EDITED BY

PAUL **WARD**

JAN MAARTEN **SCHRAAGEN**

JULIE **GORE**

EMILIE **ROTH**

The Oxford Handbook *of* EXPERTISE

THE OXFORD HANDBOOK OF

EXPERTISE

Edited by
PAUL WARD,
JAN MAARTEN SCHRAAGEN,
JULIE GORE, *and* EMILIE ROTH

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2020
Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2019950595

ISBN 978–0–19–879587–2

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Summary Table of Contents

Table of Contents

SECTION TWO: METHODS TO STUDY, TEST, ANALYSE, AND REPRESENT EXPERTISE

Section Editor: Emilie Roth

SECTION THREE: DOMAINS AND APPLICATIONS

Section Editor: Julie Gore

SECTION FOUR: DEVELOPING, ACCELERATING, AND PRESERVING EXPERTISE

Section Editor: Paul Ward

SECTION FIVE: CURRENT ISSUES AND THE FUTURE OF EXPERTISE RESEARCH

Section Editor: Paul Ward

List of Contributors

Shilo Anders, Center for Research and Innovation in Systems Safety, Vanderbilt University Medical Center, USA

Güler Arsal, Envision Research Institute, Envision Inc., USA

Winfred Arthur, Jr., Department of Psychological and Brain Sciences, Texas A&M University, USA

Jaime Auton, Centre for Elite Performance, Expertise, and Training, Department of Psychology, Macquarie University, Australia

Renato F. L. Azevedo, Department of Educational Psychology, University of Illinois at Urbana-Champaign, USA

Chris Baber, School of Computer Science, University of Birmingham, UK

Johan Bergström, Lund University, Sweden

Laura Boulton, School of Forensic and Applied Sciences, University of Central Lancashire, UK

David P. Broadbent, Centre for Cognitive Neuroscience, Division of Sport, Health and Exercise Sciences, Brunel University London, UK

Katerina Bohle Carbonell, Department of Marketing and Supply Chain Management, School of Business and Economics, Maastricht University, The Netherlands

Ashleigh Brady, Centre for Cognitive Work and Safety Analysis, Joint and Operations Analysis Division, Defence Science and Technology Group, Australia

Alexander P. Burgoyne, Department of Psychology, Michigan State University, USA

Catherine M. Burns, Department of Systems Design Engineering, University of Waterloo, Canada

Michael D. Byrne, Departments of Psychological Science and Computer Science, Rice University, USA

Cui Cheng, Michigan State University, USA

Ian Clemente, Michigan State University, USA

Harry Collins, Centre for the Study of Knowledge, Expertise and Science, School of Social Sciences, Cardiff University, UK

Gareth E. Conway, Defence Science & Technology Laboratory; UK Ministry of Defence

Lauren Crichton, People Factor Consultants Ltd, UK

Margaret Crichton (†), People Factor Consultants Ltd, UK

Phaedra Daipha, Data, Discovery & Decision Science (D3) Department, Allstate, USA

Eric Anthony Day, Department of Psychology, University of Oklahoma, USA

Örjan de Manzano, Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden

Frederic Dehais, ISAE-SUPAERO, Université de Toulouse, France

Sidney Dekker, Griffith University, Australia; and Delft University, the Netherlands

Lia A. DiBello, Workplace Technologies Research Inc., USA

Tim Dornan, School of Medicine, Dentistry and Biomedical Sciences, Queens University Belfast, UK

David W. Eccles, Department of Educational Psychology and Learning Systems, College of Education, Florida State University, USA

Robert Evans, Centre for the Study of Knowledge, Expertise and Science, School of Social Sciences, Cardiff University, UK

Bradley Fawver, Department of Health, Kinesiology, and Recreation, University of Utah, USA

Peter J. Fadde, Department of Curriculum and Instruction, Southern Illinois University, USA

David. F. Feldon, Utah State University, USA

Paul J. Feltovich, Institute for Human & Machine Cognition, USA

Ute Fischer, Georgia Institute of Technology, School of Literature, Media and Communication, USA

John M. Flach, Mile Two, LLC, USA

J. D. Fletcher, Science and Technology Division, Systems and Analysis Center, Institute for Defense Analyses, USA

Lindsey N. Foreman, Department of Psychological Science and Counseling, Austin Peay State University, USA

Joana Franco, Utah State University, USA

Aric Gaunt, Michigan State University, USA

Fernand Gobet, Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, UK

Julie Gore, University of Bath, UK

David Z. Hambrick, Department of Psychology, Michigan State University, USA

Kevin R. Harris, Department of Psychological Science and Counseling, Austin Peay State University, USA

Jop Havinga, Griffith University, Australia

Robert R. Hoffman, Institute for Human Machine Cognition, USA

Ulrich Hoffrage, Faculty of Business and Economics, University of Lausanne, Lausanne, Switzerland

Ying Hu, Michigan State University, USA

Robert J. B. Hutton, Trimetis Ltd, UK; Nottingham Trent University, UK.

Mohammadreza Jalaeian, Department of Curriculum and Instruction, Southern Illinois University, USA

Michael P. Jenkins, Charles River Analytics Inc., USA

Soojeong Jeong, Utah State University, USA

Thomas G. Kannampallil, Department of Anesthesiology & Institute for Informatics, School of Medicine, Washington University in St. Louis, USA

David R. Kaufman, DR Kaufman Consulting, USA

Alex Kirlik, Department of Computer Science, University of Illinois at Urbana-Champaign, USA

Hannah Klautke, Michigan State University, USA

Gary Klein, MacroCognition LLC, USA

Kristin Knipfer, Technical University of Munich, Germany

Dennis Kowal, Science and Technology Division, Systems and Analysis Center, Institute for Defense Analyses, USA

Daphne S. LaDue, University of Oklahoma, USA

Sean Leahy, Arizona State University, USA

Jinchao Lin, Institute for Simulation and Training, University of Central Florida, USA

Gerald Matthews, Institute for Simulation and Training, University of Central Florida, USA

Karen Mattick, Centre for Research in Professional Learning, University of Exeter, UK

Laura G. Militello, Applied Decision Science, LLC, USA

Jennifer Mishra, Kent State University, USA

Scott Moffat, People Factor Consultants Ltd, UK

Brian Moon, Perigean Technologies LLC, USA

David T. Moore (†), US Department of Defense, USA

Kathleen Mosier, TeamScape LLC, USA

Miriam A. Mosing, Department of Neuroscience and Department of Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Dan Morrow, Department of Educational Psychology, University of Illinois at Urbana-Champaign, USA

Shane T. Mueller, Department of Cognitive and Learning Sciences, Michigan Technological University, USA

Jordan Multer, Volpe National Transportation Systems Center, USA

Colm P. Murphy, Faculty of Sport, Health and Applied Science, St Mary's University, UK

Neelam Naikar, Centre for Cognitive Work and Safety Analysis, Joint and Operations Analysis Division, Defence Science and Technology Group, Australia

Anjum Naweed, Appleton Institute for Behavioural Science, Central Queensland University, Australia

Frederick L. Oswald, Department of Psychological Sciences, Rice University, USA

Kai-Philip Otte, Kiel University, Germany

Vimla L. Patel, Center for Cognitive Studies in Medicine and Public Health, The New York Academy of Medicine, USA

Erich Petushek, Department of Cognitive and Learning Sciences, Michigan Technological University, USA

Jonathan D. Pfautz, Department of Defense, USA

Steven I. Pfeiffer, Florida State University, USA

Jennifer K. Phillips, Cognitive Performance Group, USA

Rebecca M. Pliske, Dominican University, USA

Andrew Rae, Griffith University, Australia

Gemma J. M. Read, Centre for Human Factors and Sociotechnical Systems, University of the Sunshine Coast, Australia

Lauren B. Resnick, University of Pittsburgh, USA

Karol G. Ross, Cognitive Performance Group, USA

Emilie M. Roth, Roth Cognitive Engineering, USA

Jennifer Lin Russell, University of Pittsburgh, USA

Paul M. Salmon, Centre for Human Factors and Sociotechnical Systems, University of the Sunshine Coast, Australia

Faith Schantz, professional writer/independent scholar

Michaéla Schippers, Rotterdam School of Management, Department of Technology & Operations Management, Erasmus University Rotterdam, The Netherlands

Jan Maarten Schraagen, TNO, The Netherlands and University of Twente, The Netherlands

Ben Shneiderman, University of Maryland, USA

Rand J. Spiro, Michigan State University, USA

Neville A. Stanton, Human Factors Engineering, Transportation Research Group, University of Southampton, UK

Joel Suss, Department of Psychology, Wichita State University, USA

Melanie Taylor, Centre for Elite Performance, Expertise, and Training, Department of Psychology **Macquarie University**, Australia

Robert Thomson, Army Cyber Institute, United States Military Academy, USA

Fredrik Ullén, Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden

Mark Upton, myfastestmile c.i.c., UK

Jeroen J. G. van Merrienboer, Department of Educational Research and Development, Faculty of Health, Medicine, and Life Sciences, Maastricht University, The Netherlands

Fred A. Voorhorst, Oyster Lab by Alpiq, Switzerland

Guy H. Walker, Centre for Sustainable Road Freight, Heriot-Watt University, Edinburgh, UK.

Paul Ward, The MITRE Corporation, USA, and Michigan Technological University, USA

Robert L. Wears (†), University of Florida, USA

James Whyte IV, College of Nursing, The Florida State University, USA

Christopher D. Wickens, Department of Psychology, Colorado State University, USA

Mark Wiggins, Centre for Elite Performance, Expertise, and Training, Department of Psychology, Macquarie University, Australia

A. Mark Williams, University of Utah, USA

Kyle Wilson, University of Huddersfield, UK

Ryan W. Wohleber, Institute for Simulation and Training, University of Central Florida, USA

William Douglas Woody, University of Northern Colorado, USA

Sarah Yardley, Marie Curie Palliative Care Research Department, University College London & Central and North West London NHS Foundation Trust, London, UK

CHAPTER 1

AN INTRODUCTION TO THE *HANDBOOK*, COMMUNITIES OF PRACTICE, AND DEFINITIONS OF EXPERTISE

PAUL WARD, JAN MAARTEN SCHRAAGEN, JULIE GORE, AND EMILIE ROTH

Introduction

The study of expertise weaves its way through various communities of practice, across disciplines, and over millennia. Arguably, Aristotle's writings on the habitual acquisition of virtues were an important impetus for the field as one virtue was *excellence*. Ever since, the polarizing nature versus nurture debates have held center stage for researchers focused on identifying the catalyst for expertise. While this debate still rages on (e.g., *Outliers; The Talent Code; The Sports Gene*), and is frequently popularized by the media, over the past century the study of expertise has matured into a multidisciplinary field spanning scientific disciplines, such as psychology, engineering, computer science, and education.

