope of applications has expanded significantly in recent years, from radar (the one person–one machine context) to other contexts involving teams and multiple technologies, such as weather forecasters looking at multiple data types to predict hurricane tracks, soldiers searching for land mines and improvised explosive devices (IEDs), and baggage screeners. Another salient area is performance measurement, which has recently been driven to move beyond the measurement of individual performance (e.g., accuracy, errors, workload) to the measurement of performance of "complex cognitive systems." We can anticipate that research will continue to be driven by national needs, including private sector as well as government and military needs, and that important theoretical advances will stem from research that has both ecological and epistemological utility. This will continue the trend of dissolving of the simplistic "applied versus basic" distinction while retaining a promise to contribute to the basic science of psychology.

Acknowledgments

While we editors focused on the goal of creating a volume appropriate for designation as a Cambridge University Press handbook, *The Cambridge Handbook of Applied Perception Research* honors Emeritus Professor Joel Warm (see the dedication). Related to that, the foundation for this volume was laid at a Festschrift conference held in Warm's honor in May 2008. It was hosted by the University of Cincinnati and supported by the Air Force Research Laboratory and the American Psychological Association.

The editors would like to acknowledge the participation of Kenneth Boff (USAF), David Boles (University of Alabama), Melody Carswell (University of Kentucky), Marvin Dainoff (Miami University), Cynthia Dember (University of Cincinnati), Frank Durso (Georgia Tech), Colin Drury, (SUNY-Buffalo), Robert Fox (Vanderbilt University), Traci Galinsky (NIOSH), Diane Halpern (Minerva Schools at Keck Graduate Institute), Joseph Lappin (Vanderbilt University), Len Mark (Miami University), William Marras (Ohio State University), Gerald Matthews (University of Cincinnati), W. Todd Nelson (USAF), Robert Proctor (Purdue University), Roger Rosa (NIOSH), Judi See (SAIC), and Lloyd Tripp (USAF).

The editors would like to thank a number of people at Cambridge University Press, from contracting to production, then editorial and marketing, all of whom were supportive and helpful: Adina Berk, Simina Calin, Cathryn Chen, Jeanie Lee, Eve Mayer, Emily Spangler and Bhavani Ganesh.

For assistance in preparation of the indexes, the editors would like to thank Andrew Raines, Jubal Porter, Jeff Yerkes, and Jameson Beach.

PART I

Background and Methodology

1 Applied Perception Research: An Introduction to This Handbook

Robert R. Hoffman, Peter A. Hancock, Mark W. Scerbo, James L. Szalma, and Raja Parasuraman

Perception has figured prominently throughout the history of experimental psychology. Following Herbart, Brentano, Helmholtz, Fechner, and Wundt – all of whom considered and researched perception phenomena – their students, the first generation of psychologists, carried the new science of psychology to universities around the world. As the psychology of perception matured, great integrators helped the field to consolidate vast tracts of experimental findings and to identify fruitful avenues for research. This is a fundamental and enduring pursuit.

This *Handbook* covers core topics in research on perception, with an emphasis on application to "real-world" environments, primarily the workplace. Topic areas range from multisensory processing of information, to time perception, to sustained attention, to signal detection. There are also chapters on the pedagogical issues surrounding the training of a new generation of applied perception researchers.

We have cajoled contributions from scientists who are working on many emerging areas of importance, such as human-robot coordination, olfactory and tactile displays, and virtual training systems. Chapters identify and discuss current and future trends in the application of perception psychology to issues facing societies in the twenty-first century (e.g., terrorism and threat detection, medical errors, pedagogical implications of technology).

Rather than recapitulating each chapter, we briefly explain our overall organization of the *Handbook*. There are sections on the major areas of applied research, areas of theoretical and practical importance for the application of perception psychology to human performance and to the design and operation of human-technology work systems. These also address the challenges for basic research, including the problem of quantifying information, defining cognitive resources, and theoretical advances in the nature of attention and perceptual processes.

Part I, Background and Methodology, covers the history of applied perception research, basic methods such as signal detection, and other measurement techniques. Part II, Attention and Perceptual Processes, presents work on relations of perception to processes including memory, attention, and motor skills, linking these to their physiological foundations. Part III, Perception and Modality, includes chapters on cross-modal and multimodal perception, and haptics – reminding us that, of course, there is far more to perception than visual perception alone. Part IV, Perception in Context, has chapters that explore the richness of perception: perception of biological motion, perception of events, and perception of the self. Part V, Perception and Design, discusses applications of perception research for the design of artifacts. These are primarily computer and display systems for the workplace, but chapters such as that by Takamichi Nakamoto on olfactory interfaces illustrate the breadth of application of applied research on interface design. Part VI, Perception and Domains of Work and Professional Practice, includes chapters that examine applications to work domains, focusing on perceptual learning, mental workload, and sustained attention: chapters that illustrate how perception research contributes to such fields as medicine, astronautics, and robotics, to name a few. Part VII, Individual and Population Differences, discusses the development of perception and perception in special populations. The final part, Professional Topics and Issues, discusses issues of training and workforce evaluation.

We can look at the *Handbook* as a whole in terms of some themes or threads that weave through the chapters. A theme of the chapters is methods and methodology, which makes the volume true to its designation as a *Handbook*. Readers (we hope) can read a chapter and then begin their efforts at application. Therefore, contributors to this volume were required to discuss "how we do research" and to pull in brief descriptions of representative experiments or particularly important experiments or classic references. We wanted citations that would be up-to-date but also key or classic references for readers that would provide more specific guidance on "how we do it."

A second major theme is, naturally enough, that of the relationship between applied and basic science. It is generally recognized that the relations between these are complex and that each depends on the other (Hoffman and Deffenbacher, 2011; Stokes, 1997; see Helton, Kemp, and Hoffman, this volume). Each chapter in this *Handbook* presents work that relies on basic science methods and theories, or that contributes to basic science as well as applications. Therefore, the work has both epistemological relevance and ecological validity.

A third theme that weaves through chapters is the concept of perceptual learning. A number of chapters rely upon the notion that we learn how to perceive meaningful things and events in our environments. This ability is crucial for proficient performance in essentially all domains of professional expertise.

A fourth theme is the ecological approach to the study of perception. A number of chapters discuss this approach, illustrating how basic research can have ecological validity and ecological relevance.

While this *Handbook* reflects the breadth and depth of applied perception research, the coverage is by no means exhaustive.

Contributors to this volume look to a next generation of researchers, confident that they will advance perception research and applications to new accomplishments and contributions to society and humanity.

References

Hoffman, R. R. and Deffenbacher, K. A. (2011). A multidimensional analysis of the relations of basic and applied psychology. *Theoretical Issues in Ergonomic Science*. DOI: 10.1080/1464536X.2011.573013.

Stokes, D. E. (1997). *Pasteur's quadrant, basic science and technological innovation*. Washington, DC: Brookings Institution.

2 Some Highlights in the History of Applied Perception Research

William S. Helton, Simon Kemp, and Robert R. Hoffman

Introduction

The history of applied perception research is largely the history of perception research. As we (and others) have argued, the basic-applied distinction hides more than it reveals (Helton and Kemp, 2011; Hoffman and Deffenbacher, 2011; Wilson et al., 2013). All scientific research has applications, but much research in experimental psychology is driven by practical concerns in the first place. The applied orientation of modern experimental psychology was present from the beginning, as in the pragmatism of William James (1982) and Charles Sanders Peirce. While Peirce's diverse contributions to mathematics, philosophy, and logic have been widely noted (see Houser and Kloesel, 1992, 1998), Peirce was employed for a time at the U.S. Coastal Survey, where his psychophysical research was directly applicable to measurement in astronomical observation (Stafford, 1897). Alexander Graham Bell was inspired by the acoustic research of Helmholtz. He experimented in acoustics, ultimately leading to his patents for acoustic telegraphy and telephony. More than this, Bell funded American scholars to work in the laboratory of Hermann von Helmholtz. Another founder of scientific psychology, Wilhelm Wundt, trained a generation of experimental psychologists, many of whom conducted applied perception research (Rieber and Robinson, 2001; see also Hoffman and Deffenbacher, 1992).

Our goals in the present chapter are (1) to demonstrate the tight interweaving between basic perception science and the applications of the research and the related goal, (2) to demonstrate that the history of applied perception research *is* the history of perception research. We present some vignettes from different historical eras – less-turned pages from the history of perception research.

Aesthetics, Architecture, and Military Science

Ancient Greek and Roman perceptual theory was focused on a number of controversies. For example, there was a controversy over whether the visual process was essentially one of intromission or extramission. In intromission theories, such as were held by Aristotle and Democritus (Lindberg, 1976), information traveled from the outside world to the eye. Extramission theories, on the other hand, as were held by Plato, Galen, and many Stoic philosophers, asserted that some quality traveled from the eye to the objects perceived in the world. There were also different theories of odor. Plato's account held that small particles were emitted by some objects and these then traveled to the nose; Aristotle's was that only information rather than some form of matter was transmitted (Kemp, 1997). Although to us it might seem strange that anyone could seriously contemplate an extramission theory of vision or a nonparticle theory of odor, ancient reasoning about these issues was sophisticated and empirically based (e.g., Kemp, 1990, 1997; Lindberg, 1976).

There was also considerable interest in applied thinking about perception. For example, Aristotle's student Theophrastus (ca. 320 BC/1949) wrote a short work on odor, which, while making no reference to how smell is transmitted, includes a good deal on the principles of making perfume. A persistent theme in Vitruvius's (ca. 30 BC/1999) *Ten books on architecture* is how to design buildings that look good and have good acoustics. In Books 3 and 4 he laid down rules for designing temples that exhibit certain pleasing ratios and related these ratios to those of the human body. So, for example, in different styles of temple it was important to have the correct ratios of column width to column height and of column width to the distance between columns. Good Corinthian columns were said to "imitate the slenderness of a young girl" (Vitruvius, ca. 30 BC/1999, p. 55).

Vitruvius also took perceptual illusions into account. A number of perceptual illusions – for example, the moon illusion, the waterfall illusion, the apparent bending of straight sticks when half submerged in water – had been identified by early thinkers, and some theories had been put forward (e.g., Johannsen, 1971). Vitruvius did not present theories, but did give practical advice. For example, he advised that the corner columns of temples should be made thicker by a fiftieth of a diameter "because they are cut into by air on all sides and therefore seem more slender to the viewer. Thus where the eye deceives us, reasoning must recompense" (Vitruvius, 30 BC/1999, p. 50).

Vitruvius commented on wall paintings. He was distinctly unimpressed by the element of fantasy common in the frescoes painted in his own time but commented favorably on the use of trompe l'oeil in a previous period "in which they also imitated the shapes of buildings, and the projection into space of columns and pediments" (Vitruvius, 30 BC/1999, p. 91). It is clear from frescoes that have survived from the time of the early Roman Empire (see, e.g., Mason, 2011 for an online collection of Pompeii frescoes) that some fresco painters had an understanding of visual perspective and were adept at employing it. It is also worth remarking that the phenomena of visual perspective were often explained by extramission theories of human vision; Euclid's (fl. 300 BC) account of geometrical optics was an early example (Lindberg, 1976, ch. 1).

Applied perception was also an important element in ancient military thinking. Toward the end of his *Epitome of military science*, Vegetius (ca. AD 400/1996, p. 139) discussed the building of siege towers, constructions that were designed to be transported to the base of the wall of a besieged city. It was important that siege towers overtop the walls, and hence the height of the wall needed to be known. Vegetius suggested either shooting an arrow with a light calibrated thread attached into the top of the wall and then measuring the length of thread from top to base, or measuring the shadow cast by the wall at a particular time and, at the same time, putting a rod of known length into the ground and measuring its shadow. Both of these techniques acknowledged the difficulty of measuring the size of objects at a distance.

Vitruvius (ca. 30 BC/1999, p. 23, 130) also indicated a military application of acoustics. It was important to tune the sinew cords on machines such as ballistae and catapults so that they were of equal tension, and he pointed out that this was generally done by equating the pitch heard when they were plucked.

We now turn a page from more recent times.

Echolocation

Extramission theories of vision fell out of favor centuries ago (Lindberg, 1976), but it is worth noting that while such accounts may not be regarded as scientifically valid today, they are quite plausible in certain respects. For instance, Gibson's ecological optics defines affordances as properties of objects taken with reference to an observer. One might see a brick but perceive a door stop. In this phenomenology, something does travel from the eye to the object, although this is perhaps to be understood in a metaphorical sense (Hoffman et al., 1990).

Perhaps the best evidence of the plausibility of extramissive theories of perception is that bats, dolphins, and blind humans *do* emit energy, and then use the returning echoes to locate and identify objects in the world. However, the energy they use is in the form of ultrasound (infrasound in the case of elephants) rather than light.

Although the fact that bats can navigate well in the dark had long been common knowledge, Lazarro Spallanzani (1729–1799) seems to have been the first person to investigate the phenomenon. His basic technique was systematically to deprive bats of each of their five senses and see what difference it made. Unfortunately, the majority of his bats managed to navigate after any of the interventions, but Louis Jurine (1751–1819), who seems to have used a rather more effective earplug than Spallanzani, was able to establish that bat navigation was carried via audition (Galambos, 1942). Nonetheless, this result was not accepted until much later, in part because

Jurine's researches fell well short of demonstrating how bats might use hearing to navigate.

After the *Titanic* sank in 1912, there was a good deal of scientific interest in devising a warning system to detect semisubmerged hazards. As sound was already well known to travel through water and to produce echoes in that medium, some sort of acoustical device was an obvious candidate. According to Galambos (1942), Hiram Maxim (better known as the inventor of a machine gun) was convinced by the experiments of Spallanzani and Jurine that bats navigated by echo, and he suggested that ships carry low-frequency sound emitters and record the echoes. In 1914 Reginald Fessenden demonstrated that a prototype emitting 540 Hz sound could detect an iceberg off the Newfoundland coast. Shortly afterward, the German submarine offensive against Allied merchant shipping in World War 1 led to intensification of efforts to produce practical sonar devices, and by the end of the war British, French, and U.S. scientists had developed a sonar system capable of detecting submerged submarines (d'Amico and Pittenger, 2009).

By the end of the nineteenth century it was well known that electromagnetic radiation was also reflected from objects and therefore could also in principle be used for echolocation. The first practical application seems to have been devised as early as 1904 by Christian Hülsmeyer (Brown, 1999, ch. 2), who patented a variety of radar as a means for detecting (surface) ships in fog. The use of pulsed signals – useful for finding the range of objects – was an early development, and by World War II a number of countries had devised systems that could detect aircraft. During World War II, both radar and sonar detection systems were enormously enhanced.

Finding ways to improve the performance of echolocation systems suggested further psychological as well as physical research. One very general issue concerned the question of what the displays of these, or any other, systems should look like (e.g., Hoffman and Deffenbacher, 1992). A particular issue that emerged from the research on radar detection was that of understanding how a radar operator might distinguish a signal, an incoming hostile aircraft, for example, from noise such as might result from atmospheric conditions, noisy equipment, or flocks of birds. Consideration of this issue led to the development of the theory of signal detection (e.g., Green and Swets, 1966; Marcum, 1947).

Interestingly, although the original applications of the theory of signal detection were mostly in the field of perception, there were many subsequent applications in other areas, for example, in the prediction of criminal or delinquent activity (Fergusson, Fifield, and Slater, 1977).

The development of echolocation systems may have provided the clue as to how bat navigation actually worked. Hartridge (1920) described a number of his observations of bats flying in darkened rooms and noted that they were able to avoid a series of threads strung across the rooms. After his descriptions of their behavior he suggested that

> bats during flight emit a short wave-length note and that this sound is reflected from objects in the vicinity. The reflected sound gives the bat information concerning its surroundings. . . . Experiments on "sound ranging" apparatus during the war have shown that the sense of direction in man can be made use of for estimating the position in space of objects emitting sound waves. Under ordinary circumstances it is necessary to increase greatly the effective distance between the ears in order to obtain the required accuracy. But if sound waves of sufficiently short wave-length could have been used the same results could have been obtained with the normal distance between the ears. (Hartridge, 1920, pp. 56–57)

In fact, bats emit most of their acoustic energy at frequencies beyond the normal range of human hearing, and it was not until 1937, with the aid of apparatus that could convert ultrasound into audible sound, that it was shown that bats emitted ultrasonic cries. Using this apparatus, Galambos and Griffin (1942) were able to provide a satisfactory demonstration that bats did in fact use the navigation system that Hartridge had hypothesized. Further research provided much more information about the different calls that bats emit and how these vary both with the species of bat (and the ecological niche it inhabits) and with the different environmental challenges that an individual bat confronts (Jones and Holderied, 2007).

Studying bat echolocation was thus crucial to advancing underwater sonar detection. However, the field also proved important in developing navigational aids for blind people. As far back as the eighteenth century it was known that at least some blind people could detect distant objects and sometimes give reasonable estimates of how far away they were (Supa et al., 1944). As with bats, the basis for this ability was unclear, especially as blind people themselves were often uncertain how they did it. Sometimes, as in the title of the crucial paper reported in 1944 by Michael Supa, Milton Cotzin, and Karl Dallenbach, the phenomenon was described as "facial vision." However, the series of experiments that these researchers conducted established that the ability was in fact based on perceiving reflected sound. The experiments used normally sighted but blindfolded people (who proved capable of learning the task) as well as blind participants and, similarly to Spallanzani, proceeded by eliminating different cues and seeing what difference the elimination made to the ability. Using earplugs and sound screens eliminated the ability. (The final reference in Supa et al.'s 1944 paper is to Griffin and Galambos's 1941 work but no other work on bat echolocation is referenced. Nor is there any reference to sonar or radar detection. Thus, it is not clear from the paper how much use, if any, their work made of this previous research.)

As bats are clearly adapted for acoustic echolocation in a way that humans are not, there is a possibility to devise aids for blind people that use some of the methods employed by bats, and there have been a number of attempts at this. In the 1950s Leslie Kay, who had trained as an electrical engineer, worked for the UK Royal Navy on developing underwater sonar technology, and later studied bat echolocation (e.g., Kay, 1961). After taking up a university position, he devised the first of a series of devices sometimes known as sonic spectacles. Basically the early prototypes worked by emitting an ultrasonic frequency sweep signal. The echoes from the sweep were subtracted from the frequency of the signal actually being emitted at that time at receivers located at each ear and the differences fed to the ears. Thus, the pitch of the difference encodes the range of the reflecting object (a higher pitch means the object is farther away), and the relative timing in the ears codes the direction. Later developments refined the basic technology, resulting in a number of devices that are currently used by blind people, Kay was interested in developing devices that could be used from infancy by children who were born blind (see, e.g., Kay, 2000; Kay, Strelow, and Kay, 1977).

Research findings from research on bats, on underwater sonar, and on aids for the blind has been important for establishing optimal ways to perform a variety of different tasks. For example, constant frequency emission devices, such as that developed by Fessenden, are not a very effective means of determining object range, because only the initial instant at which the echo is heard carries information that can be used to apprehend range, and this may well be masked by the emitted signal itself. For this purpose, frequency modulated signals (often a sweep from relatively high to relatively low frequency) are superior, because time taken to return can be calculated for each returning frequency.

Individual bats have been shown to produce an array of different calls that are adapted to different circumstances. For example, bats flying in relatively open spaces often use relatively long calls emitted infrequently; when in cluttered environments or closing in on their prey they use more frequent calls (Jones and Holderied, 2007). Incidentally, insectivorous bats produce calls that are above the frequency range of human hearing because lower frequency sound implies wavelengths that are longer than the wing length of most moths and hence do not reflect well from them (e.g., Jones and Holderied, 2007).

Spallanzani and Jurine could not have known that their work on bat echolocation would result in signal detection theory, sonar-radar, and sonic spectacles, but we suspect their work was not merely inspired by undirected curiosity. Sometimes the applications in perception research take a while to materialize.

Sometimes, applications cross-connect people whom you might think of as being unconnected.

Color Vision and Countering Forgery

Benjamin Franklin the scientist was keenly interested in light, and color vision in particular (1765/1970). Pioneer vision researcher Thomas Young credited Franklin with helping to revive the wave theory of light.

Franklin studied his spectral afterimages using colored spectacles he made. He determined the duration of the afterimages (see Cohn, 2010; Lemay, 2009) and observed that colors fade faster than forms. Franklin corresponded with other investigators about his perceptual research, including Erasmus Darwin (grandfather of Charles Darwin). Richard Darwin (Charles Darwin's father) did his doctoral research in medicine on visual afterimages.

A description of some of Richard Darwin's research (1786, p. 314):

> Place a piece of red silk, about an inch in diameter . . . on a sheet of white paper, in a strong light; look steadily upon it from about the distance of half a yard for a minute; then closing your eyelids cover them with your hands, and a green spectrum will be seen in your eyes, resembling in form the piece of red silk.

Charles Darwin even wrote of his father (1879, p. 84), "He [Richard Darwin] published, in Vol. LXXVI of the *Philosophical Transactions* [of the Royal Society] a paper on Ocular Spectra, which [Charles] Wheatstone told me was a remarkable production for the period; but I believe he was largely aided in writing it by his father [Erasmus]." Indeed, Richard Darwin's work in vision was remarkable for the time, and he discovered a number of aspects of vision, including of microsaccades. This, and the work pioneered by Benjamin Franklin, had major implications for visual science.

Afterimages, negative color afterimages in particular, are a fascinating visual phenomenon. We now understand that negative afterimages are caused as the cone cells adapt to overstimulation and lose sensitivity. The eye's normal correction for this physiological response is to microsaccade. However, if the color source is of sufficient areal extent such that the microsaccades do not change the color to which the individual cones are subjected, the cones will eventually stop responding. In Richard Darwin's experiment, the observer looked at a brightly lit red silk piece. The cone cells reactive to the red wavelengths would be strongly activated. These red cones would eventually exhaust. When the observer looks away, the cones more reactive to the green wavelength would react more strongly and the observer would see a green afterimage.

This work was continued by Thomas Young (1801), who proposed that there were three color receptors, and this idea came to be known as the Young-Helmholtz-Maxwell theory of color perception. Thomas Young's research would also inspire Charles Wheatstone's work on optics (1879), including Wheatstone's work on depth perception (which will be discussed in the next section).

Back to Franklin. His interest in color perception may have been the result of experiments with using colored ink as a plausible counterfeiting countermeasure. In his era, his profession was referred to as Natural Philosophy, but Franklin felt that scientific experiments would result in practical inventions. As a printer, Franklin was keenly aware of the technologies for printing (he invented some of his own) and was an early advocate for paper currency. Franklin's colony of Pennsylvania issued paper currency in 1723. In 1729, Franklin (then 23 years old) wrote a pamphlet, "A modest inquiry into the nature and necessity of a paper currency." His arguments were convincing enough that Pennsylvania reissued paper currency and Franklin quickly moved into the printing of paper currency for other colonies by 1731. By the time Franklin retired from commercial printing in 1764, he had printed up to 2.5 million notes for Pennsylvania, Delaware, and New Jersey (and, in the process, had become rich).

The most pressing problems facing Franklin, and other commercial printers of paper money, were alteration and counterfeiting. Alteration involved raising the value of a paper note by modifying the denomination of the note. One countermeasure Franklin utilized was to modify the spelling of Pennsylvania for each denomination. Franklin also developed techniques for defeating counterfeiting. One technique, for example, was nature printing (using natural items, such as leaves, in the print). Recognizing that natural tree leaves are all unique, Franklin determined they would be extremely hard to duplicate if used on paper currency. Engravers would have an immense challenge to replicate the natural print. The trick was to use nature printing on a large commercial scale. For this, Franklin invented a method related to stereotyping (Newman, 1971).

We now turn a page from astronomy.

Astronomy

While the "incident at Greenwich" receives honorable mention in nearly all of the books on the history of psychology, less well known are some of the interesting details.

A main task in practical astronomy for the great seafaring nations was to record the positions of the moon and some dozens of bright "fixed" stars, for the purpose of producing the data (siderial tables) that would be used by maritime navigators. Longitude calculation could be based on compass heading, whereas the determination of latitude involved comparing the position of a fixed star to the position of that same star as observed at the latitude of an observatory, such as the Greenwich Observatory outside London, or the Prussian observatory in Königsberg. Positions on the celestial sphere had to be calculated on the basis of measurement of azimuth angle and elevation as seen by an observer located at some position on the Earth. These trigonometric calculations were based on transit times – times at which stars were at their zenith, or highest point in the sky.

Using this "right method," Astronomer Royal James Bradley cataloged the celestial coordinates for 389 fixed stars, daily longitude and latitude of the moon, and transit times for the sun and planets – the times at which they crossed the Greenwich meridian (ca. 1818; see Forbes, 1975). Transit times were determined in the following procedure.