To date, the study of expertise has been primarily concerned with how human beings perform at a superior level in complex environments and sociotechnical systems, and at the highest levels of proficiency. Early expertise researchers focused on relatively simple

Historical Development of the Study of Expertise and Communities of Practice; and an Outline of the Handbook by Paul Ward, Jan Maarten Schraagen, Julie Gore, Emilie Roth, Gary Klein, & Robert Hoffman

The author (Paul Ward's) affiliation with The MITRE Corporation is provided for identification purposes only, and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions or viewpoints expressed by the author.

tasks suited to college-level introductory courses in such fields as mechanics and mathematics, and on identifying underlying causal mechanisms—often with a special focus on the cognitive functions, processes, and requirements for operating at an expert level (e.g., Chi, Feltovich, & Glaser, 1981). Others have focused on the pathways leading to the attainment of expertise (e.g., Bloom, 1985; Ericsson, Krampe, & Tesch Römer, 1993). The more recent research has continued the search for better descriptions, and causal mechanisms that explain the complexities of expertise *in context*, with a view to translating this understanding into useful predictions and interventions capable of improving the performance of human systems as efficiently as possible.

Research on expertise is both topical and timely. As the nature of work in today's society becomes increasingly cognitive, and technological advances continue to accelerate the nature of work, the need for proficient and expert workers, and the ability to rapidly acquire skill, has increased in urgency. This is true not just for key professions that shape future society—including, industry, government, military, and healthcare—but for all walks of public and personal life where learning is central to daily activities.

The purpose of this *Handbook* is to provide a comprehensive representation of the development of this field of study. As such, we offer traditional and contemporary perspectives, and importantly, a multidiscipline-multimethod view of the state-of-the-science and -engineering research on expertise. Our aim is to present different perspectives, theories, and methods of conducting expertise research that have been influential in improving our current understanding of expert phenomena across a range of domains. Our second aim is to provide a particular focus on conveying how that understanding has been applied to address *practical* problems and societal challenges. In particular, the *Handbook* focuses on how this understanding has been translated into reliable predictions useful to society, and into effective interventions that can improve all facets of proficiency, performance, security, safety, health, and well-being.

Overview of this Chapter and Structure of the Handbook

In this chapter we provide a brief summary of the various communities of practice that have paved the way for current expertise researchers, and are formative of this *Handbook*. We then provide a synopsis of how expertise has been defined both historically and in present day. Our purpose in this chapter is threefold: To demonstrate the heterogeneity of approaches and conceptions of expertise, to contextualize current views of expertise presented in this *Handbook*, and to use these views as a springboard to examine how we should examine expertise in the future—which we address in the final chapter. Finally, we present an outline of the chapters that are presented in this *Handbook*. First, we provide a brief overview of the structure of the *Handbook*.

This *Handbook* is organized into five sections. Section I, edited by Jan Maarten Schraagen, presents frameworks, theories, and models that characterize current views of expertise. In line with the goals of the *Handbook*, multidisciplinary perspectives that range from early work in cognitive psychology to more recent work, for instance, in cognitive systems engineering are presented.

Section II, edited by Emilie Roth, presents a variety of methods, developed by researchers from different theoretical perspectives, that are vital to advancing our understanding of expert phenomena. These methods are sampled from the full range of methods used to study, test, analyze, and represent expertise and, employed collectively, would represent a truly mixed methods approach.

Section III, edited by Julie Gore, presents a diversity of application domains that offer several insights in to the nature of expertise when it matters: when the "rubber meets the road." Chapters that focus on advancing our knowledge in professions, traditional skill domains, decision making under uncertainty in naturalistic decision making (NDM) research areas, and emerging domains are presented.

Section IV, edited by Paul Ward, presents chapters with the central theme of developing, accelerating, or preserving expertise. Some of the chapters present a review of the current means by which expertise can be achieved and the constraints and barriers to maintaining expertise, whereas others present a challenge to the expertise community to begin to think differently about how we develop and support expertise. In Section V, we conclude with two chapters that examine topical issues of the value and future of expertise. We now turn to setting the scene by providing an overview of the communities of practice that have laid the groundwork for, or been most active in, pursuing studies of expertise.

Past and Current Communities of Practice and Definitions of Expertise

The history of expertise is pluralistic and multifaceted. Numerous researchers have documented our and related histories elsewhere (e.g., Amirault, & Branson, 2006; Feltovich, Prietula, & Ericsson, 2006; Hoffman & Deffenbacher, 1992; Hoffman & Militello, 2009; Klein, 2008; Chapter 17, "A Historical Perspective on Introspection," by Ward, Wilson, Suss, Woody, & Hoffman, this volume) and so we will not repeat those here. Instead, we present in this section an overview of various communities of practice that have had an impact on the development of this field of study and comment on past and current definitions and conceptions of expertise.

First, to demonstrate the impact of the study of expertise, we note that several books have already been published that cover some of the topics presented in this *Handbook*. For instance, some have reviewed expertise theories, methods, and applications, albeit with a specific or narrow focus (e.g., professional expertise—Ericsson, 2009; expertise in organizations—Wiggins & Loveday, 2015). Others have made recommendations about adopting a particular perspective on what constitutes a science of expertise (e.g., Ericsson & Smith, 1991). Another has provided a specific tribute to a particular individual (William Chase), representing the influence of this work on the field (e.g., expertise and skill acquisition—Staszewski, 2013). Yet others have presented research from a sole or specific theoretical perspective (e.g., naturalistic decision making—Hoffman, 2007; Mosier & Fischer, 2015; sociology—Mieg, 2001; philosophy—Selinger & Crease, 2006). Numerous edited texts that have examined expertise in specific application domains also exist (e.g., sport—Baker & Farrow, 2015; security—Berling & Bueger, 2015; technology—Hoc, Cacciabue, & Hollnagel, 1995; nursing—Benner, Tanner, & Chesla, 1996; science & policy—Collins & Evans, 2007). Only one other *Handbook* on expertise exists (see Ericsson, Hoffman, Kozbelt, & Williams, 2018). We think that our *Handbook* offers something different—it offers a range of perspectives and is more comprehensive in its scope, and we hope that it is more inclusive in its approach than previous publications. Moreover, in this *Handbook* we have an eye on both the past *and* the future. Many authors in our *Handbook* address questions about how future research and practice should and/or may be carried out in order to continue to make the scientific and engineering leaps needed to shape future intellectual discussions and societal contributions of this field.

Communities of Practice

When we started this *Handbook* project we asked chapter authors to send us some words describing the people who had influenced their careers, especially those that had helped position each of them as world-leading researchers in the study of expertise. What followed was a list of names constituting PhD and post-doctoral advisors, mentors, authors of scholarly works, and practitioners. The names provided did not always reflect an academic lineage—which was, in part, our original goal: to map this *Handbook* to the original pioneers of the study of expertise. When genealogical maps were provided, predecessors were not always readily identifiable as someone who one might think of as a pioneer of expertise research. Instead, predecessors and influential figures were pioneers of science, engineering, psychology, sociology, and philosophy—including theoreticians, philosophers, methodologists, practitioners, and stakeholders—that reflected multiple, often disparate, communities of practice in which we operate. Our history is much richer than just a handful of notable individuals.

Next, we describe some of the communities of practice that reflect or have emerged from these influences, and that represent the majority of perspectives, theories, methods, and applications presented in this *Handbook*. Many of these communities overlap with other communities (e.g., numerous NDM researchers use verbal reporting methods; perceptual-motor expertise researchers also conduct deliberate practice research), and some have been largely superseded by other communities or other areas of research (e.g., expert systems by AI). Although some of the early research occurred around the turn of the twentieth century, the 1970s and 1980s were a particularly busy time for the emergence of many of these communities.

Verbal Reports of Thinking Community of Practice

The verbal reports community is arguably one of the earliest communities of practice relevant to the study of expertise. The methods that unite this community have their roots in introspection—the history of which as it pertains to the study of expertise is presented in detail by Ward, Wilson, et al. (this volume). Hence, we refer the reader to that chapter for more information.

Despite Brentano's (1874/2009) observation that "outstanding people" (p. 26) are an important focus for introspective studies of psychological laws (see Ward, Wilson, et al., this volume), the history of these methods is far longer than their application to the study of expertise. De Groot (1946/1965) was arguably the first person to seriously study chess experts using methods employed by this community. However, and arguably, following Duncker (1935; who leveraged Selz's (e.g., 1924/1981) work before him), De Groot is also the source of a division in this community. On the one hand, there are those who employ verbal report methods almost exclusively as a process tracing methodology (e.g., Ericsson & Simon, 1980, 1993)—so-called think aloud reports. On the other, there are those who also employ knowledge elicitation methods to better understand expertise more broadly, as well as to represent expertise so that process-traced verbalizations can be interpreted in context (e.g., Crandall, Klein, & Hoffman, 2006; Hoffman & Militello, 2009). Generally speaking, this division also reflects those who have studied simpler tasks (e.g., chess) and those that have studied more complex ones (e.g., firefighting), respectively.

Regardless of the theoretical, methodological, and practical approach adopted, as a whole this community has given rise to several treatises that have paved the way for this community of practice to employ these methods to better understand expertise more systematically, and with greater rigor. Moreover, these methods have also played a large part in developing theories of expertise (e.g., long-term working memory theory—Ericsson & Kintsch, 1995; recognition-primed decision making—Klein, 1989). Various chapters in this *Handbook* describe these methods, their products, and/or their application to the study of expertise (e.g., Chapter 13, "Representative Test and Task Development and Simulated Task Environments," by Harris, Foreman, & Eccles; Chapter 9, "Macrocognitive Models of Expertise," by Hutton; Chapter 19, "Incident-Based Methods for Studying Expertise," by Militello & Anders; Ward, Wilson, et al.).

Skill Acquisition Community of Practice

This community of practice emerged from early research on learning (e.g., Ebbinghaus, 1885/1964) and transfer of training (e.g., Thorndike & Woodworth, 1901) where, amongst other things, researchers were interested in the *savings* from engaging in prior learning when relearning the same task, and the degree of transfer from having previously learned a similar task. One of the landmark studies with respect to expertise was conducted by Bryan and Harter (1897), who studied the acquisition of telegraphic language in field operators. Their research suggested that some skills were acquired more rapidly than others (i.e., sending rate was acquired faster than receiving rate), and habits were hypothesized to be acquired hierarchically—i.e., letters, then words, then clauses, then sentences, etc.—suggesting an automation of some aspects of the task before others are acquired. This phasing of skill acquisition is, perhaps, most representative of this community of practice, and exemplified in the framework proposed by Paul Fitts (e.g., 1964). Fitts argued that skill acquisition progresses through phases, from cognitively demanding to more direct (i.e., less verbally mediated) associations between stimulus and response that are typically accompanied with decreases in error rates and time to respond. The final phase, in contrast, is assumed to be more characteristic of automatic behavior and not to be consciously mediated. This is a point of departure for some communities of practice, who speculate that experts actively defer automating skill in order to maintain conscious control and/or access to underlying representational structures, especially those involving higher-order thinking (e.g., see Ericsson & Ward, 2007).

Many parallel lines of skill acquisition research emerged around the same time, including research focused on contextual interference (e.g., Battig, 1966), the power law of learning (e.g., Crossman, 1959; Newell & Rosenbloom, 1981; Snoddy, 1926), and reasoning more broadly (e.g., Bartlett, 1958). Some of this research has been instantiated subsequently as computational models of skill acquisition, with some emphasis on expertise and expert learning. Examples of approaches emanating (jointly) from this community of practice (and others) can be found in this *Handbook* (e.g., see Chapter 46, "Skill Decay," by Arthur & Day; Chapter 2, "The Classic Expertise Approach and its Evolution," by Gobet; Chapter 15, "Computational Models of Expertise," by Kirlik & Byrne).

Individual Differences Community of Practice

The individual differences community has its roots in early efforts (pre-WWI) to develop tests capable of measuring verbal, non-verbal, and performance-based tests of intelligence and mental ability, as well as physical and motor proficiency. The goal was often pragmatic—to be able to classify individual differences in various types of ability, to diagnose intellectual or performance disabilities, and to predict future accomplishment from current levels. Although many of these efforts would have been designed to differentiate "the best from the rest" on a range of ability measures, a landmark study in the area of individual differences as they pertain to the development of expertise was conducted by Ackerman (1988). Ackerman (1988) showed that different abilities (e.g., general intelligence, perceptual-motor speed) changed in their degree of contribution to skill development as individuals progressed through different phases of learning. As individuals become more skilled with training, ability measures tended to do less well at predicting *job* performance than those related to skill or knowledge.

More recently, however, there has been a resurgence in interest in the role of individual differences in expertise. Where previously discounted, researchers have recently asserted that some of the variance in skill is explained by domain-general factors, challenging the widely held conception that 10,000 hours of deliberate practice is sufficient to attain expertise. This community has been active in the past decade in their attempts to redress this balance (e.g., see Chapter 3, "Domain-General Models of Expertise," by Hambrick, Burgoyne, & Oswald; and Chapter 6, "Neural Mechanisms of Expertise," by Ullén, de Manzano, & Mosing, both this volume).

The Knowledge and Classic Expertise Approaches Communities of Practice

Two interrelated communities, often treated as one, emerged in the 1970s: one at the Learning Research and Development Centre (LRDC) at the University of Pittsburgh, and one at Carnegie Mellon University (CMU). Arguably, these communities of practice had the earliest concentration of effort on the experimental study of expertise, and provided specific methodologies that could be used by others to study this field. These lines of research clearly motivated those that followed to test the ideas put forward, but often at the cost of omitting other emerging frameworks that used different methods and/or had alternative theoretical orientations (e.g., Rasmussen, 1983, 1986).

Although equally known for his work in standardized testing, Robert Glaser, the first director of the LRDC, was instrumental in establishing a line of research focused on how people learn and, especially, how they progress within a particular field of expertise from novice to expert. This work led to a range of experimental and verbal report-based research (e.g., see Chi, Glaser, & Farr, 1988), including some citation classics (e.g., Chi, Feltovich, & Glaser, 1981) that examined the structure of *knowledge*, and the associated cognitive strategies and memory skills that support the transition to expertise.

Arguably, the expertise work at CMU was made possible by Herbert Simon and William Chase, and Allen Newell (e.g., Newell & Simon, 1972), which built on work by Adrian de Groot (1965; which in turn built on research by Karl Duncker and, indirectly, Otto Selz; see Ward, Wilson, et al., this volume). The *classic expertise approach* adopted by the CMU researchers is covered in some detail in the chapter by Gobet (this volume); hence, it will not be repeated here. Suffice to say, that this work, amongst many other concepts, led to the concept of *chunking*, a mechanism used by chess experts to circumvent the limits of short-term memory by grouping individual units of information (e.g., chess pieces) in to constellations of information (e.g., patterns of chess pieces). Skilled memory theory followed (e.g., see Ericsson & Staszewski, 1989), as did the concept of a retrieval structure to elaborate on the explanation of chunk storage in long-term memory, and subsequent derivations of this theory (e.g., Ericsson & Kintsch, 1995; Gobet & Simon, 1996). This community also explored differences between experts and novices in problem understanding, problem-solving strategies (e.g., forward-chaining versus backward-chaining), and differences in problem representations. As such, this work foreshadowed some of the later work on recognition-primed decision making (see Schraagen, 2018, for a discussion of the relations between these lines of work).

The work at LRDC and CMU has been extraordinarily influential, yet has tended to focus on providing computational and/or theoretical explanations of simpler domains of expertise, and of the study of single tasks. Various chapters in this *Handbook* have adopted one of the perspectives that emerged from the LRDC or CMU, or describe work that is consistent with it (e.g., Gobet; Harris et al.; Kirlik & Byrne; Chapter 39, "Expertise for the Future," by Resnick, Russell, & Schantz).

Dreyfus-ian Community of Practice

The Dreyfus-ian community emerged in the 1980s, in part, in response to the expert systems and computational model-based approaches to studying cognition, including those associated with the classic expertise approach. Rather than propose a model of expertise per se, Hubert Dreyfus extended traditional phase-based theories produced by the skill acquisition community by proposing particular stages of expertise development (see Dreyfus & Dreyfus, 1986). Learners were assumed to progress from performance that was verbal-, rule-, and problem-solving-based, to a form of thinking that was based on experience—where experts recognize important and relevant aspects of the situation and know intuitively what to do without any need to engage in effortful and deliberative thinking or problem solving. This approach has been most notably advocated in the domain of nursing (e.g., see Benner, 1984; Benner, Tanner, & Chesla, 1996) but its influence has been felt in other areas too (e.g., see Chapter 11, "Is Expertise All in the Mind?", by Baber, this volume; Chapter 40, "Learning with Zeal," by Fadde & Jalaeian, this volume; Chapter 14, "Developing Mastery Models to Support the Acquisition and Assessment of Expertise," by Ross & Phillips, this volume) (cf. Ericsson, Whyte, & Ward, 2007; Gobet & Chassy, 2008).

Social Studies of Science Community of Practice

Somewhat akin to the Dreyfus-ian community, the social studies of science community reacted to the expert systems and artificial intelligence communities' failure to consider the situated nature of human knowledge. Specifically, Harry Collins (e.g., 1990, 2018; Collins & Kusch, 1998) argued that these researchers had neglected the fact that humans and computers were socially embedded, which is needed for intelligence to emerge—because knowledge and understanding are societally determined (for a summary of this argument, see Hoffman & Militello, 2009). The strongest proponents of this view would argue that knowledge, including scientific facts, is socially constructed. For instance, rather than physical laws being universal, this community of practice might argue that they are a reflection of the received view of the world at a particular moment in time that is culturally and contextually specific, rather than a truly universal phenomenon. These views are captured in current definitions of expertise that have been forwarded by this community, such as contributory and interactional expertise—respectively, they reflect the practical competence within a domain and "the ability to master the language of a specialist domain in the absence of practical competence" (Collins & Evans, 2007, p. 14). Approaches that tap into the social nature of expertise are reviewed in multiple chapters within the *Handbook* (e.g., Chapter 4, "Studies of Expertise and Experience," by Collins & Evans; Chapter 18, "Close-to-Practice Qualitative Research Methods," by Yardley, Mattick, & Dornan).

The Deliberate Practice Community of Practice

The deliberate practice community emerged, initially, with the publication of a book by Bloom (1985) that contained several chapters examining the support mechanisms leveraged, and developmental pathways taken by current experts from a range of domains. In 1993, Ericsson et al. published a seminal paper on deliberate practice, which they defined as those solitary practice activities designed by a coach specifically to improve performance, that require substantial effort, and that are not inherently enjoyable. Ericsson et al. (1993) claimed that engagement in these activities is related monotonically to the level of expertise attained, and that the greatest improvements in performance are likely to be associated with the largest weekly amounts of deliberate practice. Therefore, individuals who have accumulated the largest number of practice hours throughout their career and consistently and deliberately engaged in high levels of practice for sustainable periods are more likely to attain expertise. The general rule of thumb to attain expert status reiterated in this research is 10 years or 10,000 hours of deliberate practice (see also Hayes, 1985; Simon & Gilmartin, 1973).

The concept of deliberate practice has permeated the expertise literature in the past quarter of a century and the influence and application of this community has been demonstrated in fields as diverse as soccer (Ward, Hodges, Starkes, & Williams, 2007), clinical psychology (Rosenberg, 2000), teacher education (Dunn & Schriner, 1999), and insurance agents (Sonnetag & Kleine, 2000). Several chapters in this *Handbook* discuss this research (e.g., Chapter 25, "Musical Expertise," by Mishra; Fadde & Jalaeian). Notably, the strong claims of this research, that deliberate practice is sufficient (rather than just necessary) to explain expertise have recently been challenged (see Hambrick et al., this volume; Ullén et al., this volume).

Perceptual–Motor Expertise Community of Practice

The perceptual–motor expertise community grew out of other communities of practice, such as: the individual differences community, and their early emphasis on general and specific motor abilities; the skill acquisition community, with its focus on a shift from deliberative to automatic behavior with increases in perceptual–motor skill; and the classic expertise approach, with its emphasis on intuition as a mechanism of perceptual automation. Arguably, the latter approach had most influence with early sports expertise researchers attempting to test chunking theory by building on Chase and Simon's recall studies using sport-specific stimuli. Poulton's (1957) work on perceptual anticipation was also influential in providing a pragmatic point of focus for researchers who realized that anticipation played a significant role in sporting expertise. Following the trend in motor control and sport psychology more broadly, a division in this community occurred with sports expertise being investigated largely from two dominant perspectives: ecological (e.g., Araujo & Kirlik, 2008) and cognitive psychology (e.g., Starkes & Allard, 1993). The work of this community led to numerous texts on expertise (e.g., Baker & Farrow, 2015; Starkes & Ericsson, 2003; Williams, Davids & Williams, 1999) and is well represented in this *Handbook* (see Fadde & Jalaeian; Harris et al.; Chapter 26, "Skilled Anticipation in Sport," by Williams, Fawyer, Broadbent, Murphy, and Ward).

Naturalistic Decision Marking and Macrocognition Community of Practice

The naturalistic decision making (NDM) and macrocognition community emerged in the mid-1980s, pioneered by Gary Klein and colleagues. The impetus for this community came from the military's desire to better understand high-stakes decisions under extreme time pressure and in dynamic, uncertain, and complex environments. The Army Research Institute for the Behavioral and Social Sciences supported several lines of NDM research during the mid-1980s, and incidents such as the 1988 *USS Vincennes* shoot-down led the US Navy to want to better understand naturalistic decisions. The first NDM conference was held in 1989 in Ohio and, over the years, the NDM community came to appreciate that its purview should not be on naturalistic decision making alone but on all aspects of cognitive activity in actual work environments (e.g., see Gore et al., 2018). Hence, NDM evolved to focus on macrocognition—or the way in which individuals cognitively adapt to complexity.