The telescope was positioned such that it pointed due north. A thread or wire (called a horary wire) was mounted inside the objective lens and was viewed through the eyepiece. Because of the telescope's north-south orientation, the horary was aligned with the meridian (a great circle passing overhead and through the Earth's rotation axis, in other words, a line of longitude projected onto the celestial sphere). The observer (who had to be lying on his back) tracked the target star by looking at it through the telescope and noted the time at the precise moment that the star crossed the horary wire, that is, reached the highest point in the sky (its zenith, hence the name "zenith sector"). The astronomer would have to look back and forth from clock to star, keeping mental track of the ticks of the pendulum (the length of which was precisely calculated so that the clicks made by the catchment mechanism would correspond to the passage of seconds).

All sources of error were of great significance to navigation. A tiny difference at Greenwich could mean miles off course to a ship far at sea. In addition, the sources of error would have to be identified, measured, and controlled if one were to resolve fundamental issues, such as whether there are motions to the so-called fixed stars, to test various predictions of Newton, to determine whether the speed of light were finite, and so on. There were a number of sources of error in the observations. The position of an instrument mounted in a loft would shift as the floor beams warped; there would be positional error due to the refraction of the Earth's atmosphere; and so on. Stars were rarely so cooperative as to cross the horary wire just at the beat of a clock. The need for the observer to keep track of two things at once led to error in the recorded transit times.

Bradley's modification to the method involved placing two more horary wires in the eyepiece, one to either side of the center one, and a side-mounted lantern to illuminate the horary wires clearly. Both modifications were aimed at supporting the observer in his task. As a star approached the first of the three horary wires within the field of view, the observer looked at the clock and noted the time. While keeping mental count of the beats of the clock, the observer would note the position of the star at the beat just before the star crossed the central horary wire, and the position of the star just after it crossed the central wire. The judged proportional difference in distance, expressed as tenths, was used to extrapolate the time in tenths of a second, when the body would have precisely crossed the central horary wire. Finally, the tenths would be added to the seconds counted before the star crossed the first horary wire line.

This method depended on the observer's spatial judgment, memory for spatial position, and ability to keep mental track of the passage of time and mentally calculate sums of fractions. The method continued in use up to the time of the fifth Astronomer Royal (1765–1811), Nevil Maskelyne. In 1794 and 1795, there worked at the Greenwich Observatory an assistant observer named Kinnebrook, whose job was to record the positions of a set of fixed stars for the purpose of producing the sidereal tables that would be used by navigators in the Royal Navy. To determine positions, Kinnebrook used the Bradley method. One evening in 1795 Maskelyne himself recorded some transit times and found differences as great as 0.5 to 0.8 second compared to Kinnebrook's records from the previous night's observations. This meant that the observatory's published sidereal tables could have been chock full of errors. After failing to yield the "correct" values (Maskelyne himself as the standard, of course), Kinnebrook was dismissed.

Decades later, a report of this incident at Greenwich caught the attention of Frederich W. Bessel, a young genius, highly respected for his research on "practical astronomy" (he eventually achieved success as the first person to estimate the distances of stars). He had been made director of the new University of Königsberg Observatory by appointment in 1810 by Wilhelm III of Prussia. After learning of the incident at Greenwich, Bessel commenced a project in which he compared his transit times with those of five other observers (at Königsberg and elsewhere). Finding reliable differences, he was able to create a "personal equation" whereby transit times of observers could be corrected in comparison to the reaction times of Bessel himself as the standard. Other observatories conducted such studies, involving what must have been dozens of astronomers over the period 1819–1860.

The perception research by Bessel was undertaken at a time when there was a push for devices to make precise measurements of time, called "chronoscopes," and recording devices called "chronographs" (see Sanford, 1889). These relied on the new technology of the electromagnet and improved the precision of observation (i.e., reducing the magnitude of personal equations an order of magnitude to a probable error of as little as plus or minus 0.05 second). Ormsby Mitchel, director of the Cincinnati College (now University of Cincinnati) Observatory, referred to the personal equation as the "personality of the eye" (1858). In 1851 he demonstrated the level of precision in astronomical recordings that could be made by taking account of the personal equation using the "American Method" of chronography, employing the galvanic barrel chronograph. As Mitchel (1860, p. 176) wrote, "The observer himself is but an imperfect and variable machine, utterly incapable of marking the exact moments required."

The technological advances in astronomy were aimed at ultimately avoiding the need to include the personal equation correction factor in determining the transit times. However, the technology made possible the first psychological research on the speed of mental processes. Bessel's work served as the model for the study of reaction time. In his very first lectures promoting scientific psychology, Wilhelm Wundt discussed Bessel's work and Johann Friedrich Herbart's notion of "complication" when two stimuli occur at the same time (1824). Wundt devised a "complication clock" for experiments allowing simultaneous judgments of the position of a moving pointer and the sound of a bell. Wundt's laboratory used the new chronoscope technology in studies of human reaction time (see Boring, 1929; Sanford, 1889; Wundt, 1903; see also Donders, 1863).

Additional applications and advances in technology interested early perception researchers.

Artillery

Hermann von Helmholtz was the grandson of an artillery officer and he himself served in the military as a physician. In the 1840s, the rifled barrel was introduced to artillery and enabled a dramatic increase in accuracy. This increase in firing accuracy, however, could only be realized if the gunner estimated the distance to the target with equal accuracy. This problem of range estimation was exacerbated in naval contexts, where less information is available to estimate distance to enemy ships. Helmholtz developed the first optical range finder, enabling more accurate range estimation by increasing optical disparity. Indeed, many of the later developers of optical range finding equipment, including Nobel laureate Albert Michelson, studied with Helmholtz.

Michelson provides an intriguing case, as when he studied with Helmholtz he was both a commissioned U.S. naval officer and funded by Alexander Graham Bell. Michelson would later work on optical range finders for the U.S. Navy. Helmholtz and his students were not alone in being interested in improving range estimation at the time. Charles Wheatstone, mentioned earlier, also explored optical issues regarding distance estimation (Wheatstone, 1838, 1852). Wheatstone, incidentally, was a musical instrument maker by trade who inspired some of the work of Alexander Graham Bell and worked for the British military to improve artillery gunners' rates of fire.

The applications of perception research have sometimes been social as well as technological. Indeed, some of the social applications are a part of the history of applied perception research we wish we could disclaim. The story does serve, however, as a cautionary tale of applied perception research.

The Uses of Early Applied Perception Research in Social Issues

In the 1850s when the early precursors of modern experimental psychology were being developed to deal with the personal equation in astronomy, the social world was very different from today. In the United States of America, for example, large numbers of people (African Americans) were slaves, Native Americans were being forcibly removed from their homelands, and women were denied equality with men.

In 1813 James Prichard published *Researches into the physical history of man*. Prichard advocated the belief that all human ethnicities had a common origin with Adam and Eve. Many abolitionists in the United States and elsewhere also advocated for the moral equality of humankind. Prichard's views were, however, quickly countered by advocates of polygenism, the view that there were multiple and independent origins of human ethnicities. In the United States, Samuel Morton, the founder of American physical anthropology, inspected mummies from ancient Egypt and determined Europeans and Africans had been distinct races three thousand years ago. As the Biblical flood occurred only one thousand years before this, Morton claimed Noah's sons could not have represented every racial variant on Earth. Morton was also an early advocate for craniometric measurements to validate a hypothetical link between cranial capacity and intelligence.

The rise of the evolutionary theory of Charles Darwin and Alfred Wallace in scientific circles reduced, but did not totally eliminate, advocacy of scientific polygenism. While slavery was banned in the developed world, some social elites sought to reinforce the status quo of European male empowerment. Researchers desiring to reinforce the social status quo scientifically began to apply the techniques developed to study the personal equation. This is exemplified in the work of Francis Galton. While some may wish to defend Galton as merely reflecting the common beliefs of his time, this is perhaps too generous, because Galton seems to have gone out of his way to defend racist and sexist beliefs while often presenting his position as neutral and objective. Galton (1873) wrote:

> On the other hand, the opinion of the present day repudiates the belief that the negro is an extremely inferior being, because there are notorious instances of negroes possessing high intelligence and culture, some of whom acquire large fortunes in commerce, and others become considerable men in other walks of life. . . . The truth appears to be that individuals of the mental caliber I have just described are much more exceptional in the negro than in the Anglo-Saxon race, and that average negroes possess too little intellect, self-reliance, and self-control to make it possible for them to sustain the burden of any respectable form of civilization without a large measure of external guidance and support.

In this article he submitted to the *London Times*, he recommended moving Chinese to Africa to replace all Africans.

In regard to sex differences, Galton conducted early research on sensitivity differences between the sexes. In one study he examined the difference in the two-point threshold on the nape of the neck (Galton, 1894). In Galton's system, perceptual sensitivity was often considered a marker of intelligence. For example, men should have quicker perceptual reaction times than women, because he considered them to be more intelligent. They should also be more perceptually sensitive. In his 1894 study, he found women to be more sensitive to touch than men. He did, however, note women's greater variability. He stated:

> I think that the recorded variability may in a very small part be accounted for by the fact that women vary much more than men in the exercise of sustained attention.... Some women are religiously painstaking, as much so as any men; but the frivolity of numerous girls, and their incapacity of, or unwillingness to give, serious attention, is certainly more marked than among men of similar ages. (1894, p. 42)

We have no reason to believe Galton was dishonest in his research, but he was by no means value-neutral. Even when the evidence contradicted his expectations and theories regarding male superiority, he managed to turn the tables on women. Given his theory that men are more perceptually sensitive, when they turned out not to be, he was able to turn women's increased variability into an indicator of their unreliability. There is little doubt that much of Galton's research agenda was to prove "scientifically" that men, especially white upper-class British men, should rule the world.

Conclusion

History of psychology is rich with many stories. Some of them have been retold so often they have become mythical. The Maskelyne-Kinnebrook incident is one such story. Many stories are rarely told, many are only partially told, and many need retelling for a next generation. Some stories are uplifting, some are surprising, and some are shameful. But all of them can, and should, be told.

The history of perception research is in large part the history of applied perception research, spanning the decades between the time that a student of Wilhelm Wundt made one of the first simulators to study the performance of railway motormen, to the time when a student of Joel Warm used new brain scanning methods to determine the mechanisms of attention. "Throughout its history there has been an interplay between practical concerns and basic laboratory science. Thus, many of the techniques that are used today in applied fields such as human factors engineering can be traced to early research in the psychology of perception. Conversely, many of the techniques that are used in basic experimental psychology can be traced to practical problems" (Hoffman and Deffenbacher, 1992, p. 32). This rich interplay suggests a need for an entirely new way of thinking about things rather than the simplistic applied-basic distinction. Research can be evaluated in terms of ecological validity, practical utility, and epistemological utility, among other dimensions (Hoffman and Deffenbacher, 1993, 2011). It is only by depending upon a richer set of distinctions that we may come to understand and appreciate the significance of sciences such as the science of perception.

References

Boring, E. G. (1929). *A history of experimental psychology*. New York: Century.

Boring, E. G. (1953). A history of introspection. *Psychological Bulletin*, 50, 169–189.

Boring, E. G. (1961). Beginning and growth of measurement in psychology. In H. Woolf (ed.) *Quantification: a history of the meaning of measurement in the natural and social sciences* (pp. 108–127). New York: History of Science Society.

Brown, L. (1999). *A radar history of World War II: Technical and military imperatives*. Bristol, UK: Institute of Physics.

Cohn, E. R. (Ed.) (2010). *The papers of Benjamin Franlkin*. New Haven, CT: Yale University Library.

D'Amico, A. D. and Pittenger, R. (2009). A brief history of active sonar. *Aquatic Mammals*, 35, 426–434.

Darwin, C. (1879): Preliminary notice. In Krause E, ed., *Erasmus Darwin* (pp. 1–127). London; Murray.

Darwin, R. W. (1786). New experiments on the ocular spectra of light and colours. *Philosophical Transactions of the Royal Society*, 76, 313–348.

Donders, F. C. (1863/1969). On the speed of mental processes (W. G. Koster, trans.). *Acta Psychologia*, 30, 412–431.

Fergusson, D. M., Fifield, J. K. and Slater, S. W. (1977). Signal detection theory and the evaluation of prediction tables. *Journal of Research in Crime and Delinquency*, 14, 237–246.

Fischer, D. H. (1970). *Historians' fallacies: toward a logic of historical thought*. New York: Harper Torchbooks.

Forbes, E. G. (1975). *Greenwich Observatory*, Vol. 1: *Origins and early history (1675–1835)*. London: Taylor & Francis.

Franklin, B. (1765/1970): Letter to Lord Kames. In A. H. Smyth (Ed.), *The writings of Benjamin Franklin,* Vol. IX: *1783–1788* (p. 25). New York, Haskell House.

Galambos, R. (1942). The avoidance of obstacles by flying bats: Spallanzani's ideas (1794) and later theories. *Isis*, 34, 132–140.

Galambos, R. and Griffin, D. R. (1942). Obstacle avoidance by flying bats: the cries of bats. *Journal of Experimental Zoology*, 89, 475–490.

Galton, F. (1873, June 5). Letter to the editor of *The Times*.

Galton, F. (1894). The relative sensitivity of men and women at the nape of the neck by Weber's test. *Nature*, 50, 40–42.

Green, D. M. and Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hartridge, H. (1920). The avoidance of objects by bats in their flight. *Journal of Physiology*, 54, 54–57.

Helton, W. S. and Kemp, S. (2011). What basic-applied issue? *Theoretical Issues in Ergonomic Science,* 12, 397–407.

Herbart, J. F. (1824–25). Psychologie als Wissenschaft. [Psychology as science]. In K. Kehrbach and O. Flügel (Eds.), *Johann Friedrich Herbart's sämtliche Werke in chronologischer Reihfolge*, Part 1, Vol. 5, (pp. 177–434). Langensalza, Germany: Hermann Beyer und Söhne.

Hoffman, R. R., Cochran, E. L. and Nead, J. M. (1990). Cognitive metaphors in the history of experimental psychology. In D. Leary (Ed.), *Metaphors in the history of psychology* (pp. 173–209). Cambridge: Cambridge University Press.

Hoffman, R. R. and Deffenbacher, K. (1992). A brief history of applied cognitive psychology. *Applied Cognitive Psychology*, 6, 1–48.

Hoffman, R. R. and Deffenbacher, K. A. (1993). An analysis of the relations of basic and applied science. *Ecological Psychology*, 5, 315–352.

Hoffman, R. R. and Deffenbacher, K. A. (2011). A multidimensional analysis of the relations of basic and applied psychology. *Theoretical Issues in Ergonomic Science*. DOI: 10.1080/1464536X.2011.573013.

Houser, N. and Kloesel, C. (Eds.) (1992, 1998). *The essential Peirce: Selected philosophical writings* (2 vols.). Bloomington and Indianapolis: Indiana University Press.

James, W. (1892). A plea for psychology as a natural science. *Philosophical Review*, 1, 146–153.

Johannsen, D. E. (1971). Early history of perceptual illusions. *Journal of the History of the Behavioral Sciences*, 7, 127–140.

Jones, G. and Holderied, M. W. (2007). Bat echolocation calls: Adaptation and convergent evolution. *Proceedings of the Royal Society B*, 274, 905–912.

Kay, L. (1961). Perception of distance in animal echo-location. *Nature*, 190, 361–362.

Kay, L. (2000). Auditory perception of objects by blind persons using bioacoustic high resolution air sonar. *Journal of the Acoustical Society of America*, 107, 3266–3275.

Kay, L., Stelow, E. R. and Kay, N. (1977). The use of electronic spatial sensors in the training of blind children. *Visual Impairment and Blindness*, 71, 174–175.

Kemp, S. (1990). *Medieval psychology*. New York: Greenwood.

Kemp, S. (1997). A medieval controversy about odor. *Journal of the History of the Behavioral Sciences*, 33, 211–219.

Lemay, J. A. L. (2009). *The life of Benjamin Franklin*, Vol. 3: *Soldier, scientist and politician 1748–1757*. Philadelphia: University of Pennsylvania Press.

Lindberg, D. C. (1976). *Theories of vision from Al-Kindi to Kepler*. London: University of Chicago.

Marcum, J. L. (1947). *A statistical theory of target detection by pulsed radar*. Santa Monica, CA: Rand Corporation RM-754.

Mason, M. K. (2011). *Roman and Etruscan wall paintings of the Tarquinia tombs and Pompeii*. Downloaded from http://www.moyak.com/papers/roman-art-etruscan.html, 7 January 2011.

Mitchel, O. (1858). On personal equation. *Monthly Notices of the Royal Astronomical Society*, 18, 261–264.

Mitchel, O. (1860). *Popular astronomy*. London: Routledge.

Morton, S. (1844). *Crania Aegyptiaca; or, observations on Egyptian ethnography, derived from anatomy, history, and the monuments*. Philadelphia: J. Penington.

Newman, E. P. (1957). Counterfeit continental currency goes to war. *Numismatist*, 70, 5–16.

Newman, E. P. (1971). Franklin making money more plentiful. *Proceedings of the American Philosophical Society*, 115, 341–349.

Rieber, R. W. and Robinson, D. K., 2001. *Wilhelm Wundt in history: The making of scientific psychology*. New York: Kluwer.

Sanford, E. C. (1889). Personal equation. *American Journal of Psychology*, 2, 3–38.

Schaffer, S. (1988). Astronomers mark time: discipline and the personal equation. *Science in Context*, 2, 115–145.

Stafford, T. (1897). On the various forms of personal equation in meridian transits. *Monthly Notices of the Royal Astronomical Society*, 57, 504–514.

Supa, M., Cotzin, M. and Dallenbach, K. M. (1944). "Facial vision": The perception of obstacles by the blind. *American Journal of Psychology*, 57, 133–183.

Theophrastus. (1949). Concerning odours. In Theophrastus, *Enquiry into plants and minor works on odours and weather signs*. Vol. 2, trans. A. Hort. London: Heinemann. Originally written between 345 and 285 BC.

Vegetius. (1996). *Epitome of military science*, trans. N. P. Milner (2nd ed.). Liverpool: Liverpool University Press. Originally written between 383 and 450 AD.

Vetruvius. (1999). *Ten books on architecture*, trans. I. D. Rowland. Cambridge: Cambridge University Press. Originally written between 30 and 20 BC.

Wheatstone, C. (1838). On some remarkable and hitherto unobserved phenomena of binocular vision. *London Royal Society Philosophical Transactions*, 128, 371–394.

Wheatstone, C. (1852). On some remarkable and hitherto unobserved phenomena of binocular vision. Part 2. *Philosophical Magazine*, 4, 504–523.

Wheatstone, C. (1879). *The scientific papers of Sir Charles Wheatstone*. London: Taylor and Francis.

Wilson, K.M., Helton, W.S. and Wiggins, M.W. (2013). Cognitive engineering. Wiley Interdisciplinary Reviews: *Cognitive Science*, 4, 17–31.

Wundt, W. (1903). *Grundzuge der physiologischen psychologie* (3 vols.) Leipzig: Verlag Wilhelm Engelmann.

Young, T. (1801). On the mechanics of the eye. Philosophical Transactions of the Royal Society of London, 91, 23.

3 Psychophysical Methods and Signal Detection: Recent Advances in Theory

Corey J. Bohil, James L. Szalma, and Peter A. Hancock

Introduction: A Brief History of SDT

Signal detection theory (SDT) represents one of the most prominent scientific developments in psychology of the past 60 years (Dember, 1998; Estes, 2002). Its application to perception began with the use of statistical decision theory for radar detection problems (e.g., Peterson, Birdsall, and Fox, 1954), and efforts to determine the sensitivity of information transmission via a sensitivity measure that was free of response bias (for an early discussion of the historical antecedents of SDT see Swets, 1973). A key insight by the pioneering researchers was that errors of commission in perception tasks are not necessarily the result of guessing, as assumed by threshold theories (Tanner and Swets, 1954).

The techniques provided by SDT have found wide application, including domains such as radiology, assessment of memory in clinical populations, and many kinds of monitoring tasks. In general, any categorical decision or diagnostic task can be evaluated using SDT, permitting separate assessment of the capacity of the decision maker to discriminate among categories (defined as perceptual sensitivity, d') and his or her cognitive bias for selecting one category over another (response criterion or response bias, β).

As a statistical model, SDT rests on a set of assumptions. These include the premises that (1) events to be detected (signals) are always embedded in a background of irrelevant sensory information (noise); (2) the distributions of noise and signal-plus-noise are of normal form and equal variance; (3) observers are both sensors and decision makers, and they adopt a criterion of sensory magnitude for deciding whether a given event is a signal or a non-signal; and (4) measures of perceptual sensitivity (e.g., d') can be treated as if they were independent of measures of response bias (e.g., β).

Even in some of the seminal work on SDT, Tanner and Swets (1954) recognized that empirical evidence indicated that the equal variance assumption is not always true, but that it is also not necessary for application of the SDT model (although it is necessary for the use of d' as a measure of sensitivity; see Swets, 1988). Further, the assumption of independence of sensitivity and response bias requires that the observing entity has no previously formed memory of signal and nonsignal events, which cannot be true of most human observers. Nevertheless, this also has not restricted the application of SDT to problems of applied perception.

Because much has already been written about SDT over its long history, we focus our attention in this chapter on two extensions of the traditional SDT model, and the implications of each of these newer approaches for topics in applied perception. First, the extension of SDT from a single to multiple dimensions is discussed, followed by a discussion of a more recent advance, fuzzy signal detection theory. We finish with some thoughts on future theoretical developments concerning the variable of time.

Multidimensional Signal Detection Theory

Traditional SDT represents the perceptual effects of a stimulus on a unidimensional scale. Even stimuli known to vary along multiple dimensions are frequently collapsed onto a single evidence axis (e.g., mammogram features such as shape, size, and darkness; Swets, 1995). This simplicity fosters broad application. Computation of signal detection indices requires only response-accuracy data: percentage of "correct" responses to "signal" and incorrect responses to "noise" events allows computation of the measures of perceptual sensitivity (d') and decision criterion (β).

Despite its long history and considerable success, the theory continues to evolve. An important development has been generalization of SDT to the multidimensional case. This serves to provide a unique representation to perceptual effects from each of a number of stimulus dimensions. Known as "general recognition theory" (GRT), the mathematics of this multidimensional extension were derived by Ashby and Townsend (1986; also see Pastore and Sorkin, 1972; Tanner, 1956). GRT is often called "decision bound theory" since the univariate decision criterion of SDT is replaced in GRT by a quadratic decision boundary that divides the perceptual space into response regions (Ashby, 1992). Note that a distinction is usually made between the theoretical framework of GRT and the phrase "multidimensional signal detection analysis" (MSDA), which refers to a method of analysis related to standard SDT but anchored in the GRT framework (see Kadlec and Townsend, 1992).

The advantage in using GRT lies in the ability to uncover interactions among percepts created by separate physical stimulus features (e.g., does perceived loudness affect the experience of timbre?). More formally, are the perceptual effects resulting from each stimulus feature statistically independent of one another? Previous techniques for answering this question relied on inconsistent definitions of perceptual independence, and on assessment methods that do not distinguish between perceptual and decisional contributions to feature interaction (Maddox, 1992). The GRT framework resolves this by providing a complete set of definitions for several forms of independence, and, in

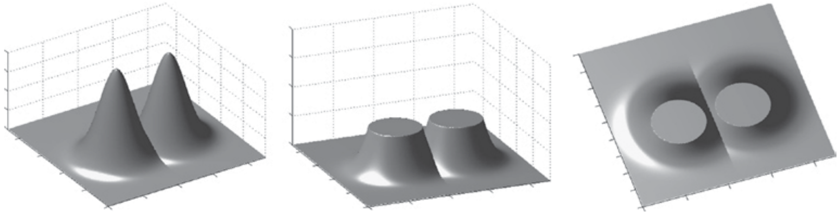


Figure 3.1. *Left panel: overlapping bivariate-normal distributions. Center panel: the same distributions with likelihood values fixed at a single value to produce contours of equal likelihood. Right panel: the same contours of equal likelihood at an angle perpendicular to the stimulus coordinate plane. The shape of each contour summarizes variance and covariance along the stimulus dimensions, as well as an emergent linear decision boundary.*

keeping with the SDT tradition, makes a distinction between perceptual and decisional effects (Ashby and Townsend, 1986; Ashby 1992).

In GRT, the term "perceptual independence" (PI) refers to the noninteraction of dimensional percepts created in presentation of a single stimulus (e.g., for a given tone, is perception of timbre correlated with perception of amplitude, or are they independent?). Across multiple unique stimuli presented at different levels of one dimension or another, GRT refers to noninteraction as "separability" (e.g., is perception of timbre the same across three stimuli with low, medium, and high amplitudes?). Within the GRT framework, these can be identified as either "perceptual separability" (PS) or "decisional separability" (DS). In other words, violations of separability can result during the perceptual stage of processing, during a decision stage that takes such perceptual effects as input, or during both.