This community gave rise to the development of a host of cognitive field research-based methods (e.g., see Crandall et al., 2006), including the critical decision method (Klein, Calderwood, & MacGregor, 1989) and a range of models of expert thinking including the recognition-primed decision-making model (Klein, Calderwood, & Clinton-Cirocco, 1986; Klein, 1989), the data/frame model of sensemaking (Klein, Phillips, Rall, & Peluso, 2006), and the flexecution model of adaptive replanning (Klein, 2007a, b). Its impact has been far reaching (see Hutton, this volume) but like other communities of practice, it has not been without its critics (e.g., see Yates, 2001). These critics argue, first, that the explanations provided for expert performance by the NDM community are descriptive rather than predictive, and that these descriptions over-represent the importance of aspects of performance that are easy to verbalize; second, that the more mundane, daily decisions that people make are ignored because of its emphasis on expert performance in complex environments; third, that NDM relies on a case study approach that limits generalization to the population at large (Markman, 2018).

Cognitive Systems Engineering Community of Practice

The cognitive systems engineering (CSE) community emerged in the early 1980s, pioneered by Neville Moray, Thomas Sheridan, and Jens Rasmussen, among others. Based on some early verbal report work on electronics troubleshooting and plant diagnosis (e.g., Rasmussen & Jensen, 1974; Rasmussen, 1981), Rasmussen (1983) developed a model of cognitive control to identify the different ways in which humans at different levels of expertise exert cognitive control of performance (i.e., knowledge based, rule-based, and skill based control) and the resulting heuristic strategies and workarounds used by skilled performers. Following some of the pioneers in ecological psychology and dynamic systems theory (e.g., Bernstein, 1967; Brunswik, 1956), cognitive systems engineers who were most influenced by Rasmussen adopted an ecological and systems perspectives with subsequent models (e.g., abstraction-decomposition hierarchy) of the work environment. This community views expertise not as an individual phenomenon or a particular stage of information processing, but rather as a coupling between an expert with a problem ecology through a representation. After two seminal books on cognitive systems engineering (Rasmussen, 1986; Rasmussen, Pejtersen, & Goodstein, 1994), Kim Vicente synthesized this work into an accessible treatise detailing Rasmussen's collection of methods, known as cognitive work analysis (see Vicente, 1999). This and related perspectives, and the associated methods in particular, are detailed throughout this *Handbook* (e.g., Chapter 20, "Cognitive Work Analysis," by Burns; Chapter 8, "Expertise: A Holistic, Experience-Centered Perspective," by Flach & Voorhorst; Chapter 10, "Cognitive Systems Engineering," by Naikar & Brady). This work has led to numerous elaborations and offshoots of this perspective that are relevant to the study of expertise, including work on joint cognitive systems, applied cognitive work analysis, and resilience engineering (e.g., see Chapter 47, "Expertise and Resilience," by Havinga, Bergström, Dekker, and Rae).

The above list of communities is not exhaustive and others surely exist (e.g., ethnomethodology, expert systems, and resilience engineering communities of practice). In the next section, we examine the different ways in which expertise has been defined in this *Handbook* across communities of practice. We begin with a short discussion of the value of performance-based definitions.

Definitions of Expertise

We would need a book dedicated to this topic to record completely the range of definitions that have been used to capture what it means to be an expert. Rather than provide an exhaustive review of this literature without doing it justice, we focus on some of the common definitions of expertise and attempt to capture the variety of definitions presented in this *Handbook*. Perhaps the most widely cited definition of expertise is one of *reliably superior performance on representative tasks*—a key tenet of the expert performance approach (e.g., Ericsson & Smith, 1991). One could argue that this is one of the best approaches for studying expertise—where measurable differences in performance between those performing at different levels of proficiency are reproduced and scrutinized under standardized and controlled conditions (e.g., see Ericsson & Ward, 2007).

This definition of expertise is exemplified in Morrow and Azevedo's chapter (Chapter 45, "Acquiring and Maintaining Expertise in Aging Populations," this volume) where they describe expertise as "superior levels of performance on representative tasks" and as "performance improvements with increasing task-related experience . . . skills, knowledge, interests, and other changes that accompany improved performance" (see also, Harris et al., this volume; Williams et al, this volume). However, one could argue that this and related performance-based definitions, while popular, do not easily apply in the study of cognitive work in many complex domains where performance cannot be measured or simulated easily, or reduced to single tasks.

Winning a chess match, serving a tennis ace, typing without error at speed, and playing a piece of music flawlessly (or innovatively) are relatively easy tasks to simulate and standardize under controlled conditions—and this is likely one reason why much research has focused on these domains. Studying performance in these domains and on such tasks may retain the functional complexity of work without compromising its ecological validity, making what is captured during experimentation a good representation of actual performance in these domains. But this does not help us to understand expertise in the vast array of complex domains (such as many of those discussed in this *Handbook*) where performance measurement on a given task is difficult, is impossible, or does not reflect the totality of the domain expertise. This is not to say performance measurement should not be part of a definition of expertise. Perhaps, it always should be, at least, when it is possible. The real question is: What do you do (and how should one define expertise) when working in the majority of complex domains where performance measurement is particularly challenging or impractical? We address this issue in more detail in the final chapter.

How does one measure, for instance, expertise in those individuals or teams whose task is to contemplate a dilemma (i.e., where there is no right or wrong answer)? What about those areas of work that are purely knowledge based, or where the environment is highly uncertain, the problem or goals are ill-defined, or where there are high stakes? Other authors in this *Handbook* suggest a slightly revised definition, still relevant to performance, but where the emphasis is on the measurement of performance goals or the associated knowledge, skills, and/or attitudes; and the emphasis is on using a multi-measures approach as well. For instance, Matthews, Wohleber, and Lin (Chapter 22, "Stress, Skilled Performance, and Expertise: Overload and Beyond") suggest that expertise is about cognitive and/or psychomotor skill competence, where those skills are central to accomplishing performance goals. Burns (this volume) suggests the emphasis should be on the "expert's deep knowledge and experience that brings about significantly different and more effective behaviors than novices." Feldon, Jeong, and Franco (Chapter 23, "Expertise in STEM Disciplines") suggest that expertise in scientific fields is more about "mastery of the knowledge and skills capable of bringing about new knowledge that meets or exceeds current standards." Moreover, Crichton, Moffat, and Crichton (Chapter 37, "Developing Operator Expertise on Nuclear Power Production Facilities and Oil & Gas Installations") adds another dimension to a definition of expertise in the nuclear and oil & gas industries, suggesting that the ability to demonstrate and implement relevant knowledge and skills competently as well as *confidently* are what really matters.

Baber (this volume) and Ross and Phillips (this volume) both draw on Dreyfus' and Dreyfus' or Klein's conceptualizations suggesting that expertise is defined by the ability to see the signal in the noise—the critical situational elements—and intuitively generate the appropriate, or at least an effective, course of action. Ross and Phillips (this volume) go one step further suggesting that in times of uncertainty expertise is defined by the ability to mentally simulate those actions and *see* their success. Emphasizing similar elements but also highlighting the social nature of expertise, Yardley et al. (this volume) add:

> Experts need a fine-tuned moral compass and the ability *to navigate complex social situations where power is at play* as well as intellectual and psychomotor skills. They have to be tolerant of ambiguity and have a capacity to withhold action or act in the face of uncertainty, based on a fine balance of risks and benefits. [emphasis added]

Collins and Evans (this volume) operationalize expertise along these same lines, and include the role of *Individual accomplishment* of the type described by Dreyfus and Dreyfus (1986) and others. In addition, they add two social dimensions: *Exposure*—the ease with which tacit knowledge within the domain of expertise can be accessed in general and from others; and *esotericity*—the ease with which social aspects of expertise can be accessed. This is consistent with Otte, Knipfer, and Schippers' (Chapter 43, "Team Reflection: A Catalyst of Team Development and the Attainment of Expertise") definition of team expertise: "the ability to effectively leverage . . . the knowledge and expertise of all team members."

In their definition of expertise, Ross and Phillips (this volume) also allude to the expert's ability to immediately recognize changes in the scenario and to flexibly apply knowledge and experience, even when the situation is novel. Baber and Flach and Voorhorst suggest that expertise is a matter of "sensitivity to environmental constraints and opportunities." These adaptive and context-sensitive components of expertise are not new. For instance, Hoffman (e.g., see Hoffman et al., 2014) has described expertise as context-dependent choice amongst alternatives and Bohle Carbonell and van Merrienboer (this volume) describe how routine expertise should be differentiated from what Hatano and Inagaki (1984, 1986) termed adaptive expertise. They describe the former as high-level performance on representative tasks, and the latter as the same but in unfamiliar situations, where adaptivity is attributed to a deeper conceptual understanding of the fit between a specific procedural skill and a specific situation (see also Ward et al., 2018).

Conversely, Wickens and Dehais (Chapter 29, "Expertise in Aviation") differentiate the type of (routine) *expert* performance of real experts from that which is more representative of competent journeymen—"the ability to successfully perform job-relevant duties and solve common problems quickly, reliably, and accurately." They suggest that, by definition, expertise conveys an "ability to successfully solve uncommon, unusually difficult, and/or strategic problems that others cannot." This perspective is consistent with Hoffman's (1998) delineation of proficiency across the skill continuum, which was amongst the first to present a proficiency scale—based on the craft guilds of the Middle Ages—that captured this adaptive feature within a definition of expertise.

Despite the limited evidence for positive skill transfer with expertise, the flexible and adaptive nature of expertise is emphasized by several authors in this *Handbook*. One view which exemplifies many is that of Resnick et al. (this volume), who suggest that experts "draw fluidly and flexibly on the information at hand and on the complex set of skills and attitudes (including the willingness to change one's mind) that comprise reasoning." We will revisit the concept of adaptive skill in the final chapter of the *Handbook*. In the next section of this chapter, we provide a brief summary of the chapters within each section of the Handbook.

Outline of the *Oxford Handbook of Expertise*

Characterizing Expertise: Frameworks, Theories, and Models

Section I provides an overview of frameworks, theories, and models used to characterize expertise. The chapters range from the classic approach to expertise, as exemplified by Chase and Simon's research on chess expertise in the early 1970s, to more recent approaches focusing on macrocognition and cognitive systems engineering. Although the study of expertise has been dominated by research in cognitive psychology, and most chapters build upon this tradition, this section also includes chapters from a sociological and neural point of view. A brief overview of each of the chapters in this section is presented next.

In Chapter 2, Gobet describes the classic expertise approach and its evolution. He starts off by briefly discussing early research on expertise that influenced the classic approach. He then describes in some detail Chase and Simon's classic papers and chunking theory. This leads the way to a presentation of some of the key experimental and theoretical research that was characterized by detailed analyses of the cognitive processes involved, use of verbal protocols, and a small number of participants. The chapter then discusses more recent theories that can be considered as outgrowths of the classic approach, providing a good opportunity to try to understand not only its key characteristic but also why it had such a large impact. The chapter concludes by a discussion of what this approach tells us about the means to address the challenges currently facing research on expertise.