It helps to understand PI, PS, and DS by comparing them in a multivariate space (using the simplest case of two dimensions). Figure 3.1 displays overlapping bivariate-normal distributions – the SDT model in two dimensions. The usual two-dimensional (2D) curves are replaced with three-dimensional "bell" shapes whose heights represent the likelihood of any given point in the space along perceptual dimensions A and B. Although this representation is illustrative, it is actually more complex than needed to evaluate PI, PS, and DS. We need only view the position of each curve (i.e., its means), as well as each of its variances and covariances (indicated by each distribution's shape). We can simplify this comparison process by slicing a section through the curves at a uniform height (i.e., likelihood value; see the middle panel of Figure 3.1) and observing the distributions with a viewpoint perpendicular to the plane created by the coordinate axes. This viewpoint makes it easy to observe distribution means, variances, and covariances (see the circular "contours of equal likelihood" in the far right panel of Figure 3.1.

To illustrate further, Figure 3.2 displays contours of equal likelihood for distributions from a feature-complete identification task (i.e., a complete

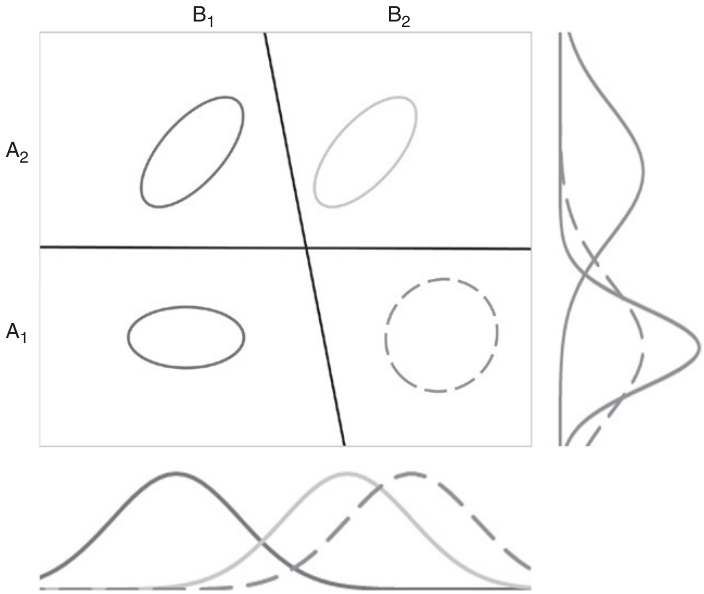


Figure 3.2. *Hypothetical contours of equal likelihood from a complete identification experiment. The marginal distributions portray category means and variances for each dimension. The horizontal line illustrates a bound consistent with decisional separability (DS), while the tilted vertical line illustrates one possible violation of DS.*

factorial design with two stimulus values on dimension A and two values on dimension B). Perceptual independence (PI), perceptual separability (PS), and decisional separability (DS) can all be observed in this figure (note also the marginal distributions presented outside the axes).

To assess perceptual independence, the shape of a single distribution (but not the location of its mean) is important. The reason for this is illustrated in Figure 3.2. Perceptual independence implies a lack of correlation between the perceptual effects for a single stimulus. In Figure 3.2, the bottom contours depict two cases of PI, but the top contours illustrate a violation of perceptual independence. The bottom right contour is circular, signifying both equal variance along each dimension for that stimulus, and zero correlation between dimensions. In the bottom left contour, the variances along each dimension are not equal (seen more clearly in the marginal distributions), but the major and minor axes of the distribution are orthogonal to the coordinate axes (indicating a lack of correlation between percepts for that stimulus and thus PI). The top distributions have equal variance along each dimension, but their tilt indicates a positive correlation between perceptual effects. For example, on a presentation of stimulus A_2B_2, if the level of B_2 is perceived to be at the high end of the distribution, then the level of A_2 will also be perceived as high.

Perceptual separability (or its converse, "perceptual integrality") is determined over a set of distinct stimuli. If, for example, the brightness of a color patch is perceived differently depending on whether presented at a low or high saturation level, then PS is violated (i.e., the dimensions are perceptually integral). This can manifest itself as either a mean shift across levels (e.g., in Figure 3.2 perception of dimension A at level A_2 depends on whether dimension B is at level B_1 or B_2) or as a change in variance across levels (e.g., perception of dimension B at level B_1 changes depending on whether dimension A is at A_1 or A_2).

Finally, an observer may render a decision about a stimulus based on integrating perceptual information from each dimension, or instead by focusing attention selectively on a single dimension. DS corresponds to the latter case (i.e., selective attention) and is indicated by a decision bound that is orthogonal to the coordinate axes (e.g., the horizontal line in Figure 3.2). The placement of this decision bound can be anywhere along an axis, just as the β value is set in univariate SDT. A violation of DS corresponds to observation of any decision bound that is not orthogonal to the coordinate axes (e.g., the diagonal bound in Figure 3.2) and implies a decision rule that integrates perceptual information from multiple dimensions. The next section reviews steps in a multidimensional SDT analysis so that PI, PS, and DS can be empirically evaluated.

Multidimensional Signal Detection Analysis

There are two approaches to analyzing data within the multidimensional GRT framework. The first – actually referred to as multidimensional signal detection analysis (MSDA) – computes variations of the standard SDT measures, d' and β, across levels of the stimulus dimensions. The second approach is to fit a set of models to data to instantiate and evaluate different GRT hypotheses (see Maddox and Ashby, 1996; Thomas 2001). Both can be used to assess PI, PS, and DS, although the techniques differ in their complexity and power. To facilitate comparison to univariate SDT, we focus most of our discussion here on MSDA, although we also touch briefly on model fitting.

MSDA requires a complete identification design (i.e., at least two stimulus dimensions with at least two levels of presentation each). Response probabilities are estimated by response proportions in a confusion matrix (e.g., using the dimensions from Figure 3.2, a 4 × 4 matrix representing each stimulus: A_1B_1, A_1B_2, A_2B_1, A_2B_2, and corresponding correct and incorrect responses: a_1b_1, a_1b_2, a_2b_1, a_2b_2).

To evaluate PS and DS, MSDA utilizes marginal response probabilities (response proportions collapsed across one stimulus-dimension level in the confusion matrix). Because this analysis focuses on effects across multiple stimuli, it is referred to as "macroanalysis." The marginal values are used to compute d' and β along one dimension, but at each level of the other dimension for comparison. For example, if d' on dimension A is the same regardless of whether computed at level B_1 or B_2, then the conclusion of PS is supported (i.e., the means and variances along dimension A are unaffected by the level of dimension B). A more trustworthy conclusion of PS is supported when the distributions form a rectangle (i.e., d' values for each dimension are unaffected by levels of the other dimension). Similarly, the decision criterion, β, is computed in the same manner and compared across levels of a dimension to determine whether DS holds.

To examine whether within-stimulus PI holds, d' values are computed using conditional-response probabilities from the confusion matrix (analysis at the level of a single stimulus is called "microanalysis"). For a single stimulus (in the bivariate case), two pairs of d' values must be derived – a pair for each dimension level that is represented in the stimulus (e.g., stimulus A_1B_2 from Figure 3.2). In each pair, a separate d' value is computed on either side of the decision bound. In the language of SDT, "hit" and "miss" rates are used to evaluate stimuli at dimension-level 2; "false alarm" and "correct rejection" rates are used to evaluate stimuli at dimension-level 1. In other words, when computing d' along dimension A at level B_2, if d' based on "hit" rate matches d' based on "miss" rate, and if the same is true when computing d' along dimension B at level A_1, then PI is supported for stimulus A_1B_2. However, this conclusion can only be reached when DS holds for each dimension; otherwise it is impossible to distinguish between violations of PI and PS. Furthermore, such evidence for PI is indirect; only violations of PI can be directly observed in any particular data set (Kadlec and Townsend, 1992).

The basic postulates of MSDA have been described by Kadlec and Townsend (1992). For a clear introduction to the mathematics of MSDA, the works of Kadlec (2001) and Wenger and Ingvalson (2003) provide excellent examples of application of the MSDA procedures (see also Farris, Viken, and Treat, 2010).

The second type of GRT analysis uses parameter estimation techniques to fit a series of models, each embodying different hypotheses about perceptual representation and/or decision bounds, to data for hypothesis comparison. This approach generates the same summary information as MSDA, but also allows evaluation of highly complex decision bounds. For example, the optimal decision bound (that maximizes long-run response accuracy) separating distributions with identical shape is linear (shown in the left panel in Figure 3.3), while unequal variances or covariance across distributions result in an optimal bound that is curvilinear (as shown in the right panel of Figure 3.3). (Excellent descriptions of this procedure can be found in Ashby and Maddox, 1993; Maddox and Ashby, 1996; Thomas 2001.) We do, however, consider examples of this approach in the next section on applications of GRT. Model fitting has been especially popular in the case of perceptual categorization research, where GRT has received the most application and validation.

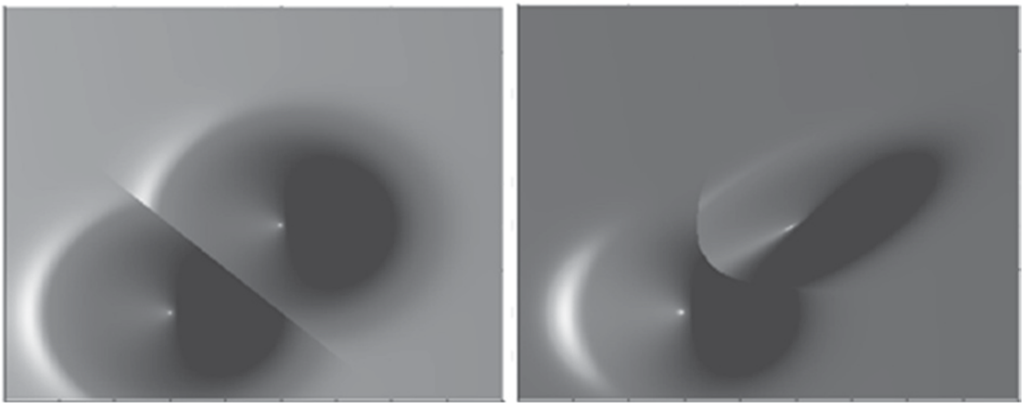


Figure 3.3. *Left panel: overlapping distributions with equal variance/covariance, resulting in a linear optimal decision bound. Right panel: overlapping distributions with nonidentical variance/covariance values, resulting in a nonlinear optimal bound.*

Applications of MSDA/GRT

The principles and techniques of perceptual independence, separability, and decisional separability have been applied to development of principles for display design (e.g., the use of configural displays). In general, the identification of the dimensionality of stimuli and how those dimensions are related to one another can be used to improve legacy displays and design entirely new displays, and in addition, to design better training procedures for both improving discriminability and adopting appropriate decision criteria.

The most prominent application of GRT applies the model-based approach to studying categorization performance in laboratory tasks with simple stimuli. GRT has provided insights into the effect of stress on performance (Markman et al., 2006; Ell et al., 2011), regulatory focus (attention to gains or losses) during decision-rule learning (Maddox et al., 2006), the influence of category base rates and payoffs on decision-criterion learning (Maddox and Bohil, 2003; Bohil and Maddox, 2003), and several other phenomena. GRT also underlies an influential neuropsychological theory that posits separate explicit- and implicit-rule learning systems in categorization (Ashby et al., 1998).

GRT has also been applied in neuropsychological research, yielding insights into how biological structures can be related to specific cognitive deficits (Maddox and Filoteo, 2007). Parkinson's disease patients, for example, are known to show deficits in tasks requiring selective attention. Applying GRT models to classification results, Filoteo and Maddox (1999) observed that Parkinson's disease patients can indeed apply a selective-attention decision criterion like the horizontal (DS) bound in Figure 3.2, but that they apply the rule with much greater variability in placement over trials than normal control participants. We have also demonstrated that Parkinson's disease patients with damage to the tail of the caudate have difficulty learning to integrate information from multiple stimulus dimensions. Amnesic patients, on the other hand, with medial temporal lobe damage can learn information-integration decision rules as well as normal control participants. GRT's ability to disentangle decisional and perceptual effects across separate stimulus dimensions facilitates assessments like these.

GRT has also been applied to other topics of clinical interest. Farris et al. (2010) recently examined the ability of males to identify signals of female sexual interest, obviously relevant for understanding incidence of sexual coercion. Men viewed photos of women expressing body language that conveyed either "friendliness" or "sexual interest." This variable was varied along with style of dress ("conservative" or "provocative"). Interactions between these dimensions potentially lead to misunderstandings – particularly among men with a history of sexual aggressiveness. Participants assessed to be at higher risk for sexual aggression were more likely to perceive an illusory correlation (i.e., they showed violations of perceptual independence) between body language and style of dress, along with increased perception of sexual interest for stimulus distributions corresponding to a provocative style of dress.

Another area of research with implications for body language interpretation is face perception (Wenger and Ingvalson, 2003; Richler et al., 2008; Cornes et al., 2011; also see Thomas, 2001 for a model-based application of GRT in this domain). Researchers have found evidence for PS, but they also find violations of DS. This suggests that faces may not be processed holistically at a perceptual level (as is often argued), but that a decisional process may be responsible for, or at least involved in, face processing that is apparently holistic (i.e., independent perceptual inputs may be integrated at a decision stage). GRT makes insights such as these possible. Future research on deception detection could take advantage of MSDA to shed light on interactions between dimensions known to influence emotion perception (Meeren, van Heijnsbergen, and de Gelder, 2005).

An application with relevance to technology design is taken from research on sensor fusion. Different sensing devices can provide unique views (e.g., images from the visible- and infrared-light spectrum) of a scene or object, and combining these into a single composite image can convey this information in a compact form that allows observers to view the inputs simultaneously. For instance, functional magnetic resonance imaging (fMRI) and computed tomography (CT) scan views can be combined into a single image when evaluating potential tumor sites. The composite view can reduce the working memory demands involved in comparing separate images by eliminating the need to shift attention.

Despite this gain, it is possible that some information is lost or obscured when images are combined. McCarley and Krebs (2006) used MSDA to explore whether one input channel may obscure information from another. Participants viewed images of an airplane fused from a combination of

Table 3.1. *The four outcomes associated with the state of the world and observer response for traditional signal detection theory*

State of the world		Response	
		Yes	No
	Signal present	Hit	Miss
	Signal absent	FA	Correct rejection

two levels of contrast (high or low) and two sensor wavelengths (long- and medium-wavelength single-band images were fused to create dual-band images). MSDA exposed violations of perceptual separability. Sensitivity (*d'*) for the contrast level of one component image was influenced by the contrast of the other component image.

There are a variety of sensor fusion techniques and many unanswered questions in this domain regarding the benefits and pitfalls of sensor fusion. As the authors suggest, GRT might provide a unifying framework for future work in this area. This example underscores the value of MSDA for understanding integrality and/or separability of stimuli known to vary on multiple dimensions. A related application may be to understand the potential for information loss due to spatial occlusion in the rapidly growing field of augmented reality. There are likely many other similar domains of application.

MSDA/GRT addresses the limitation of stimulus unidimensionality in SDT. While MSDA/GRT is one way in which SDT has been expanded and elaborated, it is not the only one. Another avenue of progress has been the integration of SDT with fuzzy set theory (Zadeh, 1965). Fuzzy signal detection theory addresses the problem associated with defining stimulus events.

Fuzzy Signal Detection Theory

Fuzzy signal detection theory (FSDT) was derived by Parasuraman, Masalonis, and Hancock (2000; order to incorporate fuzzy set concepts into SDT. As such, FSDT represents a generalization of SDT (Hancock et al., 2000), which traditionally requires a binary decomposition of the state of the world into mutually exclusive categories of signal and nonsignal. This is shown in Table 3.1. In MSDA/GRT as well as in applications of SDT to more than two categories of classification along a single dimension (for reviews regarding the latter see Macmillan and Creelman, 2005; Wickens, 2002), the underlying state of the world consists of mutually exclusive category membership for each stimulus event.

In FSDT this binary distinction between signal and noise is superseded by fuzzy mapping, such that membership of a stimulus event can be in both categories "signal" and "nonsignal." In contrast to "crisp" set theory, in which events are categorized into mutually exclusive categories (e.g., an object is either in category A or in the category "not A"), in fuzzy set theory events may be simultaneously assigned to more than one category to different degrees or proportions of membership. In the context of SDT categories, in FSDT events can have a degree of membership in each category, thus simultaneously being both signal (to a degree) and nonsignal (to a degree). Similarly, possible responses are extended from mutually exclusive categories – either "yes" or "no" – to include intermediate values (e.g., "might be a signal"). Mathematically it is convenient to represent these fuzzy memberships for the set "signal" (s) and the set "yes" (r) as varying between 0 and 1, with 0 representing complete nonmembership in the set (i.e., a "nonsignal" stimulus and response "no" in traditional SDT) and with 1 representing complete membership in the set (i.e., "signal" and "yes"). A value of 0.5 therefore corresponds to equal degrees of membership in each category.

Defining *s* and *r*: Mapping Functions

The key to FSDT is a valid function mapping the theoretical dimension (in this case, ***s*** and ***r***) to an established sensory or perceptual dimension in the operational environment. That is, the psychological construct that defines the dimension of interest (e.g., friend/foe identification, improvised explosive device [IED] detection, baggage screening, air traffic control conflicts) should be well understood. Given such a construct, a function mapping levels along this dimension to fuzzy membership values is established, either theoretically or empirically. Parasuraman et al. (2000) have suggested that a sigmoid function is often a good candidate for describing mapping functions for ***s***, although FSDT does not constrain the form of this function.

Derivation of FSDT Indices

The computational formulas for deriving fuzzy sensitivity and response bias measures are described in Parasuraman et al. (2000). In short, each event presented to the observer is assigned a value of membership in the set ***s*** and the response of the observer is assigned a degree of membership in the set ***r*** according to their respective mapping functions. These are then used to derive degrees of membership in the sets "Hit" and "False Alarm." (Membership in the sets "Miss" and "Correct Rejection" can also be computed, but these are not necessary for SDT computations since they are the complements of the Hit and False Alarm rates.) These degrees of memberships are summed over the total number of trials presented to yield hit rate and false alarm rates (HR and FAR, respectively). These can then be transformed to measures of sensitivity (*d'*) and response bias (β) using the standard SDT formulas (e.g., see Macmillan and Creelman, 2005).

Tests of FSDT

Given its relatively recent introduction, there have been few applications of FSDT to operational environments (for exceptions see Castanho, Barros, Yamakami, and Vendite, 2007; Lu, Hinze, and Li, 2011; Masalonis and Parasuraman, 2003). However, there have been several studies to test the validity of the model itself. This series of studies derived from a concern that in the presentation of FSDT, Parasuraman et al. (2000) made an interesting but untested assumption. They wrote, "because the fuzziness of the signal has already been captured in the definition of *s* and *r*, and from the fuzzy HR [and FAR], the traditional *d'* formula can be used in fuzzy SDT analysis" (p. 649). Similarly, β (or presumably any parametric or nonparametric measure of response bias) can also be computed from HR and FAR. Although these statements are computationally straightforward, it is less clear that they are theoretically tenable. This is because the computational formulas for parametric SDT measures rest on assumptions regarding the underlying decision space, shown in Figure 3.4. This representation rests on the definition of the state of the world and response set shown in Table 3.1. However, FSDT *explicitly assumes* that the underlying structure of the state of the world is *not* in such a form. Hence, it *may* be valid to transform fuzzy HR and FAR into SDT measures, but *only* if FSDT conforms to the assumptions of traditional SDT. If this is not the case, then new measures must be derived for FSDT.

Szalma and his colleagues therefore conducted ROC experiments to test the main SDT assumptions, and to determine whether an ROC function derived from a fuzzy analysis was consistent with that for traditional ROC analysis, that is, $Z_H = bZ_F\text{-}d'$ (Murphy, Szalma, and Hancock, 2004; Szalma and O'Connell, 2011; Szalma, Oron-Gilad, Saxton, and Hancock, 2006). In general, the evidence indicated that FSDT results do in fact conform to the core assumptions of SDT. In addition, there is evidence that changes in the distribution of fuzzy signals show shifts in criterion setting that are consistent with shifts in signal probability in crisp signal detection applications (Stafford, Szalma, Hancock, and Mouloua, 2003). Thus, it appears that the traditional formula for computing sensitivity and response bias can indeed be applied to fuzzy hit and false alarm rates.

However, it has also been observed that the range of criterion setting in FSDT tasks is much narrower than that observed in more traditional SDT tasks. This has been shown both empirically (Szalma et al., 2006) and in a

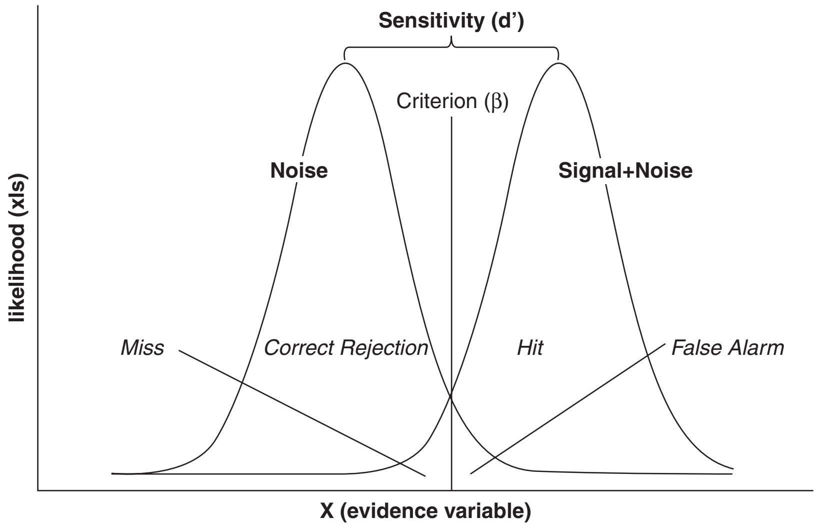


Figure 3.4. *Representation of the decision space for traditional SDT.*

recent Monte Carlo simulation (Szalma and O'Connell, 2011). One possible reason is that the typical payoff instructions used to manipulate criterion setting may not be easily applied to observers in a FSDT task. For instance, informing observers that they will be awarded 1 point for each correct detection but penalized 10 points for each false alarm would be represented in a FSDT context as "the degree to which" observers make a hit or false alarm. In the future it may be necessary to adjust standard payoff procedures to reflect the fuzziness of the stimulus dimension itself. Whatever the reason for the narrow range of criterion setting, at a theoretical level, future work should seek to clarify the meaning of a fuzzy criterion.

Application of FSDT Methods to Applied Perception

The first step in applying FSDT methods is to derive a valid mapping function relating the fuzzy membership in the set "signal" to the perceptual dimension of interest, and in deriving a corresponding response set appropriate to the task. With these functions established one can use the formula described by Parasuraman et al. (2000) to derive outcome measures. On the basis of the work of Szalma and Hancock (2013), the fuzzy hit rate and false alarm rate can be used to compute the desired SDT measures. In collecting such data, it is important to ensure that the instructions to observers define signal and nonsignal clearly and in such a way that the desired level of fuzziness is obtained in the observer's representation of the relevant perceptual dimension, and that the observers understand the appropriate response for different levels of "signalness."

In sum, fuzzy signal detection theory and general recognition theory represent different types of theoretical generalization of signal detection theory. These analytical tools currently are gradually proving their worth for applied perception research. In the next section, we consider another frontier for generalization – time – that is only beginning to be explored in detail.

The Importance of Time and Memory: Issues for Future Research and Theoretical Integration

In an early presentation of SDT (Tanner and Swets, 1954) the assumption was made that an observer only considers the information on a given trial in making a decision for that trial, such that the responses to a series of trials are independent of one another. But we must ask whether this is necessarily so. We believe that the answer is a resounding No! Furthermore, in the same way that humans learn much from trial to trial, we believe they can also learn *within* one trial – when that trial is elaborated over time.

That is, typical sensory recognition tasks provide a singular brief event of perhaps one second or less in duration, but what happens when the search happens over multiple seconds or even minutes? As pointed out by Hoffman and Fiore (2007), "Meaningful patterns sometimes exist only over time." We believe that observers progressively collect information toward a criterion at which point a response (of some nature) is emitted. As we shall note, in many complex situations, this response can also be ongoing and probabilistic rather than a one-time button press.

Thus, the critical question here is time – time for both the observation to accumulate and the response to elaborate. This dimension can be encapsulated using changing membership functions for signal (s) and response (r). At some point in time, one has to "fish or cut bait."

That is, critical decision points exist in which one has to decide whether sufficient information has been collected such that a response can be made. In typical psychological sensory experiments, this happens fast enough to be considered virtually an instantaneous decision. Few situations in complex, applied worlds impose such a temporal imperative. Even in time-critical circumstances such as aviation, it is often best to "take a moment." Thus, while a theoretical extension like GRT is a most helpful process, we shall need a GRT for time (for developments along these lines, see Pleskac and Busemeyer, 2010). On the positive front, our methods of evaluation are growing in sophistication and reliability. Our next great challenge is a theory of context to specify where and when each sequential methodological development is most relevant and diagnostically most potent.

Conclusion

The value in signal detection theory is that it provides a deeper understanding of the mental processes underlying choice behavior than can be gained via aggregate response accuracy measures. Accuracy data let us compare performance across different individuals or points in time, but tell us nothing about *why* performance appears as it does. The signal detection model parses performance data to yield a finer-grained explanation, providing separate estimates for perceptual and decisional components. Similar arguments can be made for the generalizations described in this chapter. Each promises even deeper behavioral insights than the traditional SDT model. Because SDT is a special case of each generalization, they offer higher explanatory acuity while maintaining the strengths of the traditional model.

When a theoretical generalization is proffered, it requires thorough vetting through careful experimentation. This may take the form of laboratory experiments or empirical investigations conducted in the “field” setting (see Hoffman and Deffenbacher, 2011). Ideally, this would occur before confident conclusions were drawn from research conducted on applied topics. Since its unveiling more than 25 years ago, GRT has seen substantial adoption in basic research programs, and it is gradually finding its way into applied research, as discussed previously. The MSDA approach can be applied in the same operational domains and contexts as the traditional SDT model and is relevant whenever dimensional interactions may be of interest. The more recent FSDT generalization is currently gaining its sea legs in various laboratory studies, but it too has been adopted in applied research. It likely will continue finding application in situations where yes/no judgments may be required, but where differing degrees of certainty are likely to underlie observable performance. Finally, although sequential sampling models currently capture some of the dynamic elements of choice tasks, additional work is needed to accommodate the relatively long decision time scales and within-trial learning effects potentially encountered in natural decision-making settings.

References

Ashby, F. G. and Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 37, 372–400.

Ashby, F. G. and Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154–179.

Ashby, F. G. (1992) (Ed.). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U. and Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, 105, 442–481.

Bohil, C. J. and Maddox, W. T. (2003). A test of the optimal classifier's independence assumption in perceptual categorization. *Perception and Psychophysics*, 65, 478–493.

Castanho, M. J. P., Barros, L. C., Yamakami, A. and Vendite, L. L. (2007). Fuzzy receiver operating characteristic curve: An option to evaluate diagnostic tests. *IEEE Transactions of Information Technology in Biomedicine*, 11, 244–250.

Cornes, K., Donnelly, N., Godwin, H. and Wenger, M. J. (2011). Perceptual and decisional factors influencing the discrimination of inversion in the Thatcher Illusion. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 645–668.

Dember, W. N. (1998). If you come to a fork in the road. In R. R. Hoffman, M. F. Sherrick and J. S. Warm (Eds.), *Viewing psychology as a whole: The integrative science of William N. Dember* (pp. 629–649). Washington, DC: American Psychological Association.

Ell, S. W., Cosley, B. and McCoy, S. K. (2011). When bad stress goes good: increased threat reactivity predicts improved category learning performance. *Psychonomic Bulletin and Review*, 18, 96–102.

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin and Review*, 9, 3–25.

Farris, C., Viken, R. J. and Treat, T. A. (2010). Perceived association between diagnostic and non-diagnostic cues of women's sexual interest: General recognition theory predictors of risk for sexual coercion. *Journal of Mathematical Psychology*, 54, 137–149.

Filoteo, J. V. and Maddox, W. T. (1999). Quantitative modeling of visual attention processes in patients with Parkinson's disease: Effects of stimulus integrality on selective attention and dimensional integration. *Neuropsychology*, 13, 206–222.

Hancock, P. A., Masalonis, A. J. and Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomics Science*, 1, 207–230.

Hoffman, R. R. and Deffenbacher, K. A. (2011). A multidimensional analysis of the relations of basic and applied psychology. *Theoretical Issues in Ergonomic Science*. DOI: 10.1080/1464536X.2011.573013.

Hoffman, R. R. and Fiore, S. M. (2007). Perceptual (re)learning: A leverage point for human-centered computing. *IEEE Intelligent Systems*, 1541–1672.

Kadlec, H. (2001). Signal detection theory: Multidimensional. In N. J. Smelser, J. Wright and P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences, 14082–14085*. Oxford, UK: Elsevier.

Kadlec, H. and Townsend, J. T. (1992). Signal detection analyses of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181–231). Hillsdale, NJ: Lawrence Erlbaum.

Lu, Y., Hinze, J., and Li, Q. (2011). Developing fuzzy signal detection theory for workers' hazard perception measures on subway operations. *Safety Science*, 49, 491–497.

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

Maddox, W. T., (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale, NJ: Erlbaum.

Maddox, W. T. and Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification–speeded classification relationship. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 795–817.

Maddox, W. T., Baldwin, G. C., and Markman, A. B. (2006). A test of the regulatory fit hypothesis in perceptual classification learning. *Memory and Cognition*, 34, 1377–1397.

Maddox, W. T. and Bohil, C. J. (2003). A theoretical framework for understanding the effects of simultaneous base-rate and payoff manipulations on decision criterion learning in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 307–320.

Maddox, W. T. and Filoteo, J. V. (2007). Modeling visual attention and category learning in amnesiacs, striatal-damaged patients, and normal aging. In R. W. J. Neufeld (Ed.), *Advances in clinical-cognitive science: Formal modeling and assessment of processes and symptoms* (pp. 113–146). Washington, DC: American Psychological Association.

Markman, A. B., Maddox, W. T. and Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science*, 17, 944–948.

Masalonis, A. J. and Parasuraman, R. (2003). Fuzzy signal detection theory: Analysis of human and machine performance in air traffic control, and analytic considerations. *Ergonomics*, 46, 1045–1074.

McCarley, J. S. and Krebs, W. K. (2006). The psychophysics of sensor fusion: A multidimensional signal detection analysis. In Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting.

Meeren, H. K. M., van Heijnsbergen, C. C. R. J. and de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences*, 102, 16518–16523.

Murphy, L., Szalma, J. L. and Hancock, P. A. (2004). Comparison of fuzzy signal detection and traditional signal detection theory: Analysis of duration discrimination of brief light flashes. *Proceedings of the Human Factors and Ergonomics Society*, 48, 2494–2498.

Parasuraman, R., Masalonis, A. J. and Hancock, P. A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors*, 42, 636–659.

Pastore, R. E. and Sorkin, R. D. (1972). Simultaneous two-channel signal detection: I. Simple binaural stimuli. *Journal of the Acoustical Society of America*, 51, 544–551.

Peterson, W. W., Birdsall, T. G. and Fox, W. C. (1954). The theory of signal detectability. *Transactions of the Institute of Radio Engineers. Professional Group on Information Theory,* PGIT-4, 171–212.

Pleskac, T. J. and Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117, 864–901.

Richler, J., Gauthier, I., Wenger, M. and Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 328–342.

Stafford, S. C., Szalma, J. L., Hancock, P. A. and Mouloua, M. (2003). Application of fuzzy signal detection theory to vigilance: The effect of criterion shifts. *Proceedings of the Human Factors and Ergonomics Society*, 47, 1678–1682.

Swets, J. A. (1973). The relative operating characteristic in psychology. *Science*, 182, 990–1000.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293.

Swets, J. A. (1995). Separating discrimination and decision in detection, recognition, and matters of life and death. In D. Scarborough and S. Sternberg (Eds.), *An invitation to cognitive science: Methods, models, and conceptual issues*. (pp. 635–702). Cambridge, MA: The MIT Press.

Szalma, J. L. and Hancock, P. A. (2013). A signal improvement to signal detection analysis: Fuzzy SDT on the ROCs. *Journal of Experimental Psychology: Human Perception and Performance*, 39(6), 1741–1762.

Szalma, J. L. and O'Connell, M. E. (2011). Fuzzy signal detection theory: A Monte Carlo investigation. *Proceedings of the Human Factors and Ergonomics Society*, 55, 1365–1369.

Szalma, J. L., Oron-Gilad, T., Saxton, B. and Hancock, P. A. (2006). Applications of fuzzy signal detection theory to the discrimination of morphed tank images. *Proceedings of the Human Factors and Ergonomics Society*, 50, 1716–1720.

Tanner, W. P. (1956). Theory of recognition. *Journal of the Acoustical Society of America*, 28, 882–888.

Tanner, W. P. and Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401–409.

Thomas, R. D. (2001). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception and Psychophysics*, 63, 625–650.

Wenger, M. J. and Ingvalson, E. M. (2003). Preserving informational separability and violating decisional separability in facial perception and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1106–1118.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford, UK: Oxford University Press.

Zadeh, L. A. (1965) Fuzzy sets. *Information and Control*, 8, 338–353.

4 The Measurement of Perceptual Resources and Workload

David B. Boles and Michael B. Dillard

The Resource Concept

Resources are mental capacities that both support and limit performance. As originally conceived, resources constituted a single attentional "pool" from which tasks draw. Daniel Kahneman (1973) proposed that a single task, or even two simultaneous tasks, could be performed without effective limit as long as demand did not exceed the pool's capacity. However, once single or joint demand exceeded capacity, performance was proposed to degrade in proportion to the excess demand.

The single-pool concept of resources soon succumbed to observations that interference between tasks depends on similarities between the tasks (Brooks, 1968; Klee and Eysenck, 1973; Wickens and Sandry, 1982; Wickens, Sandry, and Vidulich, 1983). This should not happen with a single undifferentiated pool, because interference is predicted to depend only on the summed overall demands of the tasks. Similarity-based interference implies that the pool of resources is instead differentiated, so that the demands of two tasks on the same subset of resources cause more interference than do their demands on different subsets of resources.

From these empirical observations emerged multiple resource theory (MRT; Wickens, 1984). In Chris Wickens's seminal cube model, resources were envisioned as constituting three dimensions. The *modality* dimension consisted of visual and auditory resources on the input side of processing, and of manual and vocal resources on the output side. The *processing code* dimension comprised verbal and spatial resources, where "code" was taken to mean one way of mentally representing a stimulus. Finally, the *stages of processing* dimension attributed resources to encoding/central processing and response stages of processing. It is noteworthy that by conceptualizing resources that could be characterized as visual or auditory, and as verbal or spatial, MRT in effect recognized that resources can be perceptual in nature.

What Are Resources?

Subsequent research showed that characterizing dual tasks in terms of the similarity of their demands along these dimensions predicted performance better than simply summing their overall demands, a clear success of the multiple resources model (Wickens et al., 1988). However, research also showed that as a particular instance of that model, MRT was incomplete. Specifically, multiple spatial codes were found to exist, with dual tasks using separate codes interfering less than if they used the same code. Evidence was likewise found for a temporal code not characterizable as either verbal or spatial (Boles and Law, 1998). In response to these developments, Boles and Law proposed that MRT be expanded by incorporating a number of resources uncovered in factor analyses of perceptual asymmetries. Subsequently this suggestion was formalized by the creation of the Multiple Resources Questionnaire (MRQ), which included 17 resources, many of them perceptual in nature (Boles and Adair, 2001a).

Fuller consideration of the MRQ follows, but one development that accompanied its creation is of particular note. This is that the adoption of a nondimensional structure – that is, a set of 17 resources instead of three dimensions as in the cube model – effectively abandoned any conceptualization of resources as being necessarily attentional in nature. To understand why, it is first important to understand how Wickens' cube model is consistent with an attentional conception of resources.

Because the model consisted of three dichotomous dimensions, attention could be viewed as splitting at most once. For example, a task presented visually, involving spatial coding and manual response, could conceivably involve a separate attentional stream from a task involving auditory input, verbal coding, and vocal response, especially if the encoding/central processing and response states of processing do not coincide in time. But if resources are instead viewed as a nondimensional set of capacities as in the MRQ, then gradations of overlap, and thus varying degrees of interference, could theoretically be observed among any number of simultaneous tasks. If so, an attentional framework is not tenable because any number of simultaneous tasks could imply not just one but multiple streams of attention.

So if resources are not best viewed as attentional, what are they? Boles and Law (1998) put forward the proposition that "orthogonal processes represent orthogonal resources." This view raises the possibility that resources correspond to the processes themselves that are manifested in the perceptual-cognitive system, and not to our ability to segregate processes into separate streams. Rugg (1986) had earlier expressed a similar viewpoint by indicating that if we could adequately specify the cognitive architecture responsible for information processing, then the number of resources would correspond to the number of active components.

Although Boles and Law (1998) stopped short of favoring the processes viewpoint over the attentional one, the processes viewpoint is more consistent with the later construction of the MRQ (Boles and Adair, 2001a). Construction in large part followed the results of factor analytic studies of lateralized perceptual processes. These studies employed fairly large sets of tasks that produce behavioral evidence of the asymmetric function of the brain, for example, by revealing differences in performance between the left and right visual fields, ears, or hands. Factor analysis of such asymmetries revealed a number of independent mental processes, most of them perceptual in nature (Boles, 1991, 1992, 1996). Given the origin of MRQ items in factor analyses of asymmetric function, one may question how resources are connected to brain mechanisms.

How Are Resources Linked to Brain Mechanisms?

If resources are process-based, and processes are brain-based, then functional imaging might profitably be brought to bear on the nature of resources. Imaging might reveal a separate brain substrate for each independent resource. In addition, within functional images of dual-task competition, evidence might be found for common resources in the form of interaction among brain areas. If activation in an area is different when two tasks are performed together than when they are performed separately, that area is a likely candidate as a site of resource competition between the tasks. While it might also represent processes related to task coordination, an appropriate experimental design could presumably differentiate these alternatives.

Unfortunately, in spite of the importance of the resource-brain linkage, research efforts in this area have been very limited. Such efforts have generally emerged from a cognitive psychology literature on the psychological refractory period, using pairs of tasks with varying temporal offset. That literature stresses the importance of a central bottleneck involving response selection and memory retrieval processes rather than bottlenecks of processes more generally (Pashler, 1994). Brain-related studies investigating the central bottleneck usually attempt to minimize perceptual interactions by using stimuli in different modalities, that is, visual and auditory (Dux, Ivanoff, Asplund, and Marois, 2006; Mochizuki et al., 2007; Sigman and Dehaene, 2008). An exception is a study by Gazes et al. (2010), which did attempt to examine perceptual process interference as part of a larger picture including response processes. Although no perceptual interference was found, the study's use of a visual tracking task and a visual color matching task can be considered to test for perceptual process interference only if it is assumed that vision is a general resource. While this assumption is made by MRT (Wickens, 1984), it is *not* assumed to be true by the MRQ, where modalities and processes are hypothesized to be integrally linked. Thus the MRQ postulates 10 separate visual resources, each involving one process (Boles, Bursk, Phillips, and Perdelwitz, 2007). Again, under the view that "orthogonal processes represent orthogonal resources," these are all potentially *independent* sources of dual-task interference.

Thus to date, attempts to relate perceptual resources to brain mechanisms have been inadequate in design and scope. For now, the view favored here is that resources are mental capacities intimately linked to a full range of processes including ones that can be characterized as perceptual, memorial, motoric, or attentional. That processes are instantiated in the brain is axiomatic, but it remains to future research to make the specific links. Such research would augment the "neuroergonomic" movement in human factors, which seeks to tie operational tasks to the neurology underlying human performance (Funke, 2011; Parasuraman, 2011; Parasuraman and Rizzo, 2007).

In the following sections, we provide a practical guide to the use of resource measures. These are typically termed *mental workload* measures because they are intended to quantify the perceptual and cognitive resources involved in tasks. In addition, we describe research that has used resource-based measures, either alone or in conjunction with other workload measures. Emphasis is on the MRQ and the Workload Profile (WP). The discussion centers on administration, the reduction and analysis of data, reliability and validity, diagnosticity, limitations, and applications. The MRQ and WP approaches are contrasted, and the possibility is considered that combining approaches will provide the fullest picture of task workload. We start with the measurement of resources.

Multiple Resources Questionnaire Approach to the Measurement of Resources

The MRQ consists of 17 rating items, which are typically completed at the conclusion of a task by its performer. Of the items, 12 can be considered fundamentally perceptual in nature. These are listed in Table 4.1, along with the other five items and the definitions of each.

All but three of these items emerged from factor analytic studies of asymmetries in task performance (Boles, 1996) and are thus believed to constitute a set of independent processes. The remaining three (the manual, short-term memory, and vocal processes) were added to the questionnaire because of other evidence that they represent resources that contribute to task performance (Adams and Biers, 2000; Fracker and Wickens, 1989; Wickens, Sandry, and Vidulich, 1983). It has been stressed that under the guiding view that "orthogonal processes represent orthogonal resources," the set of items in the MRQ is not necessarily complete, and that valid multiple resource considerations should be allowed to guide changes to the instrument as required by a particular application. These can include addition and deletion of items (Boles et al., 2007).

Boles et al. (2007) specifically indicated that in actual use, the MRQ is intended to be filled out by the performers of a task rather than workload experts. Originally the rating was done on a 0 to 4 scale (Boles and Adair,

Table 4.1. *The 17 items of the Multiple Resources Questionnaire (MRQ), with their definitions; perceptual and nonperceptual groupings are for expository purposes only and do not appear in the MRQ*

Perceptual Items

Auditory emotional process – Required judgments of emotion (e.g., tone of voice or musical mood) presented through the sense of hearing.
Auditory linguistic process – Required recognition of words, syllables, or other verbal parts of speech presented through the sense of hearing.
Facial figural process – Required recognition of faces, or of the emotions shown on faces, presented through the sense of vision.
Spatial categorical process – Required judgment of simple left-versus-right or up-versus-down relationships, without consideration of precise location, using the sense of vision.
Spatial concentrative process – Required judgment of how tightly spaced are numerous visual objects or forms.
Spatial emergent process – Required "picking out" of a form or object from a highly cluttered or confusing background, using the sense of vision.
Spatial positional process – Required recognition of a precise location as differing from other locations, using the sense of vision.
Spatial quantitative process – Required judgment of numerical quantity based on a nonverbal, nondigital representation (for example, bar graphs or small clusters of items), using the sense of vision.
Tactile figural process – Required recognition or judgment of shapes (figures), using the sense of touch.
Visual lexical process – Required recognition of words, letters, or digits, using the sense of vision.
Visual phonetic process – Required detailed analysis of the sound of words, letters, or digits, presented using the sense of vision.
Visual temporal process – Required judgment of time intervals, or of the timing of events, using the sense of vision.

Nonperceptual Items
Facial motive process – Required movement of your own face muscles, unconnected to speech or the expression of emotion.
Manual process – Required movement of the arms, hands, and/or fingers.
Short-term memory process – Required remembering of information for a period of time ranging from a couple of seconds to half a minute.
Spatial attentive process – Required focusing of attention on a location, using the sense of vision.
Vocal process – Required use of your voice.

2001a), but Finomore et al. (2006) demonstrated that the instrument became more sensitive to a mental workload manipulation when a 0 to 100 scale was used. As a result, recent uses of the MRQ have tended to use the expanded scale (Dillard et al., 2011; Finomore et al., 2008; Finomore, Shaw, Warm, Matthews, and Boles, 2013; Klein et al., 2012). The questionnaire is available in Boles et al. (2007), and the 0 to 100 scale appears in Boles (2010).

Having described the MRQ, we next turn to the issue of item reduction.

Item Reduction

There is a general consensus that even if all items on the MRQ are administered as a measure of workload, their number should be reduced when the instrument is actually scored. The aim of the reduction is to remove items that are irrelevant to the situation being assessed, reducing error variance as well as distortions in the average overall workload. Reducing the number of items also yields statistical power benefits by limiting the number of levels in the MRQ variable used in subsequent inferential analyses. This makes it more likely that a given effect will survive tests of statistical significance that correct for the number of levels of a variable (e.g., the Bonferroni procedure; Maxwell and Delaney, 1990, pp. 177–181). In fact it appears that including all 17 levels is so punishing, statistically speaking, that it can result in inconclusive MRQ outcomes (Humphrey and Adams, 2010; Klein et al., 2009).

Two reduction procedures that have been used are to retain only those items that are (a) rated above zero workload by at least 50 percent of raters (Finomore et al., 2006; Finomore et al., 2013) or (b) rated significantly greater than zero across raters, according to Bonferroni-corrected *t*-tests (Klein et al., 2012). At present it is not clear which should be the preferred approach, but both have proved useful. In our experience they frequently yield identical results. In either case, overall workload is then typically measured as the average of ratings across the retained items (Finomore et al., 2006; Finomore et al., 2013; Klein et al., 2012).

Compound Measures and Metrics for Task Comparison and the Assessment of Interference

Finomore et al. (2008) contributed suggestions for how to compare different tasks in workload when the tasks draw on differing sets of resources. Their approach was to use a joint resource profile that counted only those resources rated “at least 50 percent” in both the visual and auditory vigilance tasks. This profile was found to be sensitive to event rate, both alone and in interaction with sensory modality. Finomore et al. (2009, 2013) used the same approach when examining common resources used in simultaneous versus successive and single- versus dual-task conditions of a vigilance task. An alternative approach, used successfully by Dillard (2012), is to identify critical resources separately for each condition in the study, with all of the ones so identified then included in comparisons between conditions.

It may be noted that while assessment of the overall workload of tasks appears to be the most popular current use of the MRQ, it was actually designed to be a predictor of dual-task interference. Following several empirical comparisons, Boles et al. (2007) concluded that the best dual-task metric is overlap similarity. This is a measure obtained by examining each item in a rater's results, taking the minimum rating of the tasks in the dual-task pairing, and then summing all of the minimum ratings across items (for an example, see Dillard and Boles, 2009).

Validity

The ability of the MRQ to predict dual-task interference provides a key measure of its criterion validity. In the Boles et al. (2007) study, the correlation of the overlap similarity measure to interference between different computer games played simultaneously was found to be 0.83, a substantial result. Construct validity has also been assessed in comparison with the NASA-TLX, a generally accepted and frequently used workload measure that does not stress separate mental resources but that does include a Mental Demand subscale (Hart, 2006). The correlation between workload levels assessed by the MRQ and the full-scale NASA-TLX is most typically reported to be on the order of 0.3–0.4 (Finomore et al., 2006, 2008; Horner et al., 2011). Assuming that commonalities between the MRQ and the NASA-TLX are limited to the Mental Demand subscale, which is only one of six subscales on the NASA-TLX (Hart, 2006), a correlation of this magnitude indicates reasonable construct validity. However, such correlations are not always found (Fincannon et al., 2009) or are found for the workload of some team members in a work setting but not others (Fincannon et al., 2010).

Reliability

The interrater reliability of the MRQ has been found to range from 0.57 to 0.83 when used to assess a variety of computer-based video games as well as simple laboratory-based tasks (Boles and Adair, 2001a). However, this measure concerns only agreement among raters. As a result the estimates are misleadingly low because the task, not the individual rater, is the intended target of the instrument. A more realistic application would be to have small sets of raters rate the same tasks, with a view toward establishing the workload involved in each task. Boles and Adair (2001a) reported that when ratings are aggregated over eight or more raters, reliability is approximately 0.9. In other words, the averaged ratings of a set of eight or more raters correlates at this level with those of other sets of eight or more raters.

Boles and Adair also reported informal observations that participants' consistency in rating the items of the MRQ increases when administration is done orally rather than in written form, by reading the instructions aloud.

This was attributed to communicating a better understanding of the conjoint nature of most of the items, for example, that a task must involve processing that is both verbal *and* lexical in order for the visual lexical process to receive a high rating. On the response end, Carswell et al. (2010) formally tested the impact of vocal versus written response modalities in rating laparoscopic training tasks. When undergraduates performed the tasks and completed the MRQ while reading the items, sensitivity to task demands was higher for written than for vocal responses. However, medical students responding vocally were much more sensitive to task demands than undergraduates using either response modality, and for the medical students the MRQ was much more sensitive than a version of the NASA-TLX.

Diagnosticity

A major strength of the MRQ is its diagnosticity. With 17 resources represented, there is a potential for substantial insight into the structure of tasks and their interactions. Ideally this could be addressed by predicting the resources that will vary in usage in response to task manipulations. If the MRQ identifies the resources that are predicted to be engaged, it could then be considered a predictive instrument.

A tank warfare simulation study by Vogel-Walcutt, Schatz, Bowers, Gebrim, and Sciarini (2008) met these requirements. A high-load condition involving moving tanks was contrasted with a low-load condition using stationary tanks. A specific prediction was made that the former would show higher workload for the short-term memory, spatial attentive, spatial categorical, spatial quantitative, and visual temporal items of the MRQ. This was subsequently confirmed for all but the spatial quantitative item.

However, most assessments of the diagnosticity of the MRQ have relied on face validity. Thus Klein et al. (2005) used the MRQ to measure workload during endoscopic surgery simulation and found ratings that were significantly greater than zero for six items, all of which were judged to be diagnostic of task demands. Finomore et al. (2006, 2013) reported that of a number of resources identified by the MRQ in vigilance tasks, most matched what they might have expected for such a task, suggesting both diagnosticity and content validity. Boles et al. (2007; Experiment 2) reported that three computer-based games differed in resource demand as measured by the MRQ, in several ways that matched game characteristics. For example, two continuous-navigation games placed significantly greater demand on the manual resource than did a task requiring intermittent responses. Conversely, a word-based game placed significantly higher demand on the spatial emergent and visual lexical resources, presumably because it required "picking out" letters from a matrix and forming words.