In Chapter 3, Hambrick, Burgoyne, and Oswald review evidence concerning the contribution of cognitive ability to individual differences in expertise. Their review covers research in traditional domains for expertise research such as music, sports, and chess, as well as research from industrial-organizational psychology on job performance. The specific question that they seek to address is whether domain-general measures of cognitive ability (e.g., IQ, working memory capacity, executive functioning, processing speed) predict individual differences in domain-relevant performance, beyond beginning levels of skill. The authors note that evidence from the expertise literature relevant to this question is difficult to interpret, due to small sample sizes, restriction of range, and other methodological limitations. By contrast, there is a wealth of consistent evidence that cognitive ability is an important and statistically significant predictor of job performance, even after extensive job experience. The authors discuss ways that cognitive ability measures might be used in efforts to accelerate the acquisition of expertise.

Chapter 4 provides a sociological perspective on expertise. Collins and Evans build upon a research program they refer to as studies of expertise and experience (SEE), often referred to as the *third wave of science studies*, which treats expertise as real and as the property of social groups. This chapter explains the foundations of SEE and sets out the theoretical and methodological innovations created using this approach. These include the development of a new classification of expertise, which identifies a new kind of expertise called *interactional expertise*, and the creation of a new research method known as the imitation game designed to explore the content and distribution of interactional expertise. The authors conclude by showing how SEE illuminates a number of contemporary issues such as the challenges of interdisciplinary working and the role of experts in a *post-truth* society.

In Chapter 5, Pfeiffer discusses giftedness and talent development in children and youth, with a focus on talent development as a path toward expertise and eminence. The chapter briefly discusses a history of gifted education and then tackles some big picture issues and future possibilities. The chapter addresses a number of questions, including: Who is gifted? How are gifted individuals identified? Is giftedness domain-specific or domain-general? How malleable is giftedness? Does giftedness represent a qualitative or quantitative difference? How does the concept of expertise fit into gifted education? The author proposes a tripartite model of giftedness that offers three distinct lenses through which high-ability students can be viewed: (1) high intelligence; (2) outstanding accomplishments, and (3) potential to excel.

In chapter 6, Ullén, de Manzano, and Mosing provide an overview of some of the neuroanatomical and functional correlates of expertise, concluding that expertise is related to macroanatomical properties of domain-relevant brain regions and ultra-structural properties of both the gray and white matter. The consequence of these neural adaptations is a capacity for vastly more efficient performance of domain-specific tasks. In functional terms, this depends on multiple mechanisms that are situated at different levels of neural processing. These mechanisms include automation and alterations in functional connectivity, as well as specializations within memory systems and sensorimotor systems that optimize the processing of information which is relevant for the particular domain of expertise. The author concludes with a discussion of neural mechanisms of expertise from the perspective of new models that emphasize a multifactorial perspective and take into account both genetic and environmental influences on expertise and its acquisition.

In Chapter 7, Hoffrage provides an overview of the *fast-and-frugal heuristics* program of research on expertise. According to the program reviewed in this chapter, people—including experts—use fast-and-frugal heuristics. These heuristics are models of bounded rationality that function well under limited knowledge, memory, and computational capacities. These heuristics are ecologically rational: they are fitted to the structure of information in the environment. While studying experts in the context of this program amounts to modeling them with fast-and-frugal heuristics, studying the acquisition of expertise focuses on how laypeople learn such heuristics. Because fast-and-frugal heuristics do not require complex calculation and are typically easy to set up, this program offers a straightforward way to aid experts: After the heuristics' performance has been determined under various environmental conditions, experts can be educated about these results.

In Chapter 8, Flach and Voorhorst advance the claim that expertise is not a property of any particular stage of information processing, nor is it a property of an individual. Rather it is the property of a triadic semiotic system where the quality of performance depends on the coupling of an agent with a problem ecology through a representation. The dynamics of this coupling is akin to a self-organizing, adaptive control system. The authors argue that many of the debates about the nature of expertise arise from the different ways that people have parsed the triadic system into subcomponents (elements or dyads). Thus, the fundamental point of this chapter is that expertise is not something that can be isolated as a property of a mind, independent from a problem ecology, or vice versa.

In Chapter 9, Hutton provides an overview of so-called macrocognitive models of expertise, of which three are discussed in some detail: The recognition-primed decision model, the data-frame model of sensemaking, and the flexecution model of replanning and adaptation. Macrocognitive models are models of experienced, often expert performers and have been developed primarily from the study of decision making and cognitive work in naturalistic settings, as opposed to well-controlled laboratory experiments. They describe how people manage uncertainty and complexity in the world of work. The limitations and applications of these models are also illustrated in order to provide a future-oriented perspective on how the models might be improved and how they might be applied to support more effective cognitive work and more resilient work systems.

In Chapter 10, Naikar and Brady present a perspective of human expertise in sociotechnical systems based on the phenomenon of self-organization. Consistent with the ideals of the field of cognitive systems engineering, this perspective is based on empirical observations of how work is achieved in complex settings and incorporates an emphasis on design. The proposed perspective is motivated by the observation that workers in sociotechnical systems adapt not just their individual behaviors, but also their collective structures, in ways that are closely fitted to the evolving circumstances, such that these systems are necessarily self-organizing, a phenomenon that is essential for dealing with complexity in the task environment. Accordingly, the chapter explores in depth the theoretical and design implications of the phenomenon of self-organization for understanding and supporting human expertise in sociotechnical systems, and draws attention to the broader implications of this phenomenon for advancing a social basis for human cognition.

In Chapter 11, Baber reviews theories that explore the relationship between action and performance. These theories ask whether our cognitive activity depends on *internal representations* or whether it can be explained by our interaction with the world around us. In other words, rather than projecting a model of the world outwards in order to plan and guide our actions, these approaches see physical interaction with the world as a form of cognitive activity. These theories focus less on using mental representation and more on perception–action coupling between us and our world. Baber concludes that this points to an account of expertise which sees it as a matter of sensitivity to environmental constraints and opportunities, together with the ability to focus on optimal parameters in a given situation. From a practical point of view, he considers ways in which such sensitivity could be probed through field study and interview with experts.

This section concludes with Chapter 12 on adaptive expertise by Bohle Carbonell and van Merrienboer. The authors start by noting that the increasing number of changes at the workplace created through automation, political upheavals, and new technology frequently exposes individuals to unfamiliar situations. According to the chapter authors, mastering these situations requires individuals to possess adaptive expertise. By being an adaptive expert, individuals are able to deal with novel situations and remain performing at their original level. By drawing on recent literature, the goal of this chapter is to describe what adaptive expertise is. The authors contrast the concept of adaptive expertise with routine expertise to clarify when adaptive expertise produces superior performance to routine expertise. Subsequently they compare adaptive expertise to other expertise concepts. Following this, they describe how adaptive expertise can be developed and measured. The chapter ends with a number of recommendations of how individuals can be stimulated to develop adaptive expertise.

Methods to Study, Test, Analyze, and Represent Expertise

In Section II we present a range of methods that have been used to study expertise. These methods provide a sampling of techniques that draw from multiple theoretical perspectives and communities of practice. In combination they provide an excellent introduction to the variety of theoretical approaches and practical methods available for analyzing, representing, and testing the knowledge and skills associated with expertise. Several of the chapters provide practical *how to* guidance and describe pitfalls to avoid when conducting research on expertise.

In chapter 13 Harris, Foreman, and Eccles provide an introduction to designing representative tasks, tests, and simulated task environments for use in uncovering the basis of expertise. The authors describe how well-designed representative tasks can be used to discover the mechanisms that underlie superior performance, to stratify performers based on skill, and to discover the developmental steps involved in reaching superior levels of performance. The authors describe how representative tasks and simulated task environments can be used to understand the basis for expert performance as well as to develop training on the basis of expert performers. The authors present recent research illustrating the use of representative tasks and simulated task environments.

In Chapter 14, Ross and Phillips describe the origin, development, and application of a mastery model approach to the acquisition, representation, and assessment of expertise. The framework of the mastery model originates from the Dreyfus and Dreyfus general model of cognitive skill acquisition. Development of a mastery model is based on a semi-structured interview of experts and a qualitative analysis process. The model specifies the hallmarks of performance, characterizes the progression of skill, and provides performance indicators for each key area, categorized in five progressive levels. The chapter includes specific examples and discusses differences between the mastery model and competency modeling approaches.

In Chapter 15 Kirlik and Byrne provide a comprehensive introduction to computational models of expertise, reviewing both the foundational and contemporary body of research. The chapter discusses and provides examples of computational models built within the framework of a unified cognitive architecture as well as models that are more domain or task specific in their psychological assumptions. It highlights the requirements for effective computational modeling of expert behavior, including the need for extensive analysis, and possibly expert-level knowledge of both tasks and the environments in which expert behavior is manifest. The chapter ends with a discussion of promising future directions for research using computational modeling, as well as other emerging techniques such as neuroimaging, that can be combined to advance the scientific understanding of human expertise.

In Chapter 16 Salmon, Stanton, Walker, and Read discuss the use of hierarchical task analysis (HTA) to represent expert behavior and the factors influencing it. HTA is among the most widely used methods for conducting task analysis and a variety of ergonomics methods build on HTA outputs to provide in-depth analyses of behavior. The chapter describes HTA and its origins, discusses its strengths and weaknesses, and provides practical guidance on how to apply the method. Two rail level crossing cases studies are used to illustrate how HTA can be employed to describe and analyze both the behavior of individuals at the sharp end (i.e., at the rail level crossing itself) and the behavior of the overall sociotechnical system (i.e., the rail level crossing *system*).

In Chapter 17 Ward, Wilson, Suss, Woody, and Hoffman provide a historical examination of the philosophical roots and long-standing controversies associated with one of the most widely used methods for understanding expertise—elicitation and analysis of verbal reports. The chapter serves as a comprehensive introduction to the variety of introspective-type methods discussing their validity and utility. It begins with a historical review of the perspectives and contributions of the pioneers of introspective methods, highlighting key motivations, arguments, and disputes that have driven methodological development over the past 100+ years. It then turns to a review of current methods that rely on *thinking aloud* and other types of verbal reports to study expertise. It offers cautions and guidance on appropriate use of verbal reports, and ends with reflections on the future of introspection methodology and opportunities to improve the state-of-the-science and escape the legacies of behaviorism.

In Chapter 18 Yardley, Mattick, and Dornan provide an introduction to qualitative research methods, particularly as part of research grounded in practice. The authors argue that expertise is inherently linked to the context in which experts work. Qualitative methods enable researchers to examine the broader context, including social practices, within which expertise is manifest and to uncover and pursue unexpected findings. The chapter describes different qualitative methods and discusses their distinct contributions to understanding expertise development in professional work. The authors provide examples to illustrate how qualitative methods can be used to answer *how* and *why* questions, to disentangle the impact of different factors in complex and uncertain situations, and to explore the messiness and complexity of expertise more generally.

In Chapter 19 Militello and Anders provide a comprehensive introduction to incident-based interview methods for discovering and characterizing expertise. Incident-based methods are among the best known and easily applied methods for understanding expertise. The chapter presents four types of incident-based interview methods: critical decision method, knowledge audit, simulation interview, and cued-retrospective interviews. It describes strategies for analyzing the outputs of incident-based interviews, and provides examples of the types of products that are generated. Among practical applications described are uncovering requirements for training; identifying requirements for support tools; and developing testable hypotheses and descriptive models to inform basic research on expertise. The chapter ends with discussion of practical issues associated with the use of incident-based interview methods and experience-based advice in how to address real-world challenges.

In Chapter 20 Burns reviews the roots of the cognitive work analysis (CWA) framework and how it can be used to represent expertise. CWA methods are best known as tools for the design of novel displays that improve performance in complex domains such as process control, health, and military operations. Displays built on CWA principles have proven to be particularly effective in supporting performance in unanticipated complex situations where expert *knowledge-based* strategies are most useful. Less improvement in performance has been found in routine situations that are well supported by conventional displays and procedures. Burns argues that this may be because CWA focuses on the needs for expert performance. She points out that CWA was originally founded as part of attempts to understand human expertise and transfer the knowledge of human experts into a design so that those who are *less expert* could benefit. As a result, CWA methods are useful for understanding and transferring expertise. Burns reviews the steps of CWA and discusses their various contributions to the understanding and development of expertise. The chapter ends with a discussion of how CWA can be used to develop and transfer expertise through design.

In Chapter 21 Moon provides an invaluable perspective on how to understand and represent expertise through reflections on his own experiences in the professional practice of knowledge capture. The chapter describes the scope, uses, and origins of knowledge capture methods. Most particularly it covers protocols and methods for knowledge elicitation, focusing on the elements common to all capture methods—structure and probing questions. The issues of cognitive burden sharing across the capturer and holder, how purpose guides execution, and how constraints can shape practical developments in approaches to knowledge capture are also discussed. Throughout, the chapter offers insightful, first-hand stories of knowledge capture and provides invaluable advice in dealing with pragmatic complexities that inevitably arise when going out *into the field*. The chapter concludes by looking at future directions for the profession.

In Chapter 22 Matthews, Wohleber, and Lin examine the complex relationship that exists between stress, skilled performance and expertise. Stress generally impairs attention and working memory, increasing vulnerability to cognitive overload. The authors present a model characterizing the interrelation between stress and expertise called the standard capacity model (SCM) derived from theories of attention resources and cognitive skills acquisition. Matthews and colleagues argue that SCM, while having some empirical support, has serious limitations including neglecting contextual factors that can alter the pattern of findings. As an illustration, the authors describe the interplay between stress and expertise across four domains: test anxiety, sports performance, surgery, and vehicle driving. Consistent with the SCM, in some cases stress is associated with cognitive overload and expertise is shown to buffer the effects of stressors. However, the authors also provide evidence that expert performance is subject to domain-specific influences beyond cognitive capacity, including strategies for emotional regulation, choking under pressure, and aggressive behaviors that mediate the relationship between stress and expertise. They conclude that the relationships between stress and expertise must be examined contextually.

Domains and Applications

Section III focuses upon the wide variety of domains and applications exploring expertise. The chapters highlight great diversity of application and provide a range of insights from disparate domains. Two chapters focus on advancing professions of business and science, technology, engineering, and mathematics (STEM) education. Three chapters focus on advancing skill in domains in which expert performance has been studied traditionally: games, music, and sports. Seven chapters focus on advancing decision making under uncertainty in traditional naturalistic decision-making (NDM) areas of research: spaceflight, nuclear, oil & gas operations, healthcare, firefighting and emergency responding, weather forecasting, aviation, railroad transportation, law enforcement, and military. Two chapters focus on decision making under uncertainty in the emerging domains of cyber security and intelligence analysis. Notably, the authors who have contributed to this section are located in academia and practice across many disciplines around the world. A brief overview of each of the chapters in this section is outlined here.

In Chapter 23, Feldon, Jeong, and Franco provide a synthesis of literature which is most relevant for enhancing expertise in STEM. This synthesis focuses upon relevant findings from cognitive psychology and the psychology of science, sociology and anthropology, and educational research. The authors present the fundamental mechanisms of thinking and problem-solving practices in science and engineering that underlie expert performance within these disciplines. The chapter also examines issues pertaining to assessment and recognition of expertise in STEM fields, and the impact of training and education. The chapter ends by suggesting that further work is needed to explore and question the nature of expertise, the dynamic nature of STEM disciplines, and interdisciplinarity.

Mueller, in Chapter 24, examines word game expertise from a cognitive perspective and proposes a general taxonomic space of word games where the primary organizing axis distinguishes letter- versus meaning-centered games. His critical review of the area concludes with the hypothesis that word game expertise is supported by both practice *and* prior skills and ability, and suggests predisposition opportunity may be a fruitful framework for understanding skilled performance in this domain.

Chapter 25, by Mishra, suggests that music is a foundational domain in the development of the theory of expertise. Similar to Mueller's conclusions, Mishra reports that current thinking is that while deliberate practice is important, it is insufficient to entirely explain expertise. Mishra conjects that future research will aim to determine how genetics and practice interact in the development of experts.

In Chapter 26, Williams, Fawver, Broadbent, Murphy, and Ward provide a historical overview of research focusing on the topic of anticipation, with a particular emphasis on its importance in various high-performance domains, including sport. They review more than five decades of research which has highlighted some of the key perceptual–cognitive skills underpinning anticipation and how these interact and vary in importance from one situation to another. In the second half of their chapter, they highlight the need for methodological improvements and identify ways in which conceptual understanding may be enhanced.

Patel, Kaufman, and Kannampallil (Chapter 27) report on the study of diagnostic expertise which initially focused on characterizing the reasoning process and, later, on understanding the nature of expert knowledge and its impact on performance, including memory, comprehension, and reasoning. This chapter highlights that medical expertise is not a simple construct, and its development is characterized by non-linear growth in skills and knowledge. Facilitated by new technology, recent research has moved toward real-world studies (or a combination of both laboratory-based and naturalistic studies), with automated and often precise methods of data collection and analysis.

Chapter 28 by Wiggins, Auton, and Taylor examines the study of expertise in the context of firefighting and emergency responding. The outcomes of existing research initiatives are examined, emphasizing the importance of accurate and precise mental models acquired through active interaction within the operational environment. Future research directions are proposed that will ensure the development of a continued comprehensive understanding of the nature expertise in firefighting and emergency responding.

Wickens and Dehais (Chapter 29) explore proficiency and make the distinction between the experience of aviation professionals, often quantified in terms of hours of flight time, or flight qualifications, and expertise: proficiency at aviation tasks. They conclude from an extensive review of the literature in this area that experience of skills such as situation awareness, decision making, task management, and crew resource management may be only loosely coupled with proficiency and explore why this may be so.

Another well-established research area is covered by Roth, Naweed, and Multer in Chapter 30, who summarize methods used to uncover expertise in railroad research, followed by a review of the types of strategies that railroad workers exhibit. By providing a discussion of the impact of ongoing technological changes on the requirements for expertise the authors speculate on longer term changes in railroad technologies and the nature of expertise.

Thomson (Chapter 31) describes the historical and continuing evolution of the cyber domains, and how we can operationalize current research in cyber expertise. Research into cyber expertise is in its infancy; in fact, there is no clear definition of what constitutes cyber expertise or how it may be unique when compared to other technical fields. Thus, Thomson describes the work roles of cyber operators and reviews results from cognitive task analyses of their workplace. Finally, topics are presented for future research, including the use of realistic synthetic environments to study cyber operations with more ecological validity.

Jenkins and Pfautz (Chapter 32) focus upon intelligence analysis (IA) and highlight past research methods that have been applied to characterize the domain of IA from a high-level workflow perspective down to low-level models of analyst information processing. They argue that such characterizations of the domain provide opportunities for highlighting different characteristics of expert behaviors throughout the IA process. They conclude with implications for the design of training and propose new types of technologies to aid in IA and analogous domains.

In Chapter 33, Suss and Bolton examine expertise in law enforcement and provide guidance for those planning on conducting research in this field. Illustrative examples cover a broad range of methods and highlight the subtleties that researchers new to the domain should consider when designing and conducting expertise research.

Fletcher and Kowal (Chapter 34) review characteristics of expertise common to all domains as a context for the expertise needed by military personnel. Cognitive qualities needed for military expertise are discussed, including the emerging issue of cognitive readiness required for irregular as well as regular military operating environments. The chapter emphasizes that military expertise is similar to expertise elsewhere; however, the volatility, uncertainty, complexity, ambiguity, and lethality of the environment in which military decisions are made that affect large numbers of individuals may be unique.

Chapter 35 illustrates DiBello's work, which shows changes in society have influenced a greater need for expertise in business. DiBello's chapter provides an exceptional insight into her work with organizations over the past two decades, helping them adapt their expertise to an increasingly complex and interconnected world of business. DiBello's chapter suggests that the new *expert* in business may not be an individual at all, but rather a high performing and highly efficient team. The chapter ends with reflections about how business organizations may learn from expertise.

Fischer and Mosier's chapter (36) on teams in space captures the complexities associated with human space flights' multiteam effort requiring the coordination and collaboration not only of individuals within a team (mission control or space crew), but importantly also between teams. The chapter discusses the strategies and procedures these expert teams have established to ensure common task and team models, and to facilitate their communication and joint performance. The teamwork challenges of future long-duration space exploration are discussed, as are the continuing advancements and research needed to address them.

Chapter 37, by Crichton, Moffat, and Crichton, describes the current status of expertise development in nuclear power production and oil & gas facilities, for both routine operations and emergency response. They note simulator-based exercises increasingly being introduced. The chapter summarizes existing research into the content and format of the skills required by operators in these settings, highlighting many questions yet to be answered, including how do we measure this combination of task, duration of experience, and level of performance to determine expertise?

The final chapter of this section by LaDue, Daipham, Pliske, and Hoffman (Chapter 38) captures the latest insights on weather forecasting. The chapter summarizes four research programs ranging from organizational to individual analyses to provide unique, complementary insights about expertise in this highly technologically focused domain. Like many areas of expertise, the forecasters have extensive, complex knowledge about each type of weather process they forecast, a knowledge that may be lost if not captured and passed on to the next generation. LaDue et al. suggest that the empirical work on professional activity in context, has the potential to invigorate studies of expertise.

Developing, Accelerating, and Preserving Expertise

Section IV presents a collection of approaches that have been used, broadly speaking, to develop or maintain expertise. The chapters range in emphasis from envisioning new pathways for educating our children and teachers to leveraging our knowledge of how experts learn and adapt in complex environments. Chapters cover important topics from improving expert performance individually and in teams, to accelerating the development of expertise, maintaining or retaining expert skills, and avoiding breakdowns in system performance.