Finomore et al. (2008) further supported the claim that the MRQ has diagnostic power by showing that in a visual vigilance task, more visual resources were identified by the "at least 50 percent" criterion than were identified in an auditory vigilance task. Manual and short-term memory resources were also identified in both tasks, and both were viewed by the researchers as diagnostic. Finally, Finomore et al. (2009, 2013) pointed out that in a spatial discrimination study requiring judgments of length, as well as in a previous study emphasizing spatial processing (Finomore et al., 2006), it was the spatial items on the MRQ that were dominant in the workload estimates. Yet this was not true in a study involving detection of changes in the duration of auditory and visual pulses (Finomore et al., 2008). They concluded that these differential outcomes speak well for the content validity of the MRQ.

Together these findings suggest that the MRQ is substantially diagnostic of the mental demands imposed by a variety of tasks. However, more studies are needed that are predictive in nature, with firm bases for positing that specific resources will vary between experimental conditions.

Limitations

Although the MRQ has proved useful in a variety of situations, involving both single-task and dual-task performance, the instrument has some limitations. One is that it does not measure sources of interference that lie outside the domain of specific resources. Boles et al. (2007) pointed out that some interference appears to occur whenever two tasks are paired, regardless of their resource structure, supporting previous suggestions that there may be coordination costs in managing dual tasks, or some generalized resource that produces interference regardless of the characteristics of the tasks (Friedman, Polson, Dafoe, and Gaskill, 1982; Wickens, 1984, p. 305).

Also, Finomore et al. (2008) highlighted a content-related limitation of the MRQ, namely, the inclusion of fewer auditory items than visual ones. Although in their study the auditory emotional resource was found to be engaged at high auditory event rates, presumably reflecting the participants' own emotional reaction to closely spaced events, the MRQ really contains no items relating to processing the white noise stimuli used in the study or (for example) pertaining to auditory attention. This is part of a broader limitation deriving from the origin of the MRQ in factor analyses of processes lateralized to the left or right hemispheres of the brain. What processes emerged from those analyses depended, of course, on those that went in. Recognizing this limitation, we have repeatedly suggested that other task-related resources be added to the MRQ if believed appropriate to a particular work environment, especially if motivated by valid multiple resource considerations (Boles and Adair, 2001b; Boles et al., 2007; Boles and Phillips, 2007). A final limitation worth mentioning is that as a questionnaire, the MRQ requires disengagement from a task, followed by reflection upon it. Although the delay can be minimized, it cannot be eliminated. Therefore, as is the case with the NASA-TLX and other nonphysiological measures, the MRQ does not represent an "online" measure of workload.

Applications of the MRQ

To date most applications of the MRQ have been to the measurement of workload in single tasks as opposed to dual tasks. Joel Warm and his associates have applied the measure to vigilance decrements (Dillard et al., 2011; Finomore et al., 2009, 2013). Understanding the factors that influence workload in vigilance performance is a critical human factors concern for system reliability, the safety and well-being of military personnel, and mission success (Nickerson, 1992).

In the first of these studies to use the full 0 to 100 rating scale, observers participated in vigils in which they monitored a clocklike stimulus for the presence of a critical signal, which in the *presence* condition was a vertical line intersecting a circle at the 6 o'clock position, and not circles at the other hourly positions, but in the *absence* condition was an empty circle at the 6 o'clock position but with lines in the circles at all the other positions. The displays were updated 30 times/min, and 10 critical signals occurred per 10-minute period in a 40-minute vigil in each condition. The MRQ was completed after each vigil. The results indicated that 7 of the 17 MRQ items were rated above zero by at least 50 percent of observers. The mean rated workload across the 7 items was 45.35 out of 100 in the presence condition but 63.75 out of 100 in the absence condition, a statistically significant result indicating higher mental workload when searching for feature absence than when searching for feature presence. The results also indicated that feature search requires attention, in that the spatial attentive item was the mental process rated as involving the highest workload (Finomore et al., 2009). Similar results were reported by Finomore et al. (2013), who also noted that the MRQ but not the NASA-TLX was sensitive to the increased processing load of a multitasking condition.

The MRQ has also been used to measure workload in medical settings (Carswell et al., 2010; Horner et al., 2011; Klein et al., 2009, 2012). Some of this research has involved laparoscopic or endoscopic surgery. Klein et al. (2005) used an endoscopic simulation in which the operator could either view the surgical field (a pegboard) through Plexiglas (the control condition) or over a TV monitor, which was either aligned with the normal line of sight or had a view rotated 90 degrees (the two experimental conditions). The task was to transfer foam stars between pegs, for 12 minutes. The results showed statistically significant lower workload in the control condition than in the experimental conditions, which did not differ from one another.

A final domain in which the MRQ has been employed is virtual navigation (Fincannon et al., 2009; Sellers, Fincannon, and Jentsch, 2012). Fincannon et al. (2010) investigated the interaction of workload and spatial ability in three-person teams whose task was to use an overhead map to plan and navigate the route of unmanned aerial vehicles (UAVs). Each team consisted of an intelligence officer, a UAV operator, and an unmanned ground vehicle operator. The intelligence officer planned the routes and communicated them to the operators for execution. The MRQ was completed at mission's end. The results showed intriguing correlations between workload and spatial ability that depended on the team member. For the UAV operators, workload was reduced if the intelligence officer had high spatial ability. However, high spatial ability in the UAV operator was associated with *higher* workload in the unmanned ground vehicle operator. Examination of the MRQ item results appeared to indicate that intelligence officers with high spatial ability reduced the spatial processing and memory demands on the UAV operators. Increased spatial ability in the unmanned ground vehicle operators, on the other hand, increased vocal processing by their aerial counterparts, accounting for the higher workload.

Sellers et al. (2012) found that the MRQ and the NASA-TLX were both sensitive to increased automation of a virtual unmanned ground vehicle that navigated around civilians, with workload decreasing with increased automation. Both measures were sensitive as well to the visualization ability of the operators, revealing lower workload in individuals with higher ability.

Although the MRQ involves multiple resources, it is not the only multiple resource approach to workload measurement. Other approaches have employed the reduced set of resources incorporated in MRT.

MRT Approaches: Expert Judgment and the Workload Profile

An alternative to the MRQ is to use an MRT approach to measure perceptual resources. As already indicated, MRT implicitly recognizes that resources can be perceptual through its inclusion of visual versus auditory and verbal versus spatial dimensions. An early MRT-based approach to measuring resources was investigated by Wickens and colleagues (1988). It used expert judgment to estimate resource demand in eight "channels" employed in flight simulation tasks and in an accompanying secondary task. Two channels involved left and right manual responses, but the remaining six can be viewed as perceptual: window display, electronic map display, print, voice (i.e., voice recognition), verbal processing, and spatial processing. Each of the flight tasks, which varied in difficulty, involved the pilot using a perspective display of terrain to navigate a simulated light aircraft at low altitude for a 15- to 20-minute flight. Each of the secondary tasks, which was performed simultaneously with the flight task, involved mentally manipulating and responding to unrelated signals requiring either mental rotation or numerical computation. A workload expert rated the amount of conflict expected for each task pairing in each of the eight channels. These were then summed using a mathematical model that also included overall demand within each channel irrespective of conflict. Performance was subsequently tested under actual dual-task conditions, and across task pairings was found to correlate at nearly a 0.50 level with the predictions of the model (Wickens et al., 1988).

Unfortunately this promising beginning received little if any follow-up, perhaps because of obvious difficulties in applying the methodology. Probably the most vexing is the use of an expert to predict conflict between the channels, raising as it does questions as to what constitutes a multiple resource expert, and whether the expert's judgments show both intra- and interrater reliability. In addition, one can also question whether an expert is in a position to judge the demand on users who show individual differences in ability and experience. Also, aggregating conflict ratings with overall demand obscures the contribution of each component, although their effects could presumably be extracted with further analysis of the data.

These same difficulties pertain to a similar MRT-based approach called the Workload Index (W/INDEX; North and Riley, 1989). It too used expert judgment, in this case to estimate the additive workload across attentional channels, the penalty due to conflicts within channels, and the penalty due to conflicts between channels. All of these were then entered into a summative equation meant to predict performance.

More recently, the MRT was applied in a different way in an approach called the Workload Profile (WP) (Tsang and Velazquez, 1996). Here participants rate a task along eight workload dimensions, five of which can be viewed as at least partially perceptual in nature (perceptual/central stage of processing, spatial code, verbal code, visual input, and auditory input). Thus the WP, like the MRQ, uses the judgment of actual participants rather than workload experts. In an experiment conducted by Tsang and Velazquez, workload was assessed in a Sternberg working memory task, a continuous tracking task, and a dual task involving both. Multiple regressions showed that the WP dimensions collectively accounted for 18–26 percent of single-task variance, rising to 43–71 percent if participant effects (individual differences) were accounted for. The dimensions accounted for about 10–15 percent of variance in the dual task, rising to 30–56 percent when individual differences were accounted for.

Comparison of the MRQ and WP

The MRQ and WP are resource-based and use participant ratings to measure workload. However, they also differ in important respects, most notably in the prevalence of resources defined by conjoint properties such as being both visual *and* lexical (highly prevalent in the MRQ, and not present in the WP), and in the number of posited resources (17 in the MRQ, and 8 in the WP). Empirical comparison of the approaches would be desirable, but to date only that conducted by Phillips and Boles (2004) appears to exist. Using three computer games, participants rated resource use with both the MRQ and WP. The two instruments were found to predict the interference between simultaneously performed pairs of games equally, although the MRQ did this with less variability over participants than the WP. This somewhat inconclusive outcome could perhaps be resolved by repeating the study, selecting dual tasks that differ in dimensions that are differentiated by one but not the other instrument. For example, the MRQ has five spatial resources, and the WP has only one. Also, the WP differentiates between tasks that generically emphasize the central/perceptual stage of processing and those that do not, while the MRQ draws no such distinction. Including such points of difference in a dual-task study might well identify the conditions best suited to each instrument.

Diagnosticity of the WP

The diagnosticity of the WP has been investigated using canonical discriminant analysis (Fréard et al, 2007; Rubio et al., 2004; Tsang and Velazquez, 1996). Essentially, this technique determines which variables capture the greatest amount of variability in a dependent measure or set of measures. The technique organizes them into orthogonal dimensions called *canonical variates*. As applied to diagnosticity, the aim is to determine to what extent the items of a workload scale capture differences between the conditions of a study. For example, using WP items, Tsang and Velazquez (1996) identified canonical variates that distinguished between tracking and memory tasks, between easy and difficult tasks, between first-order and second-order tracking, and between single and dual tasks.

The diagnosticity of WP relative to other workload instruments was examined by Rubio, Díaz, Martín, and Puente (2004). Two of the instruments were non-resource-based measures of workload, namely, the NASA-TLX and the Subjective Workload Assessment Technique (SWAT; Reid and Nygren, 1988). They found that canonical discriminant analysis was able to distinguish between single and dual tasks using NASA-TLX data. However, using the WP and SWAT data, such analysis distinguished not only between single and dual tasks, but also between memory and tracking tasks. These results suggest greater diagnosticity for the WP and SWAT than for the NASA-TLX. Using a similar statistical approach, Fréard et al. (2007) also reported that the WP was more diagnostic than the NASA-TLX, especially in distinguishing purely visual from auditory and mixed-modality tasks.

Rubio et al. (2004) compared the WP with two nonresource workload measures, the NASA-TLX and the SWAT. Participants performed a Sternberg memory search task, a tracking task, and a dual task combining both. As global measures of workload, all three intercorrelated above 0.90, indicating high convergent validity. In addition all showed substantial correlations with timed performance over the various conditions. However, the NASA-TLX outperformed both the WP and SWAT in predicting tracking error performance.

Limitations

Finally, we note that the limitations of the MRQ also apply to the WP, as well as the NASA-TLX and SWAT. None is an "online" measure, and all are limited in the number of included resources. It is also doubtful that any can claim to be absolute measures of workload, in the sense that a given quantity on a measured scale can be deemed "excessive" (Wickens and Hollands, 2000, p. 457).

Having considered the MRQ and MRT approaches, we next take up how approaches might be combined to provide a fuller picture of workload.

Combined Approaches

In contrast to examining how one workload instrument measures up to another, we can also ask whether there is any advantage to combining existing instruments or even pieces of them. Potentially this could paint a richer view of workload. Of relevance are studies that have used at least one resource-based measure as part of the test battery. However, few have analyzed and presented the data in terms of the value added by combining workload measurement instruments.

Phillips and Boles (2004) used two resource-based workload measures in the same study, namely, the MRQ and WP. As noted, the instruments equally predicted interference between games, although the MRQ showed lower variability across subjects. Item-by-item comparison of ratings on the MRQ and WP show substantial similarity. However, the WP appears to have been sensitive to auditory signals in one of the games while the MRQ was not (although the statistical significance of the result is unclear). The general lack of auditory items in the MRQ was noted previously as a limitation, so these results suggest that the WP could be combined with the MRQ to provide additional auditory resource sensitivity.

A second indication that there may be a benefit to combined measures is from the vigilance study of Finomore et al. (2008), which combined one resource-based measure (the MRQ) with one nonresource measure (the NASA-TLX). It was reported that while a number of the visual resources of the MRQ correlated with NASA-TLX subscales, the significant correlations were most typically to the Mental Demand subscale. The implication is that the two instruments measure partially overlapping aspects of workload, with the MRQ providing more diagnosticity than the Mental Demand subscale, but with other subscales of the NASA-TLX measuring unique workload variance. The results suggested particularly large unique variance contributions through the Performance and Frustration subscales.

Finally, Dillard et al. (2011) found that the MRQ complemented previous findings using the NASA-TLX in challenging the mindlessness model of vigilance, which states that vigilance decrements in performance are due to attentional lapses (Manly, Robertson, Galloway, and Hawkins, 1999; Robertson et al., 1997). Instead of responding to occasional critical signals as in a traditional vigilance format, with the sustained attention to response (SART) task observers continuously respond to the more frequent neutral events but withhold responding to rare critical signals.

Dillard et al. tested the claim that the SART task is an engine for the promotion of mindlessness in a study requiring observers to monitor a simulated air traffic control display divided into quadrants. Within each quadrant was an icon representing a jet flying a circular path around the center of the display. For critical signals, which occurred 48 times in a 40-minute vigil, one jet was headed the direction opposite the others, indicating that it was on a collision course. In different conditions of the study, observers either responded to neutral events and withheld response to the critical signals or responded to the critical signals and withheld response to the neutral events. Contrary to the prediction of the mindlessness model, workload as assessed by the MRQ was the same in the two active conditions, and was at a substantial level, as opposed to significantly lower workload ratings in the control condition, which involved watching the display with no work imperative. Eye tracking measures supported this conclusion, with the active conditions demonstrating more visually active search of the display in comparison to the control condition. Both outcomes were in agreement with previous research using the NASA-TLX. Such findings carry operational implications by suggesting that work systems should be designed to reduce operator workload instead of introducing additional cues to gather attention, because operators in vigilance tasks are highly taxed and not mindless. Thus both the conclusions and their operational implications were strengthened through a combination of resource-based (MRQ) and non-resource-based (eye movement and NASA-TLX) approaches.

Taken together, these findings suggest that combining resource-based instruments such as the MRQ and WP or combining resource-based with nonresource workload instruments can give a fuller picture of workload than is provided by a single instrument. However, research on this topic is still in its infancy. Certainly combining measures can have little added cost; for example, the MRQ and a short version of the NASA-TLX, skipping a scaling procedure, each takes less than five minutes to complete (Horner et al., 2011). In contrast, the scaled version of the NASA-TLX and the WP both take longer, perhaps on the order of 15–30 minutes each (Rubio et al., 2004; Tsang and Velazquez, 1996).

Conclusions

Under any model of multiple resources, perceptual resources are proposed to constitute a major influence on performance. Visual and auditory resources, perceptual/central processing resources, and conjoint modality and process resources have all been proposed and found to be both measurable and predictive of task performance (Boles and Adair, 2001a, 2001b; Boles et al., 2007; Rubio et al., 2004; Tsang and Velazquez, 1996; Wickens et al., 1988). The role of attention has generally been relegated to the background. Our own approach has effectively abandoned attention as an overarching framework. It is hoped that future research will conclusively test the possibility, raised here, that resources are mental entities intimately linked to a full range of processes including ones that can be characterized as perceptual, memorial, motoric, or attentional.

Resources have been measured by expert judgment (Wickens et al., 1988) and by questionnaire, but it is questionnaire approaches that have recently dominated. The MRQ consists of 17 resource items, 12 of which are perceptual in nature (Boles and Adair, 2001a, 2001b; Boles et al., 2007). As an instrument it typically shows moderate correlations with the non-resource-based NASA-TLX, especially to the Mental Demand subscale (Finomore et al., 2008; Horner et al., 2011). The MRQ has been found an excellent predictor of interference between simultaneously played computer games, an indication both of its construct validity and of its ability to measure perceptual resources (Boles et al., 2007). Another strength of the questionnaire is its diagnosticity with respect to the resources demanded within tasks (Boles et al., 2007; Finomore et al., 2006, 2008, 2009, 2013; Klein et al., 2005; Vogel-Walcutt et al., 2008).

The WP is an eight-item questionnaire, five of which are at least partially perceptual in nature. It has been found to account for a significant proportion of single-task variance in performance (Tsang and Velazquez, 1996). There are suggestive indications that it may be more generally sensitive to auditory resources than the MRQ (Phillips and Boles, 2004). Its value is also supported by a demonstration of its discriminant validity, showing that it can differentiate single from dual tasks, memory from tracking tasks, and visual from auditory and mixed-modality tasks (Fréard et al., 2007; Rubio et al., 2004; Tsang and Velazquez, 1996).

We end with the suggestion that further research be directed not only toward contrasting the relative strengths of resource-based instruments of workload, but toward combining approaches. Different resource-based as well as non-resource-based approaches may be sensitive to somewhat varying sources of performance variability. Applied research can only benefit from pooling these sources.

Acknowledgment

This chapter was written while Michael Dillard held a National Research Council Research Associateship Award at the Air Force Research Laboratory/711th Human Performance Wing.

References

Adams, C. M. and Biers, D. W. (2000). ModSWAT and TLX: Effect of paired comparison weighting and weighting context. In Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society. Santa Monica, CA: Human Factors and Ergonomics Society, pp. 6.610–6.613

Boles, D. B. (1991). Factor analysis and the cerebral hemispheres: Pilot study and parietal functions. *Neuropsychologia*, 29, 59–91.

Boles, D. B. (1992). Factor analysis and the cerebral hemispheres: Temporal, occipital and frontal functions. *Neuropsychologia*, 30, 963–988.

Boles, D. B. (1996). Factor analysis and the cerebral hemispheres: "Unlocalized" functions. *Neuropsychologia*, 34, 723–736.

Boles, D. B. (2010). Multiple Resources Questionnaire. Downloaded from http://bama.ua.edu/~dboles/MRQ.expanded.pdf.

Boles, D. B. and Adair, L. P. (2001a). The Multiple Resources Questionnaire (MRQ). *Proceedings of the Human Factors and Ergonomics Society*, 45, 1790–1794.

Boles, D. B. and Adair, L. P. (2001b). Validity of the Multiple Resources Questionnaire (MRQ). *Proceedings of the Human Factors and Ergonomics Society*, 45, 1795–1799.

Boles, D. B., Bursk, J. H., Phillips, J. B. and Perdelwitz, J. R. (2007). Predicting dual-task performance with The Multiple Resources Questionnaire (MRQ). *Human Factors*, 49, 32–45.

Boles, D. B. and Law, M. B. (1998). A simultaneous task comparison of differentiated and undifferentiated hemispheric resource theories. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 204–215.

Boles, D. B. and Phillips, D. B. (2007). A reply to the methodological and theoretical concerns of Vidulich and Tsang. *Human Factors*, 49, 50–52.

Boles, D. B., Phillips, J. B., Perdelwitz, J. R. and Bursk, J. H. (2004). Application of the Multiple Resources Questionnaire (MRQ) to a complex gaming environment. *Proceedings of the Human Factors and Ergonomics Society*, 48, 1968–1972.

Brooks, L. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 22, 349–368.

Carswell, C., Lio, C. H., Grant, R., Klein, M. I., Clarke, D., Seales, W. and Strup, S. (2010). Hands-free administration of subjective workload scales: Acceptability in a surgical training environment. *Applied Ergonomics*, 42, 138–145.

Dillard, M. B. and Boles, D. B. (2009). Interference between visuospatial dual tasks and the effects of training. *Proceedings of the Human Factors and Ergonomics Society*, 53, 1888–1892.

Dillard, M., Boles, D.B., Funke, M.E., Funke, G.J., Finomore, V., Dukes, A. et al. (2011). The sustained attention to response task does not promote mindlessness in vigilance. Proceedings of the International Symposium on Aviation Psychology 16, 309–314.

Dillard, M. B. (2012). *Transfer of training and the role of mental workload in expert performance*. (Doctoral dissertation, University of Alabama).

Dux, P. E., Ivanoff, J. G., Asplund, C. L. and Marois, R. (2006). Isolation of a central bottleneck of information processing with time-resolved fMRI. *Neuron*, 52, 1109–1120.

Fincannon, T. A., Evans, W., Phillips, E., Jentsch, F. and Keebler, J. (2009). The influence of team size and communication modality on team effectiveness with unmanned systems. *Proceedings of the Human Factors and Ergonomics Society*, 53, 419–423.

Fincannon, T. D., Ososky, S., Jentsch, F., Keebler, J. and Phillips, E. (2010). Some good and bad with spatial ability in three person teams that operate multiple unmanned vehicles. *Proceedings of the Human Factors and Ergonomics Society*, 54, 1615–1619.

Finomore, V. S., Shaw, T. H., Warm, J. S., Matthews, G., Riley, M. A., Boles, D. B. and Weldon, D. (2008). Measuring the workload of sustained attention: Further evaluation of the Multiple Resources Questionnaire. *Proceedings of the Human Factors and Ergonomics Society*, 52, 1209–1213.

Finomore, V.S., Shaw, T.H., Warm, J.S., Matthews, G. and Boles, D.B. (2013). Viewing the workload of vigilance through the lenses of the NASA-TLX and the MRQ. *Human Factors*, 55, 1044–1063.

Finomore, V. S., Shaw, T. H., Warm, J. S., Matthews, G., Weldon, D. and Boles, D. B. (2009). On the workload of vigilance: Comparison of the NASA-TLX and the MRQ. *Proceedings of the Human Factors and Ergonomics Society*, 53, 1057–1061.

Finomore, V. S., Warm, J. S., Matthews, G., Riley, M. A., Dember, W. N., Shaw, T. H. and Scerbo, M. W. (2006). Measuring the workload of sustained attention. *Proceedings of the Human Factors and Ergonomics Society*, 50, 1614–1618.

Fracker, M. L. and Wickens, C. D. (1989). Resources, confusions, and compatibility in dual-axis tracking: Displays, controls, and dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 80–96.

Fréard, D., Jamet, E., Le Bohec, O., Poulain, G. and Botherel, V. (2007). Subjective measurement of workload related to a multimodal interaction task: NASA-TLX vs. Workload Profile. In J. Jacko (Ed.), *Human-computer interaction: HCI intelligent multimodal interaction environments: Lecture notes in computer science* (Vol. 4552, pp. 60–69). Berlin and Heidelberg: Springer-Verlag Berlin Heidelberg.

Friedman, A., Polson, M. C., Dafoe, C. G. and Gaskill, S. J. (1982). Dividing attention within and between hemispheres: Testing a multiple resources approach to limited-capacity information processing. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 625–650.

Funke, M. E. (2011). Neuroergonomic and stress dynamics associated with spatial uncertainty during vigilance task performance. Doctoral dissertation, University of Cincinnati. Downloaded from http://etd.ohiolink.edu/view.cgi?acc_num=ucin1313773072.

Gazes, Y., Rakitin, B. C., Steffener, J., Habeck, C., Butterfield, B., Ghez, C. and Stern, Y. (2010). Performance degradation and altered cerebral activation during dual performance: Evidence for a bottom-up attentional system. *Behavioural Brain Research*, 210, 229–239.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX): 20 years later. *Proceedings of the Human Factors and Ergonomics Society*, 50, 904–908.

Horner, R. D., Szaflarski, J. P., Jacobson, J., Elder, N., Bolon, S., Matthews, G. et al. (2011). Clinical work intensity among physician specialties: How might we assess it? *Medical Care*, 49, 108–113.

Humphrey, C. M. and Adams, J. A. (2010). General visualization abstraction algorithm for directable interfaces: Component performance and learning effects. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 40, 1156–1167.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Klee, H. and Eysenck, M. W. (1973). Comprehension of abstract and concrete sentences. *Journal of Verbal Learning and Verbal Behavior*, 12, 522–529.

Klein, M. I., Lio, C. H., Grant, R., Carswell, C. M. and Strup, S. (2009). A mental workload study on the 2d and 3d viewing conditions of the da Vinci surgical robot. *Proceedings of the Human Factors and Ergonomics Society*, 53, 1186–1190.

Klein, M. I., Riley, M. A., Warm, J. S. and Matthews, G. (2005). Perceived mental workload in an endoscopic surgery simulator. *Proceedings of the Human Factors and Ergonomics Society*, 49, 1014–1018.

Klein, M. I., Warm, J. S., Riley, M. A., Matthews, G., Doarn, C., Donovan, J. F. and Gaitonde, K. (2012). Mental workload and stress perceived by novice operators in the laparoscopic and robotic minimally invasive surgical interfaces. *Journal of Endourology*, 26, 1089–1094.

Manly, T., Robertson, I. H., Galloway, M. and Hawkins, K. (1999). The absent mind: Further investigations of sustained attention to response. *Neuropsychologia*, 37, 661–670.

Maxwell, E. E. and Delaney, H. D. (1990). *Designing experiments and analyzing data: A model comparison perspective*. Belmont, CA: Wadsworth.

Mochizuki, H., Tashiro, M., Gyoba, J., Suzuki, M., Okamura, N., Itoh, M. and Yanai, K. (2007). Brain activity associated with dual-task management differs depending on the combinations of response modalities. *Brain Research*, 1172, 82–92.

Nickerson, R. S. (1992). *Looking ahead: Human factors challenges in a changing world*. Mahwah, NJ: Erlbaum.

North, R. A. and Riley, V. A. (1989). W/INDEX: A predictive model of operator workload. In G. R. McMillan, D. Beevis, E. Salas, M. H. Strub, R.

Sutton and L. Van Breda (Eds.), *Applications of human performance models to system design, Defense Research Series* (Vol. 2, pp. 81–89). New York: Plenum.

Parasuraman, R. (2011). Neuroergonomics: Brain, cognition, and performance at work. *Current Directions in Psychological Science*, 20, 181–186.

Parasuraman, R. and Rizzo, M. (2007). Introduction to neuroergonomics. In R. Parasuraman and M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 3–12). New York: Oxford University Press.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116, 220–244.

Phillips, J. B. and Boles, D. B. (2004). Multiple Resources Questionnaire and Workload Profile: Application of competing models to subjective workload measurement. *Proceedings of the Human Factors and Ergonomics Society*, 48, 1963–1967.

Reid, G. B. and Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam: North-Holland.

Robertson, I. H., Manly, T. Andrade, J., Baddeley, B. T. and Yiend, J. (1997). "Oops!": Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, 35, 747–758.

Rubio, S., Díaz, E., Martín, J. and Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: An International Review*, 53, 61–86.

Rugg, M. D. (1986). Constraints on cognitive performance: Some problems with and alternatives to resource theory. In G. R. J. Hockey, A. W. K. Gaillard and M. G. H. Coles (Eds.), *Energetics and human information processing* (pp. 353–371). Dordrecht, The Netherlands: Martinus Nijhoff.

Sellers, B.C., Fincannon, T. and Jentsch, F. (2013). The effects of autonomy and cognitive abilities on workload and supervisory control of unmanned systems. *Proceedings of the Human Factors and Ergonomics Society*, 56, 1039–1043.

Sigman, M. and Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during dual-task performance. *The Journal of Neuroscience*, 28, 7585–7598.

Tsang, P. S. and Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39, 358–381.

Vogel-Walcutt, J.J., Schatz, S., Bowers, C., Gebrim, J.B. and Sciarini, L.W. (2008). Augmented cognition and training in the laboratory: DVTE system validation. *Proceedings of the Human Factors and Ergonomics Society*, 52, 187–191.

Wickens, C. D. (1984). *Engineering psychology and human performance* (1st ed.). Columbus, OH: Charles E. Merrill.

Wickens, C. D., Harwood, K., Segal, L., Tkalcevic, I. and Sherman, B. (1988). TASKILLAN: A simulation to predict the validity of multiple resource models of aviation workload. *Proceedings of the Human Factors Society*, 32, 168–172.

Wickens, C. D. and Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.), Upper Saddle River, NJ: Prentice Hall.

Wickens, C. D. and Sandry, D. L. (1982). Task-hemispheric integrity in dual task performance. *Acta Psychologica*, 52, 227–248.

Wickens, C. D., Sandry, D. L. and Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25, 227–248.

5 Methods and Applications of Eye Tracking

F. Jacob Seagull

Introduction

Knowing where people are looking can provide insight into their thought processes and shed light on the workings of their mind (Pillsbury, 1908). The appeal of eye tracking is intuitive – the ability to track a person's gaze opens a world of possibilities. Indeed, today's eye-tracking technology makes a range of research questions possible and feasible. The technology is impressive and can seem almost magical.

Early eye-tracking devices were expensive, and some of the earliest were invasive, requiring users to wear special contact lenses. Noninvasive measures can be as simple as filming a person doing a task and manually coding his or her gaze locations on the basis of estimating the person's direction of gaze. In the early twentieth century Frank Gilbreth used such a technique to study typing. Four decades later, gaze patterns of pilots were tracked by the same means (Fitts, Jones, and Milton, 1950). Over the years, eye tracking has been used in a number of domains, including psychology, biology, motor learning, marketing, and others. For more extensive review of the history of eye tracking, please see Richardson and Spivey (2004).

In the past decade, with increases in computing power and marked improvements in camera miniaturization, eye tracking is becoming easier and more affordable. Software for analyzing data from eye-tracking sessions is getting more powerful as well as easier to use. Numerous vendors provide off-the-shelf packages that provide hardware and software that promise complete eye-tracking solutions.

This chapter focuses on practical considerations, and less on background or eye-movement theory. It will focus on the domains of human factors and applied human performance and the practical considerations aimed at assisting the inexperienced eye-tracking researcher to understand the relevant considerations in the design and conduct of eye-tracking research. To that end, the purpose of this chapter is to provide an overview of eye-tracking capabilities, including strengths and weaknesses of different methods, so that the reader can begin to carry out eye-tracking research in a cogent manner. Having clear understanding of capabilities will allow realistic establishment of research plans. Forming a research question, selecting hardware that is appropriate for the setting, and planning data analysis properly will facilitate easy data analysis and interpretation, paving the way to a successful applied research endeavor.

Types of Gaze Tracking Studies

The breadth of meaning derived from analysis of gaze behavior is bounded only by the resourcefulness and ingenuity of investigators. That being said, some questions can be asked and answered in a straightforward manner using gaze tracking. While there are variations in the specific instantiations of these master patterns of gaze studies, a majority of well-executed eye-tracking studies in the field of applied human factors fall into one of these few categories. The division between applied human factors and basic cognitive science is not sharp, as other chapters in this *Handbook* have pointed out. However, studies that are more purely cognitive science will not be covered in this chapter. Many seminal works regarding reading, cognitive developmental, and basic visual processes that merit study are outside the scope of this chapter.

Proof of Concept

As with many novel technologies, eye tracking is often applied to new situations in which a gaze-tracking system may provide a means of accomplishing a goal that would be impossible by conventional means, or that improves an existing way of accomplishing the goal. Such studies abound, such as using eye tracking as a "hands-free" input method to computer systems or an assistive technology (i.e., an "eye mouse"). Trackers can facilitate supervision of activity that is crucially dependent on vision, such as allowing trainee surgeons to monitor what the attending surgeon is looking at during surgery (Chetwood et al., 2012) or exploring ways to integrate eye tracking into surgical training environments (Atkins et al., 2012). Such proof-of-concept pieces tend to describe applications of eye tracking or technological innovations as opposed to eye-tracking research. This type of investigation may or may not include quantitative analysis of gaze data.

Descriptive Research

This most basic type of research simply presents descriptions of visual scanning. What do people look at when they drive, perform surgery, read, fly a plane, walk through an unfamiliar environment? While these types of exploratory studies may provide some insight, the studies are more valuable when there is refined hypothesis testing. These types of studies can provide insight into the links observed between the activity or object in the environment and the response of the gaze-tracked user.

Determining or Differentiating Expertise

As expertise develops, so do meaningful patterns of behavior (see Risko and Kingstone, and Hagen and Tanaka, this volume). As schemas develop through practice and study of a task, so too do patterns of eye movement. Some of the earliest studies of gaze showed that there were measurable differences in visual scanning between experts and novices. Experienced pilots exhibit lower variability and more consistent patterns of scanning of cockpit instruments than do less experienced pilots or trainees (Kramer et al., 1994; Kasarskis et al., 2001). Expert neurologists differ from less experienced neurologists (Matsumoto et al., 2011). In scanning radiology images of the brain, the scanning patterns of experienced neurologists were correlated with clinically relevant, nonsalient areas of the brain images, while the scanning by other medical practitioners (e.g., medical students, nurses) was highly correlated only with the visual salience of cues (Matsumoto et al., 2011). Better chess players exhibit different scanning than less experienced players, who show a smaller visual span (ability to see more than one square in a single fixation) and more revisiting (rescanning) squares more often than experts (Blignaut, Beelders, and So, 2008). New drivers have longer fixations and greater variance than experienced drivers (Chapman, 1998).

This type of study can be used to provide insight into strategies as a function of experience or proficiency level. Beyond proficiency level differences, strategies can differ between experts as a function of subspecialization. For example, surgeons, anesthetists, and nurses exhibit different scanning strategies while reviewing video of the same surgery (Xiao et al., 1999). Authors often cite knowledge of these discovered strategies as potentially useful for training novices to increase their proficiency. Many seem to use eye tracking to understand what experts do and train people to use the implied cognitive strategies, such as through Recognition Primed Decision training (Klein, 1998), or perceptual training of tennis players (Williams, Ward, Smeeton, and Allen, 2004). Few have actually tried to train people to follow specific scan patterns, such as Shapiro and Raymond (1989), who trained people to play video games using efficient scan strategies or inefficient strategies and showed a transient advantage to training with efficient scan strategies.

A second use for the study of eye movements is to employ measures of eye-tracking patterns as a diagnostic tool to evaluate people in terms of their proficiency level. Because the differences in scanning behavior are generally robust and consistent, it is possible to pass the gaze data through a statistical analysis to determine the degree to which they resemble an expert's performance. For example, surgeons' scan patterns while performing minimally invasive surgery (on a simulator) were shown to be a reliable method for classifying experts and novices (Ahmidi, Ishii, Fichtinger, Gallia, and Hager, 2012), and eye tracking improves the ability to discriminate between experts and trainee pilots in evaluation flights (Kramer et al., 1994). One should note that before initiating eye-tracking research, it may be worth considering that a stopwatch is a simple device that often can reliably distinguish between experts and novices without complicated statistical models – experts generally perform visual scanning tasks faster than novices. When time is not a reliable measure, there often are other measures of competence that can be used. Eye tracking does, however, provide an objective measure of performance when other measures are not available.

Comparing (Task/Artifact) Designs

Hypothesis testing is a cornerstone of good science. In applied sciences, hypothesis testing often involves the comparison of two alternative artifacts, or two strategic approaches to solving a problem or completing a task. Eye tracking can be used effectively as a measure that can differentiate between two alternatives. For example, eye tracking can suggest which drug-label design leads to more efficient scanning of pertinent information and is less likely to lead to misreading a drug label (Bojko, 2006). Tracking data can show the effects of a traffic avoidance display on airplane pilot scanning of the cockpit instruments. Eye tracking is often used in marketing studies to examine the effects of product designs, such as to compare the effectiveness of Web-page designs, or print advertisements, or product labels in shopping/commerce environments (Bojko, 2013). Nearly any visual change in an environment can be investigated through eye-gaze analysis. Even nonvisual interventions can be studied, such as the effect of noise or distraction on scanning patterns. Eye tracking has potential for use as a dependent measure in any study of human behavior. One should be cognizant of the limits in interpretability of eye-tracking measures, however. Distilling complex scanning behaviors down to simple metrics can constitute a formidable challenge.

This taxonomy of typical studies is just a sample of the breadth of issues that eye-tracking research is commonly used to address (for reviews, see Duchowski, 2009; Richardson and Spivey, 2004). Familiarity with these types of studies can help you refine and improve the design of your own eye-tracking research.

Technology Overview

If you have a question that you will use eye-tracking data to answer, you must have eye-tracking hardware/technology at your disposal. Perhaps the first, most fundamental consideration regarding hardware is whether the eye-tracking optics will be connected physically to the person being tracked or will be mounted remotely, away from the person. Within these two categories of eye trackers there are many variations.

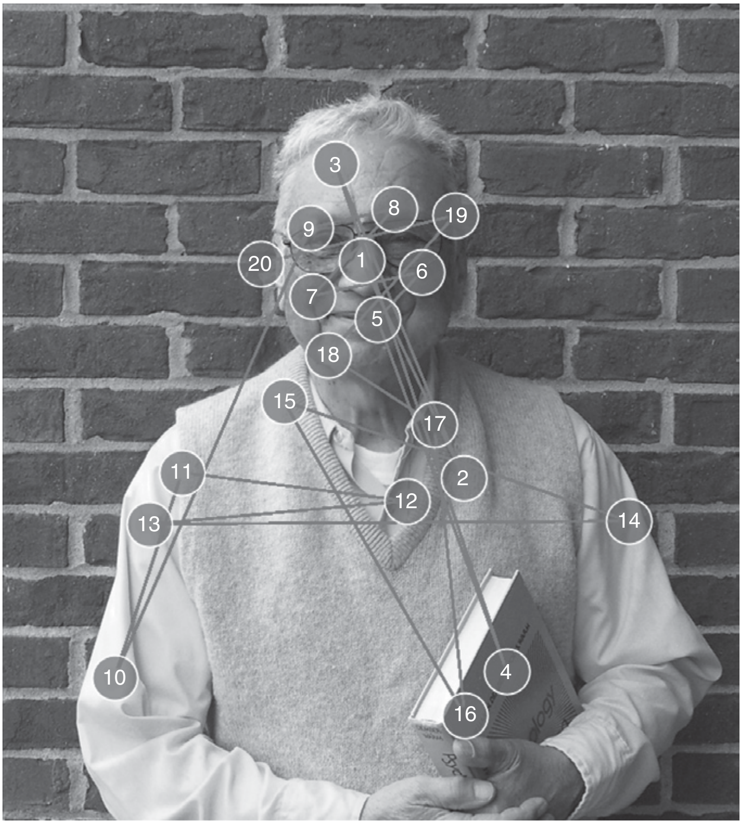


Figure 5.1. *An example of an eye movement trace recorded by gaze tracking. (Photo courtesy of E. Warm; eye movement analysis courtesy of Feng-GUI.com.)*

In most systems, there are two sources of information used by the eye tracker, an eye camera and a scene camera. The eye camera captures the location and orientation of the user's eyes. The scene camera provides information about the object of the user's gaze. The scene camera may be a camera or any other source of information about the user's environment used to correlate gaze data to information in the environment, such as screen captures from a computer monitor, for instance. An example of a combined gaze and scene scanning pattern is shown in Figure 5.1.

Remote Optics

The most simple eye-tracking configuration is the single-camera remote optics setup. In this configuration, a single "eye camera" is pointed at the user's eye to capture a clear image of the eye, typically from a distance of a few feet (though the distance can vary greatly through the use of telephoto camera lenses). Infrared cameras and infrared illuminators are often used to facilitate ease of image processing. Remote-optic systems can be extremely simple and low cost (at the expense of compromises in accuracy). Technology as simple as a Web camera and open-source software can be used for rudimentary tracking (Li and Parkhurst, 2006).

There are more sophisticated and accurate systems in which the camera is fixed in place and the user's head must remain stationary, often anchored through the use of a "bite-bar" that the user bites on (inconvenient and often uncomfortable, but extremely stable), or a chin rest or forehead support (more convenient, but less effective). The amount of allowable movement of the user's head varies greatly, depending on the technology used. Some require the user to remain motionless, while others are able to track the user's head within a constrained volume of space. Tracking the location of the user's head and eyes can be accomplished purely through image-processing techniques that identify the user's head and eyes in the eye camera's raw video feed. Tracking can also be achieved through image processing combined with mechanical mechanisms to adjust the camera's position so that it is centered on the user's head and eyes. Additionally, optical, magnetic, or other stand-alone tracking mechanisms can be used to ensure that the eye camera captures an adequate view of the user's eye(s).

Remote optics can be integrated into a computer display, which can be convenient for gaze tracking in human-computer interaction studies or other studies in which the task is centered on a computer screen (e.g. reading, Web browsing, computer-based simulations). There are few functional differences between integrated and nonintegrated optics. Integrated optics can simplify the setup of a system, and an integrated system is less conspicuous to the users and may lead to more natural behaviors.

Remote systems can include additional cameras that are used to triangulate the location of the user's head and eyes, to improve accuracy, and to expand the functional area in which tracking can occur.

A more recent trend in remote optics for eye tracking is the use of multiple camera systems arrayed around an environment. Adding cameras can expand the tracking space and allow accurate tracking in a volume of space that is, theoretically, arbitrarily large. Video-game technologies, such as the Kinect, are also contributing to efforts to improve eye tracking. As game controllers continue to provide inexpensive and more robust means of tracking users' biometric information, eye-tracking systems can capitalize on these data as part of the tracking system.

Head-Mounted Optics

The other main type of system is referred to as a "head-mounted" system, in which there is physical contact between the user and the tracking system. Typically, a camera is mounted on headgear (headband, cap, or glasses) and pointed either directly at the user's eye or at a partially reflective surface that points to the eye. Users can see through the partial mirror to the scene, and the eye camera captures a front-on view of the eye. Because of the close proximity and the physical coupling of the camera to the eye, head-mounted systems can provide clearer and more stable images of the eye.

Head-mounted systems often include a scene camera mounted in conjunction with the eye camera to capture images of the environment toward which the user is facing. Eye data and scene data can be combined to indicate the relative location of the user's gaze. Absolute location of the user's gaze can be calculated in such systems through the addition of either a head-tracking system that links the head position to a stabile/absolute frame of reference, or through the use of optical anchors in the environment that can be recognized through image processing. Optical anchors can be passive patterns such as two-dimensional barcodes or visual targets that are easily parsed, or through active anchors that use synchronized signaling such as blinking infrared emitters.

Head-mounted systems that are mobile are now proliferating. A user can move through an open environment wearing a self-contained eye-tracking system. The success of this type of system is providing expanded opportunities for eye tracking in natural work environments.

Both head-mounted and remote optics systems can track a single eye or both eyes of a user. While single-eye tracking provides the direction of gaze, binocular systems can provide estimations of both gaze direction and gaze depth, through calculation of the eyes' vergence – the point at which the two eyes' directions of gaze intersect. An advantage of binocular systems is that they can compensate for parallax. Parallax errors occur when an object is being tracked at a distance that is different from the distance used for calibrating the tracker. For example, focusing your eyes at a point directly in front of you a few inches in front of your nose will result in different eye positions than if you are looking at a point close to the horizon. Binocular systems will accurately track gaze location in these situations, while single-eye systems will not.

Tracking Methods

The specific techniques by which the tracker converts the eye image to a calculation of gaze direction vary greatly. They all attempt to calculate the relationship between the characteristics of the eye to one another or to other parts of the anatomy, most often as the relation between the pupil and reflections of light off the cornea, but also can include the relation of the pupil to head position, or nose position. One major type uses "dark pupil" imaging, in which the image of the pupil of the eye is a dark spot, as it appears when we see an eye in everyday life. Light pupil trackers reflect infrared light off the retina so that the pupil appears light, similar to the "red eye" seen sometimes in flash photographs. The bright pupil imaging can facilitate image processing because of the increased contrast between the iris and the pupil. Bright pupil technology is more difficult in outdoor environments, or other places that have infrared light sources that can interfere with imaging.

Recent trends in tracking systems have included open-source projects that have provided basic source code to create low-cost remote (Li and Parkhurst, 2006) or head-mounted (Li, Babcock, and Parkhurst, 2006) trackers using commercially available hardware and open-source software. The technological building blocks for eye tracking can be attained from Web cameras or other low-cost digital imaging devices, face recognition software, or video game controllers such as the Kinect or Wii. Some smart phones are already using their built-in cameras to “know” when users are looking at the screen.

Considerations for Selection

There are a number of considerations when buying an eye-tracking system. It is important to understand the requirements for the tracking task to determine which characteristics of the tracking system are essential. Because of the close dependence between hardware requirements and task requirements, some of these considerations apply to task design as well. Criteria for selection of a tracking system may include the following.

Accuracy and Precision

This is the angular resolution and consistency with which a tracker functions. Accuracy varies widely between systems. Different technologies range from high accuracy of less than 1 degree of visual angle to 5 degrees or more. When considering the accuracy needed, consider the size of the visual target the user will be looking at, the distance to the target, and the proximity of the target to other targets. Low tracking accuracy is acceptable if the target areas are large (measured by the visual angle they subtend, not absolute size) and are separated by adequate distance.

Mode of Data Processing

This is whether gaze calculations are carried out in real time or in postprocessing. It is important to know the extent to which you can monitor the tracker’s accuracy and status during data collection. Many systems allow real time calculation and monitoring of eye gaze while the user is being tracked. This allows you to check the calibration of the tracker during the experiment session. Other systems allow you to monitor the eye camera to see whether good data are being extracted from the image but do not integrate the data with the scene camera in real time. Some combine eye data and scene data only in postprocessing. Lack of real time monitoring is most often an issue in mobile, head-mounted systems, where miniaturization and mobility are a trade-off with functionality.

Tracking Frequency

This is the number of times per second that a tracker calculates eye position. For neuroscience applications, high-speed trackers are used for real time calculation of eye gaze at 250 Hz, 1000 Hz, or more. Where high-speed tracking is not needed, 24 Hz, 30 Hz, or 60 Hz systems are often sufficient. Some systems provide adjustable frame rates.

Field of View

This is how far the user's gaze can deviate from looking straight ahead. This determines what angle of the environment the user can survey while being tracked accurately. Wide field of view is important in unconstrained settings, but less so where the area of visual interest is very narrowly defined by, for example, a computer screen in human-computer interaction studies.

Tracking Volume

This is the extent to which the users can move their head in space while being accurately tracked. Some tasks require the user to move around a workspace in order to complete the task. In order to collect valid data, trackers must be able to track the user within a space that allows the user to function normally.

Tracking Stability

This is the robustness of the tracker's ability to track in different environments and over time. How well does a given tracker work with people with brown eyes, or people with glasses, or on children, or in diverse lighting conditions? Trackers can be used in a range of conditions and with a range of users. Not all trackers function well in all circumstances. How long will the tracker maintain good calibration, and will calibration remain good if the user accidentally bumps into the tracker during a task?

These are a few of the questions that may be relevant to ask about a tracker's compatibility with a proposed task. Relevant questions depend on the prospective use of the tracker. While difficult to quantify, tracking stability is nonetheless a critical consideration. Hands-on experience with a tracking system in an analogous environment is often the only way to estimate this important consideration.

Task and Environment

The task being studied and the environment in which it is carried out have a great influence on the type of eye tracking that should be performed and the probability of successful data capture. There are a number of aspects of the task and environment that influence this. Deciding on which characteristics of the environment, task, and user behaviors are relevant is an important consideration that can influence the potential for success.

Just as in many other scientific endeavors, the ease of measurement is often related to the degree to which it is possible to isolate the activity of interest from extraneous influences. Narrowly focused investigations yield the "cleanest" results, but often at the price of reducing the external validity of the conclusions (the applicability of the results to broader contexts).

Visual Tasks

There are many tasks of potential interest in which the user's activity is nearly entirely visual in nature. Examples of visual tasks include reading prose, labels, or warning signs; viewing printed advertisements; scanning maps or schematic diagrams; viewing photos of faces or other visual scenes; or viewing a movie, didactic instructional session, or social interaction. Each of these tasks may be of interest as a component of a larger set of tasks in an applied setting. Many of the core cognitive sciences have used this type of isolated eye-tracking task to answer fundamental questions about human cognition and development (Gredebäck, Johnson, and van Hofsten, 2009). Eye tracking can be extremely effective and straightforward in these types of tasks, as analysis of the tracking is not complicated by actions the user may take. The environment of interest is often constrained to a small field of view such as a piece of paper or single video/computer screen. Static images further facilitate straightforward analysis. By using tightly controlled visual environments and constraining the task to visual activities, highly accurate and detailed information can be collected about users' gaze patterns.

Visual-Manual Task

Applied settings often include a user interacting with an environment or system in order to complete a task manually. Eye-gaze information can supplement a researcher's understanding of the strategies that the user employs to complete the task. Eye-tracking data can be combined with records of the user's activity (e.g., key presses in a computer system; steering behaviors in a driving or flying task; hand movements in complex manual tasks such as surgery or part assembly). Incorporating user activity into eye-tracking paradigm can complicate the analysis of eye data.

Activities may be prompted by external events. For example, in monitoring a process-control application, a user may need to open a valve when pressure reaches a set limit. Eye-tracking analysis can be centered around the events initiated by the process being monitored. This facilitates comparisons between users by creating a common reference point. Eye-tracking data are often segmented into epochs or windows surrounding events, such as different phases of an operating room anesthetic procedure (induction, intubation, maintenance, emergence, etc.) or phases of flight (taxi, takeoff, climb, cruise, descent, final approach, landing).

When activity is user-initiated or self-paced, eye data may be linked to the onset of the user activity. User-initiated activity may not be linked to external events; thus establishing contingencies between gaze data and user activities can create challenges.