Resnick, Russell, and Schantz (Chapter 39) argue that current methods of classroom teaching may be unfit for purpose—they do not adequately prepare students to thrive and excel in a complex world. As an alternative, they discuss the role of argumentation in developing expert reasoning skills, and point to an emerging body of research which suggests that teaching these skills in the classroom can lead to the acquisition and retention of general knowledge, beyond the topics taught through discussion. They review the ways in which *dialogic* reasoning can be used as a form of teaching to support the development of argumentation skill, and discuss some of the barriers to extending these methods beyond those students already considered gifted. In addition, they examine the role of educational, organizational, and social systems in facilitating a transition toward greater adoption of these methods in the classroom and beyond.

In Chapter 40, Fadde and Jalaeian provide an overview of the concept of deliberate practice and review the associated research in teacher education, medicine/surgery, and sports. They examine the difference between domains that have a culture of *practice*—like sports and music—versus a culture of *study* or a culture of *experience*. They highlight that while strong correlations have been found between expertise level and domains that have a culture of practice, much weaker correlations have been found in those with a culture of study or experience, such as education or professions. Accordingly, they present an alternative to deliberate practice, termed deliberate performance, which captures the kinds of learning activities in which professionals might engage to deliberately improve their performance. Last, Fadde and Jalaeian compare three models of training based on related research and review their effectiveness and conclude that consciously incorporating deliberate practice during college-based professional education and deliberate performance during the career work of professionals (who typically have little time to *practice*) can accelerate the development of professionals to expert levels.

In Chapter 41, Spiro and colleagues examine expertise in complex and ill-structured domains from the perspective of *cognitive flexibility theory* (*CFT*). Their emphasis is on *adaptation* in modern situations that deviate from novelty in relatively ordinary yet unexpected ways. Spiro et al. build on and extend the adaptive skill framework proposed by Ward et al. (2018) by further specifying how one prepares for situational novelty via meta-features of an *adaptive mindset* that generalize across cases in ways that content does not. Spiro's view also specifies how these features support the novel rearrangement of previously encountered case features in ways that are adaptive to new situations. Computer-supported case-based learning environments are used as a means to apply CFT to *expertise acceleration*, and a theoretical rationale and empirical examples are provided for structuring these computer systems in terms of the principles of case and concept selection and sequencing. *New modes of deliberate practice* that foster *adaptive readiness* are proposed, including skill at situation-adaptive assembly of knowledge and experience for *adaptive performance* that require a rethinking of what constitutes deliberate practice. Spiro and colleagues conclude with a discussion of a wide range of practical *implications* to the accelerated fostering of adaptive response to novel situations.

In Chapter 42, Moore and Hoffman present a view of proficiency scaling in the domain of intelligence analysis. They highlight the role of what are termed essential competencies, and detail the many distinct analytical roles entailing a specialization of expertise in this domain. Moreover, they discuss models of analyst reasoning and knowledge as a function of proficiency level and consider the stability of individual differences in styles, and distinctiveness of approaches to critical thinking across proficiency levels. In an era when the intelligence community is calling for robust measures of performance, they review how analysts make sense of situations and events for which there is no single cause, and discuss the role of human agency and motivations in causal reasoning. Last, they present implications of this research for training future analysts.

Otte, Knipfer, and Schippers put forward the claim, in Chapter 43, that team reflection is a major driver for the development and attainment of expertise in teams. They define team reflection as the collective evaluation of prior team activities, review the associated research, elaborate the mechanisms that link team reflection to expertise in teams and discuss multiple catalysts of team reflection. In the final part of their chapter, they investigate the shortcomings of previous team reflection research. These include the level at which this research has been analyzed and the short- and long-term consequences of engaging in this activity. They make suggestions for future research in this area to deepen our understanding of the effects of team reflection on the development of expertise in teams.

In Chapter 44, Petushek, Arsal, Ward, Upton, Whyte, and Hoffman focus on the role of mentoring in the development of expertise and discuss the multifactorial nature of mentoring, coaching, and related learning enhancement methods. They pursue a specific goal of unpacking the complex interactional relationship between developmental functions and roles to more fully describe what it means to be an effective coach/mentor. They review the meta-analytic and empirical evidence supporting the effectiveness of developmental and mentoring-type roles on a range of outcomes, such as job performance and career progression, and provide some specific examples of studies that have documented mentoring activities in action. They conclude by summarizing some of the major issues yet to be resolved in this field, and make specific suggestions for how to advance this area of research.

In Chapter 45, Morrow and Azevedo focus on the relationships between expertise and aging. They begin their chapter by considering how experts excel on domain-relevant tasks despite cognitive limitations. Further they examine how these expertise-related advantages develop, providing possible ways that adults can offset age-related cognitive constraints to maintain performance in later years. Their chapter centers around two key issues related to expertise and aging: the extent to which superior levels of performance can be maintained by experts as they continue to age; and the extent to which knowledge and skill associated with experience can offset age-related declines in abilities and function.

Arthur and Day (Chapter 46) focus on the issue of skill and knowledge decay and retention and how this intersects with expertise. Their review highlights several important conclusions: First, decay is affected more by interference than forgetting of information. Second, decay is highly dependent on task and situational factors.

Third, there is less decay on complex tasks than is observed for simple tasks. Fourth, retention is a function of expertise—it is stronger with practice, elaborative rehearsal, and greater mastery of the task. Fifth, retention and transfer are distinct concepts. Sixth, there is a limited amount of empirical research on decay and expertise in complex real-world performance domains. They conclude by suggesting that much could be learned by closer integration of the literature on expertise and skill decay and make recommendations for how to proceed in this regard.

In Chapter 47, Havinga, Bergström, Dekker, and Rae present an argument for expertise from a resilience engineering perspective, suggesting that this has changed the value of expertise from meeting required standards to helping organizations adapt. They begin by introducing the concept of resilience and its application to safety in sociotechnical systems. Then they explore how to manage expertise in complex systems, considering both its costs and benefits to engineering resilient organizations. Their review considers the role of expertise at multiple levels of the system, including frontline workers, teams, and management, and on an organizational systems level.

In Chapter 48, Conway and Gore focus on the role of expertise in developing complex policy interventions for government. They begin by providing some context on the nature of government work and its challenges; and then examine the role of expertise across the system, including the overlap in expertise between different roles, how this has evolved over time, and how the evidence is derived, from whom, and how it is shaped by values. They highlight the difficulties of identifying and incorporating true expertise into policy interventions and consider whether expertise itself is under threat in this process. Last, they identify research gaps that if addressed would support the further professionalization of the policy function in government.

Current Issues and the Future of Expertise Research

Section V presents two chapters that address the current and future challenges of expertise research. In Chapter 49, Klein, Shneiderman, Hoffman, and Wears highlight the seeming irony that although expertise is increasingly sought out and needed in today's society, at the same time several communities have actively begun to disparage experts! They present a series of arguments that demonstrate why their criticisms are misguided and assert that the criticisms made can help the research community discover better methods for supporting experts and for developing expertise.

In the final chapter, Ward, Schraagen, Gore, Roth, Hoffman, and Klein discuss some of the future directions that the field of expertise studies might take in order to continue to allow people to thrive in a world whose complexities are ever increasing. In particular, they present a particular view of expertise focused on adaptive skill—a concept that has often been discussed but is an empirically neglected aspect of expertise research—as a potential remedy for advancing the field and better preparing individuals to cope with the uncertainties and complexities of tomorrow's society.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General 117*(3), 288–318.

Amirault, R. J., & Branson, R. K. (2006). Educators and expertise: A brief history of theories and models. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds), *The Cambridge handbook of expertise and expert performance* (pp. 69–86). Cambridge, UK: Cambridge University Press.

Araujo, D., & Kirlik, A. (2008). Towards an ecological approach to visual anticipation for expert performance in sport. *International Journal of Sport Psychology 39*(2), 157–165.

Baker, J., & Farrow, D. (2015). *Routledge handbook of sport expertise*. London: Routledge.

Bartlett, F. C. (1958). *Thinking: An experimental and social study*. London: Allen & Unwin.

Battig, W. F. (1966). Facilitation and interference. In E. A. Bilodeau (Ed.), Acquisition of skill (pp. 215–244). New York: Academic Press.

Benner, P. (1984). *From novice to expert: Excellence and power in clinical nursing practice*. Upper Saddle River, NJ: Prentice-Hall Health.

Benner, P., Tanner, C. A., & Chesla, C. A. (1996). *Expertise in nursing practice: Care, clinical judgment and ethics*. New York: Springer.

Berling, T. V. & Bueger, C. (2015). *Security expertise: Practice, power, responsibility*. London: Routledge.

Bernstein, N. (1967). *The coordination and regulation of movements*. New York: Pergamon.

Bloom, B. (1985). *Developing talent in young people*. New York: Balantine.

Brentano, F. (2009). *Psychology from an empirical standpoint*. London: Taylor & Francis. (Original work published 1874.)

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd edn). Oakland, CA: University of California Press.

Bryan, W. L., & Harter, N., (1897). Studies in the physiology and psychology of the telegraphic language. *Psychological Review 4*, 27–53.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science 5*(2), 121–152.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.) (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum.

Collins, H. (1990). *Artificial experts: Social knowledge and intelligent machines*. Cambridge, MA: MIT Press.

Collins, H. (2018). *Artifictional Intelligence: Against humanity's surrender to computers*. Cambridge, MA: Polity.

Collins, H., & Evans, R. (2007). *Rethinking expertise*. Chicago: University of Chicago Press.

Collins, H., & Kusch, M. (1998). *The shape of actions: What humans and machines can do*. Cambridge, MA: MIT Press.

Crandall, B., Klein, G., & Hoffman, R. R. (2006). *Working minds*. Cambridge, MA: MIT Press-Bradford.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics 2*, 153–166.

De Groot, A. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton. (Original work published 1946.)

Dreyfus, H. L. & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: The Free Press.

Duncker, K. (1935). On problem solving. *Psychological Monographs 58*(5), 1–113. Translated by Lees, L. S. (1945; *Zur Psychologie des produktiven Denkens*).

Dunn, T. G. & Schriner, C. (1999) Deliberate practice in teaching: What teachers do for self-improvement. *Teaching and Teacher Education* 15, 631–651.

Ebbinghaus, H. (1885). *On memory: A contribution to experimental psychology* (H. Ruger and C. Busenius, Trans.). New York: Teacher's College, Columbia University. (Reprinted 1964.)

Ericsson, K. A. (2009) (Ed.). Development of professional expertise. New York: Cambridge University Press.

Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (2018) (Eds). *The Cambridge handbook of expertise and expert performance* (2nd edn). Cambridge, UK: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long term working memory theory. *Psychological Review 102*(2), 211–245.

Ericsson, K. A., Krampe, R. T. h., & Tesch-Römer, C. (1993) The role of deliberate practice in the acquisition of expert performance. *Psychological Review 100*, 363–406.

Ericsson, K. A., & Simon, H. (1980). Verbal reports as data. *Psychological Review 87*(3), 215–251.