Field of View (FOV) Considerations

In eye-tracking research, the relevant operational environments can be as small as a computer screen, cell phone screen, or piece of paper; they can be as large as an entire operating room, airplane cockpit, or even boundless open environment. Knowing, understanding, and choosing an environment and task for eye tracking will determine what hardware configuration can be most appropriate. Limiting the area of interest investigated to the smallest meaningful FOV possible can simplify the technical demands on the eye-tracking system. For example, if you are interested in users' visual scanning when interacting with a computer-based schematic diagram, an eye tracker that only tracks accurately across the small FOV subtended by the screen would be adequate – scanning behavior beyond the screen would not be relevant to the studied task.

Duration of Tracking

Eye trackers can lose their calibration over time and may need periodic recalibration. To ensure adequate calibration, short task durations are preferable. Trackers can be recalibrated between tasks, improving data quality.

Dynamism of Environment

Environments in which the user's head is stationary and the objects in the environment do not move or change provide a static environment for eye tracking. In static environments, the angle between a user's eyes and head correlate very closely to locations within the environment. In contrast, when the user's head moves freely, or when the elements in the visual environment move dynamically (such as people in a work team, or objects of interest on a computer screen) such a correlation does not exist. In dynamic environments, additional steps must be taken to link the eye movements to gaze location (which is a combination of eye orientation, head orientation, and head location) with respect to objects in the environment.

If you have a fixed environment and limited space in which you operate, then there are markers that can be used to "anchor" the eye-tracking data. There are systems for driving or flying environments and other constrained workstations that successfully use markers to make anchoring an automated process. With a data set that is anchored to the external frame of reference, coding fixations into categories corresponding to object or locations in the environment can be an automated process.

In environments such as human computer interaction, where the visual environment may change depending on actions such as the user scrolling down a document or navigating to a new Web page, coordination of eye gaze and screen display requires software based synchronization (Duchowski, 2009; Nielsen and Pernice, 2009).

In the absence of these types of methods of synchronizing the eye data with the external world, the data analysis process involved can be cumbersome and labor intensive. Coding the fixations into meaningful categories may require frame-by-frame manual analysis of the video record of gaze fixations and manual classification of fixations into the various areas of interest (regions in the environment that are functionally grouped together).

Eye trackers still work best when a person does not look very far peripherally (deviating from straight ahead). Trackers have an operational FOV. They work best at a FOV similar to that of their calibration matrix (and for monocular trackers, similar distance). Many “real world” tasks require looking far right, left, and so on (Sanders, 1970). Some trackers may not record reliably under such conditions. When possible, try to structure your task to allow straight-ahead viewing. Manual tasks such as using a keyboard, flight yoke, or steering wheel may anchor or constrain the participant’s body to a given orientation. This may prevent the participant from reorienting the body to peripheral stimuli and lead to higher likelihood of the angle of gaze deviating significantly from straight-ahead.

Recommendations for Applied Research

Technical constraints on eye tracking vary between different tracking systems. However, in general, providing a controlled, consistent environment can facilitate ease of collection, analysis, and interpretation of eye-tracking data. Paradigms that facilitate eye-tracking research include tasks requiring a small FOV, tasks conducted in a static environment, tasks of short duration, tasks that permit a fixed head position, tasks that take place at fixed distance from the participant, tasks that have a discrete onset and are system-initiated. Paradigms that complicate eye-tracking research include tasks that have a wide FOV, are conducted in a dynamic environment, are of extended duration or are continuous, require free head and body movement, have displays that are of variable distance from the participant, and do not have a discrete onset or are user-initiated.

This maximally constrained paradigm of a static, small FOV environment and short duration, discrete tasks will (depending on the tracking technology) tend to produce the highest quality eye-tracking data. Changing any dimension of this paradigm will necessitate additional technology or techniques to compensate for the lack of controls. The appropriate technology or techniques may only compensate only partially, resulting in a degradation of data quality.

Data Analysis and Interpretation

After data have been collected, analysis takes place. Eye-tracking data are traditionally divided into *fixations*, where the eye is looking at a location; *saccades*, where the eye is transitioning rapidly from one location to the other; and *smooth pursuit* movements, where the eye is following the motion of an object in the environment.

The "objects" in the environment must be defined in order to determine where fixations occurred. When the environment is structured and static, and when the gaze-tracking data can be linked to absolute coordinates in the environment, software can be used to classify each fixation automatically into one of the defined objects or areas of interest in the environment. When the environment is dynamic, or when the gaze data are recorded relative to the user's perspective (without link to absolute coordinates), gaze-tracking data may need to be classified manually by reviewing video data of the user's gaze frame by frame to determine where the user was looking. Needless to say, this can be an extremely labor-intensive process.

Instead of identifying each object in an environment, a strategy often employed in data analysis is the use of broader areas of interest. For example, in studying driving applications, areas of interest could include the dashboard (speedometer, fuel gauge, odometer, etc.), console (radio, heating controls, etc.), mirrors (rear view and side view), passenger, and through-the-window view area. The degree of specificity/granularity of the areas of interest should be determined by the accuracy and precision of the tracking system, combined with the characteristics of the task and the environment.

Quantitative Analysis

After the gaze data have been coded into fixations, saccades, and pursuits, further analysis is needed to interpret the meaning of the data. The eye tracker creates a record of the user's gaze, but understanding what that record means requires further processing/interpretation.

There are simple measures that are commonly used for basic interpretation.

Fixation Counts or Proportions

How often did the user look at a given object? Was it more or less often than at other objects in the environment? The number of fixations that are made to an object can be an indicator of the importance of the object, and these numbers can be compared between objects. It can be argued that, in general, if someone is gazing at an object, he or she is paying attention to that object. This is by no means always true (Posner, Snyder, and Davidson, 1980), but for applied research, it is most often true. In applied research, if a person is looking at an object, then that object is the most likely target of the user's attention.

Fixation Duration

When a user looks at an object, how long does he or she look at it before looking away (average duration)? Over the course of the data collection, how long did the user spend looking at that object (total duration). These two duration measures are also used as an indicator of an object's importance.

Interpretation of duration of frequency of fixation can be a challenge and depends on the task context in which the fixations occur. In the study of aviation displays, for example, it has been said that fixation frequency is an indicator of the importance of a display item, while fixation duration is an indicator of how difficult it is to interpret the meaning of the item (Senders, 1984). This interpretation is particularly apt for tasks that involve scanning displays for information about a task that is being carried out in the environment. For example, altitude indicators will be checked often, but the fixations will be brief during instrument scans. Conversely, an unfamiliar icon that appears on the aviation display will be fixated longer – until it is recognized – independently of its inherent importance. The interpretation of fixation frequency and duration in other types of tasks, such as viewing objects that are being manipulated in a work environment, can be quite different. Fixation duration can be an indicator of an object's importance or salience (Nielsen and Pernice, 2010). A user will spend the most time looking at the face of the person who is talking, and less time looking at the face of someone not talking.

Saccades can be the source of information as well. Understanding the pattern of saccades can indicate patterns of information gathering and information dependencies. Frequent transitions (saccades) between two objects can indicate a link between the objects in the user's mind. Patterns of transitions can be interpreted as indications of performance strategies. Expert pilots have more regular scan patterns than novices. Expert attending surgeons have more regular scan patterns than trainees. The regularity of the scan patterns can be an indication of expertise in some cases.

There are countless data analysis and data modeling techniques that can be used to interpret eye-tracking data. Goldberg's work (Goldberg and Helfman, 2014) outlines more sophisticated analysis techniques for visual comparison tasks. Sophisticated sequential analysis techniques can be used. Statistical modeling of experts and novices can be used to classify performance (Ahmidi et al., 2012). There are eye data that can be used as a measure of workload (Camilli, Nacchia, Terenzi, and Di Nocera, 2008). For those interested, further reading provides in-depth descriptions of such techniques (Duchowski, 2009).

Qualitative Analysis

In some applied settings, presentation of qualitative data can communicate powerful messages. There are a number of ways of communicating eye-tracking data in a qualitative, graphical manner effectively. Heat maps show the result of a user's scan of an image or environment. Areas that were fixated for longer amounts of time show "hotter" colors. An example is in Figure 5.2. Heat maps are often used to show the effectiveness of a Web page design or print ad design. If important information is not "hot," the design may not be effective. Circle-line diagrams and scan-path maps show the scan pathways of the user. Eye movements leave a "breadcrumb trail" or line that is overlaid on the image of the environment/screen, and at the conclusion of the tracking session, the scan-path map shows all of the lines indicating patterns of movement. Scan-path mapping can also indicate fixations by drawing circles at the point of fixation, with longer fixations represented by larger circles.

Playing back the raw eye-tracking data can be an effective way of communicating intuitively to system engineers or Web page designers about what users looked at. The scan paths can be presented as a "replay" of the user's scan path. Patterns can be more salient by having the breadcrumb trail appear as the eye moves from one fixation to the next, showing the linking lines and the fixation circles retracing the user's gaze pathway.

What to Expect from Applied Eye-Tracking Research

Once a tracker has been selected, a task designed, and an analysis planned, the collection of tracking data can commence. The collection of data can present challenges, as well.

With current technology, it is not unusual to find that a relatively large proportion of the users you recruit to participate will not produce usable data. For planning purposes, one can assume 30–50 percent of the users will not produce usable data, so your planned population should be 50–100 percent larger than your required sample size. Some users will have eyes that do not track well because of their shape, size, or color or because they wear glasses. As a result, it will not be possible to calibrate them and get a valid tracking signal. Others may calibrate well initially, but tracking will fail to produce a signal intermittently during the task.

Some data may be lost because of the structure of the task, or because of a user's strategy in completing the task. Some users will keep their head still

Figure 5.2. *An example "heat map" representation of eye movement data. (Photo courtesy of E. Warm; eye movement analysis courtesy of Feng-GUI.com.) (See Plate 1.)*

and rotate their eyes, while others may rotate their heads more and their eyes less. Depending on the tracker, these different strategies may produce poor-quality data. If important visual activity falls outside the accurate range of the tracker, data will be lost.

Trackers may be poorly calibrated initially or lose accurate calibration over the duration of a task. Poor calibration can lead to inaccuracy in which fixations are recorded with a tracking error that is constant (e.g., fixations are recorded as consistently 5 degrees to the left of the user's actual fixations) or nonlinear errors, in which there are regions in the visual field that are not well mapped to true eye fixations.

Calibration may deteriorate because head-mounted optics are bumped or slide out of position over time. This can lead to a loss of signal or to low accuracy.

Accuracy of the system may not be sufficient to analyze the tracking data at the desired level of detail. Having coding categories that rely on small areas of interest or ones that are in close proximity to one another may result in unusable data. When planning a coding strategy for eye-tracking studies, categories may need to be able to be reduced from highly detailed areas of interest to more general, consolidated ones when calibration or resolution is reduced.

Conclusions

There is a wide variety of application areas for eye-tracking research, and an equally large array of eye-tracking technologies. As the technologies continue to evolve and become more accessible, the barriers to using eye data in applied perception research will be swept away, and eye tracking may become a ubiquitous component of technologies such as mobile devices and computer interfaces. Understanding the ways to structure experimental tasks to facilitate meaningful analysis of gaze data will continue to be a challenge for eye-tracking research, even in the absence of technical barriers. It is hoped that the information presented in this chapter will help establish a way of thinking about eye-tracking research that creates a technology-independent foundation on which sound research can be built.

References

Ahmidi, N., Ishii, M., Fichtinger, G., Gallia, G. L. and Hager, G. D. (2012). An objective and automated method for assessing surgical skill in endoscopic sinus surgery using eye-tracking and tool-motion data. *International Forum of Allergy and Rhinology*, 2(6), 507–515.

Atkins, M. S., Tien, G., Khan, R. S., Meneghetti, A. and Zheng, B. (2012). What do surgeons see: Capturing and synchronizing eye gaze for surgery applications. *Surgical Innovation*, 20(3), 241–248.

Blignaut, P. J., Beelders, T. R. and So, C.-Y. (2008). The visual span of chess players. In B. S. Smith and F. U. Jones (Eds.), ETRA '08: Proceedings of the 2008 Symposium on Eye Tracking Research and Applications. New York: *Association for Computing* Machinery, pp. 165–171.

Bojko, A. (2006). Eye tracking study on the impact of the manufacturer's logo and multilingual description on drug selection performance. In Proceedings of the 50th Annual Meeting of the Human Factors and Ergonomics Society Annual Meeting. Santa Monica, CA: Human Factors and Ergonomics Society, pp. 1112–1116.

Bojko, A. (2013). *Eye tracking the user experience*. New York: Rosenfeld Media.

Camilli, M., Nacchia, R., Terenzi, M. and Di Nocera, F. (2008). ASTEF: A simple tool for examining fixations. *Behavior Research Methods*, 40(2), 373–382.

Chapman, P. R. (1998). Visual search of driving situations: Danger and experience. *Perception*, 27, 951–964.

Chetwood, A. S., Kwok, K. W., Sun, L. W., Mylonas, G. P., Clark, J., Darzi, A. et al. (2012). Collaborative eye tracking: A potential training tool in laparoscopic surgery. *Surgical Endoscopy*, 26(7), 2003–2009.

Duchowski, A. T. (2009). *Eye tracking methodology: Theory and practice* (2nd ed.). London: Springer.

Fitts, P. M., Jones, R. E. and Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, 9(2), 24–29.

Goldberg J. H. and Helfman, J. (2014). Eye tracking on visualization: Progressive extraction of scanning strategies. In W. Huang (Ed.), *Handbook of Human Centric Visualization*. Springer, pp. 337–372.

Gredebäck, G., Johnson, S. and von Hofsten, C. (2009). Eye tracking in infancy research. *Developmental Neuropsychology*, 35(1), 1–19.

Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A. and Wickens, C. D. (2001). Comparison of expert and novice scan behaviors during VFR flight. Paper Presented at the 11th International Symposium on Aviation Psychology, Columbus, OH. Downloaded from http://www.humanfactors.illinois.edu/ReportsandPapersPDFs/isap01/proced01.pdf, 8 August 2012.

Klein, G. (1999). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Kramer, A., Tham, M., Konrad, C., Wickens, C. D., Lintern, G., Marsh, R., Fox, J. and Merwin, D. (1994). Instrument scan and pilot expertise. Paper Presented at the Proceedings of the 38th Annual Meeting of the Human Factors and Ergonomics Society. Santa Monica, CA: Human Factors and Ergonomics Society, pp. 36–40.

Li, D., Babcock, J. and Parkhurst, D. J. (2006). openEyes: A low-cost head-mounted eye-tracking solution. Paper Presented at the Proceedings of the 2006 Symposium on Eye Tracking Research and Applications. New York: Association for Computing Machinery, pp. 95–100. Downloaded from http://thirtysixthspan.com/openEyes/lI_etal06.pdf, 8 August 2012.

Li, D. and Parkhurst, D. J. (2006). Open-source software for real-time visible-spectrum eye tracking. Proceedings of the 2nd COGAIN Conference, pp. 18–20. Downloaded from http://thirtysixthspan.com/openEyes/lI_parkhurst06.pdf, 8 August 2012.

Matsumoto, H., Terao, Y., Yugeta, A., Fukuda, H., Emoto, M., Furubayashi, T., et al. (2011). Where do neurologists look when viewing brain CT images? An eye-tracking study involving stroke cases. *PLoS One*, 6(12), e28928.

Nielsen, J. and Pernice, K. (2010). *Eyetracking web usability*. Berkeley, CA: New Riders.

Pillsbury, W. B. (1908). *Attention*. New York: MacMillan

Posner, M. I., Snyder, C. R. and Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, 109(2), 160–174.

Richardson, D. C. and Spivey, M. J. (2004). *Eye tracking: Characteristics and methods*. In G. Wnek and G. Bowlin (Eds.), *Encyclopedia of biomaterials and biomedical engineering* (pp. 568–572). New York: Marcel Dekker.

Sanders, A. F., (1970). Some aspects of the selective process in the functional visual field. *Ergonomics*, 13, 101–117.

Senders, J. W. (1984). *Visual scanning processes*. Hillsdale, NJ: Lawrence Erlbaum.

Shapiro, K. L. and Raymond, J. E. (1989). Training of efficient oculomotor strategies enhances skill acquisition. *Acta Psychologica*, 71, 217–242.

Williams, A. M., Ward, P., Smeeton, N. J. and Allen, D. (2004). Developing anticipation skills in tennis using on-court instruction: Perception versus perception and action. *Journal of Applied Sports Psychology*, 16, 350–360.

Xiao, Y., MacKenzie, C., Orasanu, J., Spencer, R., Rahman, A. and Gunawardane, V. (1999). Information acquisition from audio-video-data sources: an experimental study on remote diagnosis. *Telemedicine Journal*, 5, 139–155.

6 Applied Perception and Neuroergonomics

Matthew Rizzo and Raja Parasuraman

Introduction

Vision is critically important in many everyday tasks such as map reading, driving, and scanning computer displays. Neuroergonomics can provide methods and theories that can enhance our understanding of visual performance in such settings. This chapter describes the relevant methods and theories from psychophysics, neuropsychology, and cognitive neuroscience that can be used in the evaluation of different aspects of visual perception and cognition. These techniques can be used to understand how normal variation in visual perception and cognitive abilities influence performance on occupational and everyday visual tasks. They can also be used to examine how impairments due to neurological or psychiatric disorders, aging, drugs, and so on, affect visual performance.

Psychological research in visual perception has a rich history of making contact with findings from neuroscience. Whether one goes back to the time of Helmholtz in the nineteenth century, or to the 1950s, when David Hubel, Torsten Wiesel, and others made their groundbreaking discoveries of lateral inhibition, the study of visual perception in particular has benefited greatly from an understanding of the neural structures that support vision. Many recent texts on vision reflect this joint use of evidence from both behavioral and brain science (e.g., Palmer, 1999; Ullman, 1997).

The picture is somewhat different when one considers the functions visual perception serves – to sense, recognize, navigate, and act in the everyday world – not only in relation to natural objects but also with human-made artifacts such as tools, cars, aircraft, and computer displays. Neuroscience tools and methods are providing greater understanding of visual tasks carried out in naturalistic settings. Vision is critically important in such everyday tasks as understanding a map, crossing the street, reading a book, driving a car, or apprehending complex dynamic information displays in modern aircraft. In such settings people – drivers, pilots, nuclear plant operators, and other workers – must monitor multiple inputs from central and peripheral vision and the other senses, allocate attention among environmental targets and distracters, and correct their errors while also monitoring personal status or other factors that might compromise performance. Furthermore, individuals with impairments of visual perception, cognition, and memory, due, for example to poor sleep, fatigue, medication effects, illicit drugs, or alcohol, are liable to make poor decisions based on faulty inputs, increasing their risk of errors and injuries. A minority of individuals in the workplace may also have neurodegenerative or other medical disorders (whether diagnosed or undiagnosed) that can also affect their visual performance.

Such *applied* perception issues are the focus of the chapters in this volume. In this chapter we discuss how findings from neuroscience, specifically cognitive neuroscience and neuropsychology, can inform the application of perception theory and research. The consideration of neuroscience for "real-world" issues falls within the domain of neuroergonomics, which has been defined as the study of the human nervous system in relation to performance at work and in everyday settings (Parasuraman, 2003; Parasuraman and Rizzo, 2007). The central premise is that research and practice in human factors and cognitive engineering can be enriched by consideration of theories and results from neuroscience.

Some decades ago such a claim would have been considered far-fetched because our knowledge of human (as opposed to animal) brain function was limited and restricted to only the simplest aspects of behavior. But with the phenomenal growth of cognitive neuroscience, it is increasingly the case that theories of human performance can be constrained or extended by consideration of findings from the neurosciences. Neuroergonomics can therefore provide added value, beyond that available from traditional neuroscience and conventional ergonomics, to our understanding of brain function and behavior as encountered in work and in natural settings (Parasuraman, 2011).

An example is the application of theory and research to issues entailed by sensory or cognitive impairments. Humans vary widely in their visual abilities and skills; moreover, some individuals may have visual disorders that reduce or distort information, discrimination, judgment, understanding, and performance. Such impairments can increase the risk of errors, injuries, and other adverse outcomes, as described in a recent review of the role of vision in everyday activities (Wood, Owsley, and Rubin, 2011).

Measuring these abilities and determining how weaknesses in perceptual domains map onto job tasks are directly relevant to the practice of neuroergonomics and benefit from an understanding of the conditions that commonly affect human visual performance. We do not provide a comprehensive survey of issues in visual perception that are relevant to performance in the workplace. Rather, we discuss selected topics that provide examples of how both normal variation in visual functioning, as well as visual impairments, and an understanding of their neural bases, have implications for the design of workplaces and for the training of workers.

We begin with relatively basic properties of the visual system such as visual acuity and color perception and progress to increasingly complex functions such as motion processing. Finally we cover "higher-level" functions such as visual attention and mental workload, closing with a discussion of multisensory functions and operational considerations. In each case we describe the typical psychophysical, neuropsychological, or cognitive neuroscience methods that are used to evaluate these visual and cognitive functions.

Visual Acuity and Contrast Sensitivity

Visual Acuity

Often the only visual criterion used to assess a worker's fitness (e.g., a locomotive engineer) is visual acuity, which can be easily measured using visual tests. One such test is the Early Treatment Diabetic Retinopathy Study chart (Ferris, Kassoff, and Bresnick, 1982) for assessing pattern vision (letters), which has been used in several major clinical trials sponsored by the National Eye Institute. Impairments in acuity decrease performance on many tasks and can be corrected with lenses, although there can be distortions associated with improper refraction, bifocals, or difficultly getting used to the prescriptions. What the most reasonable cutoff scores should be for acuity on certain tasks is an underinvestigated question. For example, many would likely agree that visual impairments of 20/200 should preclude automobile driving, but some states use a 20/40 cut off score, which may be both inappropriate and unfair. Pilots are often held to higher standards.

Spatial Contrast Sensitivity

This is the ability to perceive patterns (such as gratings or letters) presented at different contrasts and may be a better predictor of worker competency (with regard to visual tasks) than simple visual acuity. Monocular and binocular contrast sensitivity are easy to measure using a wall chart (Pelli, Robson, and Wilkins, 1988) that provides a measure of low to medium spatial frequency sensitivity (i.e., near the peak of the human contrast-sensitivity function). Contrast sensitivity can be measured under standard photopic conditions and under low visibility conditions by viewing the contrast sensitivity chart through low transmittance filters. This approach may be useful for screening individuals who have trouble seeing in low-light conditions, but it remains a research issue. People who are aware of their visual defects may compensate well enough to perform safely during the daytime.

Glare

Glare can impair visibility on sunny days in the summer, in the winter from reflections off the snow, and at night from streetlights, halogen lamps, or oncoming vehicles. Glare discomfort is a complex phenomenon with no single good measure, and glare is not always related to performance impairment since people can often compensate (e.g., by squinting or averting their gaze). However, glare sensitivity (or disability) is tied to performance impairment and is more directly relevant to worker safety. Corneal surgery to mitigate the need for corrective lenses may increase glare discomfort and disability. Older people with cataracts may report rings of glare around glowing objects like lamps and headlights. People may also have impaired low-luminance vision, independent of glare. Glare disability can be tested using a commercially available Brightness Acuity Tester, which floods a viewer's eye with light as the viewer tries to read a visual acuity chart (Holliday, Trujillo, and Ruiz, 1987). Whether these phenomena affect worker performance and safety is an open research question for many domains of work.

Color

Perception of light by normal humans can be characterized using a three-dimensional "color space." (For a discussion of a variety of ways to represent the variables that determine color – alternative color "spaces" – see Palmer, 1999). The three axes of this space are an achromatic dimension and two chromatic axes, which can be related to the activity of the three retinal cones. Relative differences in the activity of the three retinal cones give rise to chromatic differences. Thus, along the deutan chromatic axis, the activity of the S-cone (maximally sensitive for short wavelength light) remains constant, but the activity of the M-cone (medium wavelength) changes with respect to that of the L-cone (long wavelength). Along the tritan axis, the ratio of M-cone to L-cone activity remains constant, while S-cone activity varies. Variations along these two chromatic dimensions cause differences in hue, which are related to the perceived dominant wavelength (i.e., red, yellow, blue) as well as to differences in saturation, which are related to the purity of the spectral composition, with less saturated shades representing greater mixture of white light with the dominant hue (e.g., pink, rose, red). Differences in brightness ("luminance" for light sources, "reflectance" for objects that reflect light) occur when light changes along the third dimension, the achromatic axis. Along this axis, the activity of all three cones changes together, such that the ratio of one cone's activity relative to another remains constant. Conventional color diagrams such as the Commission Internationale de L'éclairage (CIE) chart (Jones, 1943) depict the colors within a plane of the two chromatic dimensions, omitting brightness.

No single cone can distinguish the wavelength of the light it receives, and wavelength sensation is not the same as color perception. Perceived color is the result of the interaction between the appearance of surfaces and the human visual apparatus and is not inherent in objects (Wyszeki and Stiles, 1982). Most objects we see are not light sources but are surfaces that reflect, emit, or transmit light. The wavelengths that they transmit to our eyes are determined by both the reflectant properties of the object and the illuminating light in the scene (Boynton, 1992). When the illuminant changes, so does the wavelength composition that reaches our eye from the object. Yet, the perceived colors of objects are remarkably stable under different lighting conditions, a phenomenon known as "color constancy" (e g., an orange seen under a yellow light still looks orange). Therefore, wavelength composition alone cannot determine perceived color (Land, 1986). Rather, the relative proportion of received light at each wavelength reflected by the object determines color. For complex everyday scenes (versus the displays used in the perception laboratory), the nervous system attempts an indirect estimate of the illumination (if we may speak of the nervous system as a computational mechanism). This may involve averaging the spectral luminance across large regions of the scene, deducing what kind of lighting is present, and then "discounting the illuminant" from the wavelengths reflected by a given object (Land et al., 1983). This generates color as a stable property of objects, over a fairly wide range of natural illumination.

Color cues allow us to parse information in scenes from chromatic boundaries. These cues allow for the recognition of targets in natural scenes amid glare, shadows, camouflage, and transparencies, which can reduce conspicuity and mask object borders (Dominy and Lucas, 2001). While these optical phenomena would seem to be important for work, many employers do not require color-vision testing in operator assessments. For example, factory work that requires assembly of color-coded components may be severely affected by color vision loss. Weather meteorological displays often include several colors, often more than necessary (Hoffman et al., 1993, recommended no more than 12), so that undiagnosed deficiencies in color discrimination in weather forecasters who use these displays could affect their work. Of course, the need for color discrimination ability depends on the task, and color-impaired workers can often use cues other than color to discriminate and recognize objects. In the case of automobile driving, traffic signals are often mounted in a standard vertical order so that a knowledgeable driver can infer stop (red on top), go (green on bottom), and caution (yellow in the middle) despite color-vision loss.

Impaired color vision may be detected with pseudoisochromatic color plates, as in the Ishihara test, or AO-14 test (Rizzo and Barton, 2005). Color sorting or matching tests provide more detailed evidence of impaired color perception. Typical sorting tests include the Farnsworth-Munsell 100 Hue Test and the shorter items of Farnsworth D-15 and the Lanthony New Color Test, which test hue discrimination, and the Sahlgren Saturation Test, which measures saturation discrimination. Congenital color weakness and blindness typically affect discriminations along the red–green color axis, and less commonly, the blue–yellow axis. The degree of deficit can range from complete to partial defects. Lighting to maximize perception of colors and chromatic boundaries can improve operator performance in work environments (Ferguson, Major, and Keldoulis, 1974).

Visual Fields

The binocular visual fields normally subtend more than 180° across and about 100° vertically. The fovea subtends about 3° and has the highest visual acuity. The macula or parafovea spans about 10° and also participates in visual tasks that demand fine visual resolution such as reading text, maps, signs, dials, displays, and gauges. The peripheral visual fields extend beyond this and have low visual acuity but good temporal resolution and motion detection. Standard perimetry tasks such as Goldmann (dynamic) and Humphrey (static automated) perimetry minimize attention demands to gain maximal estimates of sensory ability (Rizzo and Barton, 2005). In these tests the viewer is asked to report the detection of light spots of varying intensities or sizes presented against a white background in different parts of the field, with his or her head and eyes held still and focused on a centrally located fixation spot. The results provide a map of the viewer's visual sensitivity from the center to the periphery of the visual field in all quadrants, including areas of weakness or blindness.

Visual field defects may arise at the level of the eye or the brain. The effects on worker performance depend on the location of the defect within the visual fields and the specific types of processes affected within the abnormal region. The many possible degrees of visual field loss correspond to different lesions in the visual pathways. Visual field size is a predictor of performance on a variety of tasks that require visual search, such as safely driving a motor vehicle, searching for a target in a visual array such as a specific face or person in a crowd, or scanning for information in text or video displays. Some individuals with acquired visual field defects may experience a "hole" in their vision (Allen, 2000). The added task of having to remember to search for critical information in the areas of impaired vision creates an extra cognitive load or interference, tantamount to the burden of multitasking. A visual fields defect at or near the fovea due, say, to glaucoma or retinal detachment, causes serious problems for many tasks because of impaired ability to fixate fine details. Keyhole or tunnel vision with fields spanning just 20° is another serious problem. For example, persons with retinitis pigmentosa who have marked constriction of the peripheral visual fields may be unable to detect objects that approach from the side. But there is a vast gap between these situations and full monocular or binocular fields and the effects of many patterns of loss on worker performance are largely unexplored.

There is a paucity of research on operator performance and errors due to cerebral visual field loss. Lesions of the primary visual cortex (in Brodmann area 17, or V1) or white matter produce defects in the visual fields opposite the side of the lesion (see Rizzo and Barton, 2005, for a detailed review). These defects are homonymous in that they occupy the same hemifield in each eye (because of the reversal of real-world images by the lens and crossing of nasal fibers of the optic nerve), meaning the defects in the two eyes are nearly identical when superimposed. Hemianopia refers to loss of half of the visual field. Operators with hemianopia cannot see objects on one side of fixation. A visual field defect that is restricted to the upper or lower quadrant of a hemifield is known as a quadrantanopia. A lesion below the calcarine fissure of the primary visual cortex results in an upper quadrantanopia. A lesion above the calcarine fissure causes a lower quadrantanopia and can be a problem for reading, navigation, and many tasks in the workplace. Damage to the macular representation in V1 is troublesome because it may interfere with ocular fixation, visual scanning, and the ability to process visual spatial details.

Stroke, trauma, and tumor commonly cause the cerebral lesions that produce visual field defects. These lesions often extend into the prestriate cortex (Brodmann areas 18 and 19 or area V2/V3),adjacent temporal lobe, and parietal lobe (Rizzo and Barton, 2005). The resulting perceptual defect is less well localized than those caused by V1 lesions.

These defects have been explained using the heuristic of parallel processing in two visual systems originating in area V1: a “what” pathway and a “where” pathway (Ungerleider and Mishkin, 1983). Lesions in these pathways can impair visual processes independently of V1-type visual field defects. Damage in the ventral occipital lobe and adjacent temporal regions along a “what” pathway is associated with defects of visual recognition (visual agnosia), color perception (cerebral achromatopsia), and reading (acquired alexia). These conditions can impair worker performance, even in the absence of a visual field defect. Damage along the occipital-parietal “where” pathway is associated with eye and hand control, disordered visuospatial attention, and impaired motion processing (cerebral akinetopsia) (Rizzo, Nawrot, and Sparks, 2008). Patients with Balint syndrome often have bilateral parietal lobe lesions (due to stroke or a visual variant of Alzheimer disease) and are severely incapacitated (Rizzo and Vecera, 2002). Patients with hemineglect, a neurological syndrome most often associated with a lesion of the right parietal lobe, often fail to attend to stimuli in the left hemifield. Some of these individuals are looking but not seeing (Rizzo et al., 2001), a problem that also affects individuals with sleep deprivation or metabolic disorders (Rizzo, 2011).

Information from parallel pathways is also processed outside the visual cortex. Damage to the prefrontal cortex may impair mechanisms for “executive attention” and working memory that briefly maintain visual information (such as the location and identity of other vehicles near the driver’s car) so that it is available for use (Rizzo and Vecera, 2002). Damage to the cerebellum may impair neural mechanisms that distinguish between image movement across the retina and self-movement (Nawrot and Rizzo, 1998), which is important for perception of heading, collision detection, and related abilities.

Structure, Motion, and Depth

Binocular stereopsis depends on the separation of the eyes in the head, casting different projections of images of an object on the two retinas, and provides unambiguous cues for perception of object structure and distance relative to an observer. Binocular steroacuity can be measured using hand-held cards containing polarized images or red-green images (anaglyphs) that are viewed, respectively, with polarized lenses or spectacles with one green and one red lens. Such impairments may affect up to 10 percent of the general population. There is little evidence, however, on the extent to which such impairments affect performance on many job tasks. This is probably because information on object structure and depth is so critical for interacting with objects and obstacles that our brains use multiple redundant cues besides binocular stereopsis. These cues include accommodation, convergence, binocular disparity, motion parallax, texture accretion/deletion, convergence of parallels, position relative to the horizon, relative size, familiar size, texture gradients, edge interpretation, shading and shadows, and aerial perspective (Palmer, 1999).

Perception of structure-from-motion or kinetic depth may fail in patients with visual cortex lesions due to stroke or early Alzheimer's disease (Rizzo et al., 1995). Structure-from-motion deficits in drivers with brain lesions are associated with increased risk for safety errors and car crashes in driving simulation scenarios (Rizzo et al., 2001). Recovery of depth from motion relies on relative movements of retinal images. For motion parallax, relative movement of objects is produced by moving the head along the interaural axis. Impairments of motion parallax may be a factor in vehicle crashes or falls to the ground in persons with cerebral impairments, who must make quick judgments with inaccurate or missing perceptual information regarding the location of surrounding obstacles, and may contribute to crashes involving alcohol intoxication (Nawrot, 2001). Displacement of images across the retina during self-motion (egomotion) produces optic flow patterns that can specify the trajectory of self-motion with high accuracy (Gibson, 1979; Warren et al., 1989). Perception of three-dimensonal structure from motion is also affected by normal aging (Jiang, Greenwood, and Parasuraman, 1999) and perception of heading from optical flow patterns can also decline.

Drugs such as marijuana (tetrahydrocannibinol [THC]) and ecstasy (MDMA, or 3,4-methylenedioxymethamphetamine) also impair these functions, presumably because of chronic effects on cholinergic receptors (with

THC) and serotoninergic/5-hydroxytryptamine-2 receptors (with MDMA) (Rizzo et al., 2005, 2009). Processing of visual motion cues also may be impaired in patients taking antidepressants, such as nefazodone hydrochloride, that block serotonin reuptake (Horton and Trobe, 1999).

Detecting and acting to avoid impending collision events require information on egomotion and approaching objects. For example, objects on collision paths with an automobile driver tend to maintain a fixed location in the driver's field of view, whereas safe objects will translate to the left or right side (Vaux et al., 2009). Where two roads intersect but one is sharply curved (hence at a fixed location in the driver's field of vision), apparent time to contact is underestimated. Older individuals are not as accurate as younger individuals at detecting an impending collision during deceleration. They are less adept at determining whether an approaching object in a simulator scenario will crash into them (Vaux et al., 2009). Performance is worse for longer time-to-contact conditions, possibly because of a greater difficulty in detecting the motion of small objects in the driver's field of view. Judgments on time to contact can be measured in actual driving tests using radar detectors (Pietras et al., 2005).

Biological Motion and Action Understanding

When objects that move in the visual environment are other people or animals, their motion is referred to as biological motion. In addition to the neural mechanisms of motion described previously, the brain appears to have specialized neural circuits for understanding the movements of biological organisms (Blake and Shiffrar, 2007). Perception of biological motion is important in many everyday tasks, including communication and social interaction with others. Rapidly detecting and identifying the actions and movements of other people are also critical in many civilian and military surveillance environments. For example, sensors on semiautonomous unmanned vehicles and other platforms are increasingly being used to provide video or infrared images to remotely located operators (Cooke, Pringle, Pedersen, and Connor, 2006; Parasuraman, Cosenzo, and de Visser, 2009). More traditionally, closed-circuit television monitors are found in prisons, airports, highways, and busy city streets. Surveillance images show people or vehicles in motion and engaged in various activities. Such information can be used to identify individuals who pose potential threats or to determine the potential for danger in crowd control situations.

Coding the actions of other people is a key function of the superior temporal sulcus (Thompson and Parasuraman, 2012). This brain region may integrate neural signals coding for motion and object form and then send representations of actions to higher brain regions such as the prefrontal cortex. It is through the coordinated interaction of these regions that a viewer is able to understand and interpret the movement trajectories of others, thereby gaining an understanding of their intentions.

Perception of biological motion appears to be a capability at birth, as evidenced by studies in human and animal neonates (Simion et al., 2008). Yet, despite its early development and robustness, biological motion is influenced by attention, particularly when stimuli are ambiguous, are degraded, or overlap one another (Thompson and Parasuraman, 2012). Under such circumstances, attention may be needed to resolve ambiguity, boost processing of a degraded stimulus, or force a choice between competing stimuli. For example, using functional magnetic resonance imaging (fMRI), Safford et al. (2010) showed that activation of the superior temporal sulcus was strongly modulated when participants had to attend to biological motion (e.g., human stick figures performing jumping jacks) in the presence of overlapping, competing nonbiological stimuli (e.g., tool motion). Biological motion perception is also impaired when participants have to perform a dual task (Thornton, Rensink, and Shiffrar, 2003) or perform a vigilance task over an extended period (Parasuraman et al., 2009), but only when stimuli are visually degraded, as they might often be in tasks such as monitoring video feeds over noisy satellite or unmanned vehicle channels.

These findings have neuroergonomic implications for the workplace. One important implication is that monitoring video screens for suspicious or dangerous behaviors can be an attentionally demanding task and is not impervious to the effects of competing stimuli or tasks or to the occurrence of vigilance decrement. This should be particularly the case when displays are noisy, when multiple stimuli are present, or when inputs are degraded. Furthermore, surveillance can involve monitoring for behaviors that are ill defined, ambiguous, and highly context-dependent, and which might be even more attentionally demanding than the clearly defined tasks studied within the laboratory. Consequently, personnel in these and other security work settings may benefit from specific types of attentional training. Such training could take the form of showing operators videos of scenes and events and asking them to detect particular actions while concurrently searching for another nonmotion feature, such as the face of a specific individual or the color of his or her clothing.

Visual Attention

Attentional factors are important not only in perception of biological motion but in many other perceptual tasks (Nakayama and Joseph, 1998), making the analysis of visual attention key to understanding performance at many everyday visual tasks. Conventional visual field tests such as Goldman and Humphrey perimetry may overestimate functional ability in individuals engaging in real-world tasks that demand peripheral vision. The Useful Field Of View (UFOV^R) is a compound measure that depends on speed of processing and attention (divided and selective) and has been interpreted as the visual area from which information can be acquired without moving the eyes or head (Ball et al., 1993). The attended field of view is similar to the useful field of view except drivers are allowed to move their eyes and head (Coeckelbergh, 2002). The efficiency with which drivers can extract information from a cluttered scene (such as a busy work environment) begins to deteriorate by 20 years of age (Sekuler et al., 2000). A fundamental mechanism underlying Useful Field of View deficits appears to be a failure to disengage attention (Cosman et al., 2012).

Another important visual attention phenomenon is “change blindness,” which refers to changes to objects or scenes that are often missed (Rensink, O’Regan, and Clark, 2000). Individuals are also less able to consolidate, perceive, and remember information stored in working memory (inattentional amnesia), particularly with dynamic displays (as in the “attentional blink”; Rizzo et al., 2001), high information load (Rizzo et al., 2009), and irrelevant distracters (Kramer et al., 2001; O’Regan, Rensink, and Clark, 1999). Change blindness occurs even in persons who have normal vision but who are unable to detect critical changes in a scene because of a brief visual disruption. The disruptions can include saccades, flickers, blinks, camera cuts, or gradual image changes. Change blindness probably depends on visual working memory and spatial attention. Change blindness is more likely when working memory is occupied by other information or working memory capacity or duration is impaired (e.g., because of aging, neurological disease, drugs, or fatigue), and it reduces the ability to perceive salient changes in traffic related scenes (Rizzo et al., 2009)

The attentional blink is another type of blindness that can occur in people with normal vision. When we identify a visual object, our ability to perceive a second object is impaired for several hundred milliseconds (because visual working memory is still occupied by the first object when the second arrives). This period, known as the attentional blink, is not due to an eye blink and can be measured in a laboratory setting using a rapid serial visual presentation of visual targets (often a sequence of letters) on a computer monitor. The attentional blink can be greater in patients with a variety of brain lesions because of reduced temporal processing speed and working memory (Rizzo et al., 2001). Increased attentional blink may impair a worker’s ability to perceive information from a continuous stream of targets, obstacles, and distracters.

Executive attention switches the focus of attention among critical tasks, as in automobile drivers tracking the road; monitoring the changing locations of neighboring vehicles; reading signs, maps, traffic signals, and dashboard displays; checking the mirrors; and choosing to be distracted by noncritical secondary tasks such as cell phone conversation or texting. This can involve switching attention between disparate spatial locations, local and global object details, or different visual tasks and is thought to rely on mechanisms in the prefrontal areas (Cosman and Rizzo, 2012).

Sustained Attention

The tests of visual attention described previously typically apply to situations where people need to select information in the presence of distractors or switch between multiple targets. However, many activities require that attention be maintained on a task for long periods, particularly so that infrequent but critical targets are not missed. Such sustained attention or vigilance tasks have been the object of investigation in many studies (Warm, Parasuraman, and Matthews, 2008). Neuroimaging studies using fMRI have shown that a network of brain regions in the right hemisphere, including the parietal and frontal cortex, are critically involved in sustained attention tasks (Parasuraman et al., 1998). Because of some limitations of fMRI procedures with respect to the study of long-term vigilance (observers need to remain motionless and the scanning environment is very loud), several investigators have used lower-cost but less sensitive alternatives to fMRI, such as transcranial Doppler sonography and functional near infrared spectroscopy.

Transcranial Doppler sonography employs 2 MHz pulsed ultrasound signals to monitor cerebral blood flow velocity in the intracranial arteries of the brain, typically the middle cerebral artery, which can be located in each hemisphere through a "transtemporal window" near the temple (Aaslid, 1986). This neuroimaging technique measures the difference in frequency between the outgoing and reflected energy as it strikes moving red blood cells. The small size of the transducer allows for real-time measurement of cerebral blood flow that is relatively insensitive to head motion. Using transcranial Doppler sonography, Warm and colleagues reported a series of studies of vigilance (for reviews, see Warm et al., 2008; Warm and Parasuraman, 2007; Warm, Finomore, Vidulich, and Funke, this volume). A consistent finding is that the vigilance decrement is paralleled by a decline in blood flow velocity over time, relative to a baseline of activity just prior to beginning the vigilance session. The parallel decline in vigilance performance and in blood flow velocity is found for both visual and auditory tasks (Shaw et al., 2009), pointing to the involvement of a supramodal attentional system.

The result has been interpreted within a resource model of vigilance, in which the decrement is attributed to the depletion of attentional resources with increasing time at work. A critical finding in support of the resource theory (Kahneman, 1973) – as opposed to a generalized arousal model – is that the blood flow change occurs only when observers actively engage with the vigilance task. When observers are simply asked to monitor a display passively without a work imperative for the same time as in an active attention condition – a case of maximal underarousal – blood flow velocity does not decline but remains stable over time.

The degree of blood flow decline in these studies is also modulated by task demands, particularly in the right cerebral hemisphere. The finding of lateralized effects coincides with the results of the positron emission tomography (PET) scan and fMRI studies, which also point to a right hemispheric system in the control of vigilance (Parasuraman et al., 1998), and it rules out the possibility that the hemovelocity effects reflect gross changes in heart rate variability, blood pressure, and cardiac output, because these changes are not likely to be lateralized.

Detection performance in vigilance tasks can be improved by providing observers with consistent and reliable cues to the imminent arrival of critical signals, with the extent of the decrement being reduced or eliminated (Wiener and Attwood, 1968). Such cueing effects are particularly effective when signals are difficult to discriminate because of poor viewing conditions. For example, as described in the section on biological motion, Parasuraman et al. (2009) found that observers exhibited a vigilance decrement when required to discriminate critical signals representing particular combinations of hand movements with objects in video scenes (e.g., a hand picking up a gun in order to fire it versus a similar movement in order to transport the gun), but only when the videos were visually degraded. Under such viewing conditions, precueing the critical signals reduced the vigilance decrement and, when the cue was perfectly predictive of the signal, eliminated the decrement.

If changes in cerebral blood flow are indicative of the success or failure of sustained attention, then perfectly reliable cueing should reduce or eliminate the decline in cerebral blood flow over time on task. This prediction was tested by Hitchcock et al. (2003) using a simulated air traffic control display. Critical signals for detection were pairs of aircraft traveling on a potential collision course. Observers monitored the simulated air traffic control display for 40 minutes under one of four levels of cue reliability – 100 percent reliable, 80 percent reliable, 40 percent reliable, or a no-cue control. Detection performance was stable in the 100 percent reliable cueing condition but declined over time in the remaining conditions, so that by the end of the vigil, performance efficiency was clearly best in the 100 percent group followed in order by the 80 percent, 40 percent, and no-cue groups. These results were accompanied by similar changes in blood flow velocity in the right hemisphere. As was the case with detection probability, the hemovelocity scores for the several cueing conditions were similar to each other during the early portions of the vigil, but showed differential rates of decline over time, so that by the end of the vigil, blood flow was clearly highest in the 100 percent group followed in order by the 80 percent, 40 percent, and no-cue groups.

The goals of neuroergonomics include understanding aspects of human performance in complex work systems with respect to the underlying brain mechanisms and providing measurement tools to study these mechanisms (Parasuraman, 2003). From this perspective, the use of transcranial Doppler sonography measures of cerebral blood flow to assess vigilance can be considered a success. The vigilance studies have revealed a close coupling between vigilance performance and blood flow, and they provide empirical support for the notion that blood flow may represent a metabolic index of information processing resource utilization during sustained attention.

Attention, Multitasking, and Mental Workload

Having considered visual attention and sustained attention, we turn to a broader consideration of attentional factors that come into play in complex visual tasks, of the type encountered in transportation environments such as automobile driving and aviation. Such activities typically involve *multitasking*, whether among the relevant subtasks (such as lane control, speed, and traffic monitoring in driving) or between the primary task and other tasks (such as querying a Global Positioning System – [GPS]-based navigation device or answering a cell phone). The rapid proliferation of portable technologies has increased opportunities for multitasking in everyday life. People can surf the Internet, text each other, listen to music, download videos, and obtain navigational instructions, all while on the move and engaged concurrently in other activities. Such activities can increase the cognitive load on a human operator, thus making its assessment an important issue in many human factors applications (Wickens and McCarley, 2007).

Studies of multitasking have shown that mental workload typically increases to high levels, in some cases to the point where safety is compromised. Many studies have shown that driving performance is degraded when drivers converse on a cell phone (even when using hands-free devices), a performance deterioration attributed to competition for central attentional resources rather than motor interference (Horrey and Wickens, 2006). Resource explanations for performance changes can be circular, because performance is proposed to vary with application of resources, but resources are inferred from performance changes (Navon, 1984). Accordingly, measures are needed that can provide independent assessment of resource competition during multitasking. Neuroimaging techniques such as fMRI and event related potentials can provide such measures. Just and colleagues (2008), for example, used fMRI to compare simulated driving carried out alone and concurrently with an auditory sentence verification task. Even though the two tasks typically activate nonoverlapping cortical networks, the parietal cortex, a brain region activated when driving alone, showed lower activation with the addition of the comprehension task. These and other multitasking driving studies using different methods, such as event related potentials (Strayer and Drews, 2007), provide strong neural evidence for the resource interpretation for multitasking decrement.

Behavioral measures, such as accuracy and speed of response, have been widely used as workload measures. However, neuroergonomic measures offer some unique advantages, including the ability to obtain covert information continuously in work settings where overt behavioral measures may be relatively sparse, as in highly automated airline cockpits (Kramer and Parasuraman, 2007). Another reason is that such measures can be linked to neural theories of attention, thereby allowing for the development of neuroergonomic theories that in turn can advance practical applications involving mental workload assessment. For example, fMRI (Just et al., 2003) and neuropsychological (Previc, 1998) studies have supported the theoretical distinctions among perceptual/cognitive, verbal/spatial, and focal/ambient visual processing resources, which are components of Wickens's (2002) multiple-resource model.

Multisensory Perspectives

While this chapter has focused on visual perception and cognition, other sensory modalities are clearly important in the workplace. Hearing loss is a common outcome in noisy workplace settings (e.g., military, industrial). Hearing aids improve apprehension of face-to-face conversation, but are not as effective for discriminating and localizing salient signals in noisy environments (such as the oft cited cocktail party, a car with engine noise, warning signals, ambient conversation, and Doppler cues from nearby vehicles). Multisensory integration is also clearly important in settings where objects and hazards have multiple sensory attributes (Lees et al., 2010). In the latter case, the defect may come into play at higher levels.

As described previously, attentional abilities are critical for the continuous direction of attention to relevant features in operational environments. Instances of "looking but not seeing" are also more likely during prolonged monitoring of displays, as air-traffic controllers, health care workers, or baggage screeners are required to do (Rizzo et al., 2001). Eye movements can index information processing and perceptual failure in such task settings and depend on the stimulus and its context for visual search (Kramer and McCarley, 2007). Eye movement analyses can assess fixation duration (dwell time), distance, location (in regions of interest), scan path length, and likelihood of transitions between successive fixations. The transportation industry provides a good example for applications. With vehicle operation on straight roads in low traffic, drivers tend to fixate around the focus of expansion, between the road and horizon, in the direction of forward travel. During curve entry, drivers tend to fixate around the tangent to the inner curve unless