Ericsson, K. A., & Simon, H. (1993). *Protocol analysis: Verbal reports as data* (rev. edn). Cambridge, MA: MIT Press.

Ericsson, K. A., & Smith, J. (1991). Prospects and limits of the empirical study of expertise: an introduction. In K. A. Ericsson & J. Smith (Eds), *Toward a general theory of expertise: Prospects and limits* (pp. 1–38). Cambridge, UK: Cambridge University Press.

Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds), *Complex information processing: The impact of Herbert A. Simon* (pp. 235–267). Hillsdale, NJ: Lawrence Erlbaum.

Ericsson, K. A., & Ward, P. (2007). Capturing the naturally-occurring superior performance of experts in the laboratory: Toward a science of expert and exceptional performance. *Current Directions in Psychological Science 16*(6), 346–350.

Ericsson, K. A., Whyte, J., & Ward, P. (2007). Expert performance in nursing: Reviewing research on expertise within the framework of the expert-performance approach. *Advances in Nursing Science 30*(1), E58–E71.

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds), *The Cambridge handbook of expertise and expert performance* (pp. 41–67). Cambridge, UK: Cambridge University Press.

Fitts, P. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243–285). New York: Academic Press.

Gobet, F., & Chassy, P. (2008). Towards an alternative to Benner's theory of expert intuition in nursing: A discussion paper. *International Journal of Nursing Studies 45*(1), 129–39.

Gobet, F., & Simon, H. A. (1996). Templates in chess memory. A mechanism for recalling several boards. *Cognitive Psychology 31*(1), 1–40.

Gore, J., Ward, P., Conway, G., Ormerod, T., Wong, W., & Stanton, N. (2018). Editorial. Naturalistic decision making: Navigating uncertainty in complex sociotechnical work—an introduction to the Special Issue. *Cognition, Technology, & Work*. Available from: http://static.springer.com/sgw/documents/1629808/application/pdf/CTW+SI+on+NDM+2018.pdf

Hatano, G., & Inagaki, K. (1984). Two courses of expertise. *Research and Clinical Center for Child Development Annual Report 6*, 27–36.

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds), *Child development and education in Japan* (pp. 262–272). New York: W. H. Freeman.

Hayes, J. R. (1985). Three problems in teaching general skills. In S. F. Chipman, J. W. Segal, & R. Glaser (Eds), *Thinking and learning skills*, Vol. 2: *Research and open questions* (pp. 391–405). Hillsdale, NJ: Lawrence Erlbaum.

Hoc, J-M., Cacciabue, P. C., & Hollnagel, E. (1995). *Expertise and technology: Cognition & human–computer cooperation*. Hillsdale, NJ: Lawrence Erlbaum.

Hoffman, R. R. (1998). How can expertise be defined? Implications of research from cognitive psychology. In R. Williams, W. Faulkner, & J. Fleck (Eds), *Exploring expertise* (pp. 81–100). New York: Macmillan.

Hoffman, R. R. (Ed.). (2007) *Expertise out of context: Proceedings of the sixth international conference on naturalistic decision making*. New York: Lawrence Erlbaum.

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

Hoffman, R. R., & Militello, L. G. (2009). *Perspectives on cognitive task analysis: Historical origins and modern communities of practice*. New York: Psychology Press.

Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D. (2014). *Accelerated expertise: Training for high proficiency in a complex world*. New York: Psychology Press.

Klein, G. (2007a). Flexecution, as a paradigm for replanning, part 1. *IEEE Intelligent Systems 22*, 79–83.

Klein, G. (2007b). Flexecution, part 2: Understanding and supporting flexible execution. *IEEE Intelligent Systems 22*, 108–112.

Klein, G. (2008). Natruralistic decision making. *Human Factors 50*(3), 456–460.

Klein, G., Phillips, J. K., Rall, E. L., & Peluso, D. A. (2006). A data/frame theory of sensemaking. In R. R. Hoffman (Ed.), *Expertise out of context: Proceedings of the 6th international conference on naturalistic decision making* (pp. 113–155). New York: Lawrence Erlbaum.

Klein, G. A. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in man-machine systems research* (Vol. 5, pp. 47–92). Greenwich, CT: JAI Press.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. In *Proceedings of the Human Factors Society 30th annual meeting* (Vol. 1, pp. 576–580). Santa Monica, CA: Human Factors Society.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics 19*(3), 462–472.

Markman, A. B. (2018). Combining the strengths of naturalistic and laboratory decision-making research to create integrative theories of choice. *Journal of Applied Research in Memory and Cognition 7*, 1–10.

Mieg, H. (2001). *The social psychology of expertise*. Mahwah, NJ: Lawrence Erlbaum.

Mosier, K. L., & Fischer U. M. (Eds). (2015) *Informed by knowledge: Expert performance in complex situations*. New York: Psychology Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Lawrence Erlbaum.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Poulton, E. (1957). On prediction in skilled movements. *Psychological Bulletin 54*, 467–478.

Rasmussen, J. (1981). Models of mental strategies in process plant diagnosis. In J. Rasmussen & W. B. Rouse (Eds), *Human detection and diagnosis of system failures*. New York: Plenum Press.

Rasmussen, J. (1983). Skill, rules and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics SMC-13*(3), 257–266.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. New York: North-Holland/Elsevier.

Rasmussen, J., & Jensen, A. (1974). Mental procedures in real-life tasks: A case study of electronic troubleshooting. *Ergonomics 17*, 293–297.

Rasmussen, R., Pejtersen, A. M., & Goodstein, L. P. (1994). *Cognitive systems engineering*. New York: Wiley & Sons.

Rosenberg, J. I. (2000). Reconstructing clinical training: Self-reflection, deliberate practice and critical thinking skills. *International Journal of Psychology 35*, 415.

Schraagen, J.M.C. (2018). Naturalistic decision making. In L. J. Ball & V. A. Thompson (Eds), *The Routledge international handbook of thinking and reasoning* (pp. 487–501). London and New York: Routledge.

Selinger, E., & Crease, R. P. (2006). *The philosophy of expertise*. New York: Columbia University Press.

Selz, O. (1981). *Die gesetze der produktiven und reproduktiven geistestätigkeit kurzgefasste darstellung*. Bonn. Reproduced in N. H. Frijda, & A. D. de Groot (Eds), *Otto Selz: His contribution to psychology* (pp. 21–75). The Hague: Mouton. (Original work published 1924.)

Simon, H. A., & Gilmartin, K. (1973). A simulation of memory for chess positions. *Cognitive Psychology 5*, 29–46.

Snoddy, G. S. (1926). Learning and stability: A psychophysiological analysis of a case of motor learning with clinical applications. *Journal of Applied Psychology 10*, 1–36.

Sonnetag, S., & Kleine, B. M. (2000) Deliberate practice at work: A study with insurance agents, *Journal of Occupational and Organizational Psychology 73*, 87–102.

Starkes, J., & Allard, F. (1993). *Cognitive issues in motor expertise*. Amsterdam: North-Holland.

Starkes, J. & Ericsson, K. A. (2003). *Expert performance in sport: Advances in research on sport expertise*. Champaign, IL: Human Kinetics.

Staszewski, J. J. (2013). *Expertise and skill acquisition: The impact of William G. Chase*. New York: Psychology Press.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement on one mental function upon the efficiency of other functions, I: Extended training and experience are required to achieve high levels of proficiency. *Psychological Review 8*, 247–261.

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

Ward, P., Gore, J., Hutton, R., Conway, G., & Hoffman, R. (2018). Adaptive skill as the *conditio sine qua non* of expertise. *Journal of Applied Research in Memory and Cognition 7*(1), 35–50.

Ward, P., Hodges, N. J., Starkes, J. L., & Williams, A. M. (2007). The road to excellence: Deliberate practice and the development of expertise. *High Ability Studies 18*(2), 119–153.

Wiggins, M. W., & Loveday, T. (Eds). (2015). *Diagnostic Expertise in Organizational Environments*. London: Routledge.

Williams, A. M., Davids, K., & Williams, J. G. (1999). *Visual perception and action in sport*. London: E & FN Spon.

Yates, J. F. (2001). "Outsider:" Impressions of naturalistic decision making. In E. Salas & G. Klein (Eds), *Linking expertise and naturalistic decision making* (pp. 9–33). Mahwah, NJ: Lawrence Erlbaum.

SECTION I

CHARACTERIZING EXPERTISE

FRAMEWORKS, THEORIES, AND MODELS

SECTION EDITOR: JAN MAARTEN SCHRAAGEN

CHAPTER 2

THE CLASSIC EXPERTISE APPROACH AND ITS EVOLUTION

FERNAND GOBET

Introduction

Chase and Simon's research on chess expertise, published in three papers (Chase & Simon, 1973a, b; Simon & Chase, 1973), was highly influential and defined research on expertise for the following decades. It was followed by a spate of research on expertise, particularly in Pittsburgh, which tended to share a number of characteristics: close links with cognitive psychology and adherence to the then-dominant information-processing paradigm, tests of Chase and Simon's chunking theory with elegant experiments, detailed analyses of the processes involved, use of verbal protocols, and small number of participants. I will call this approach the *classic expertise approach*.

The aim of this chapter is to describe this approach and to understand why it was (and still is) so influential. After distinguishing it from *non-classic* approaches, I briefly discuss early research on expertise that influenced—to various extents—the classic approach. I then describe in some detail Chase and Simon's classic papers and chunking theory. This leads the way to a presentation of some of the key experimental and theoretical research. Given the limited space available, this presentation will by necessity be very selective. The chapter then discusses more recent theories that can be considered as outgrowths of the classic approach, providing a good opportunity to try to understand not only its key characteristic but also why it had such a large impact. The chapter concludes by a discussion of what this approach tells us about the means to address the challenges currently facing research on expertise.

It might be helpful at this point to contrast the classic approach with other influential research paradigms that predated it. Alfred Binet is arguably the first psychologist to have studied expertise experimentally. He developed clever methods for studying great calculators (Binet, 1894b, 1966) and magicians (Binet, 1894a). In particular, he pioneered a number of chronometric techniques and instruments for understanding how skilled magicians create visual illusions in magic tricks. His research on calculators, including the experimental tasks he developed, would have a direct impact on skilled memory, one of the theories developed within the classic approach. By contrast, his research on chess players (Binet, 1894b), which relied on interviews and introspection, was not as influential, although it did emphasize the role of memory and knowledge in expertise.

In the tradition of research known as judgment and decision making, Meehl's (1954) review of the literature criticized expertise in psychotherapy and psychiatry. His central points were that experience is not a good predictor of expertise and that the diagnoses reached by simple mathematical models have a much higher reliability than those reached by human clinical experts. Research into individual differences, sometimes called differential psychology, used psychometrics to study expertise and creativity. In this approach, superior performance is considered as mostly due to innate talent (Djakow, Petrowski, & Rudik, 1927; Galton, 1869). Lehman (1953) was a prime example of the historiometric approach, which uses the statistical analysis of archival documents to study outstanding performance. He found interesting patterns in the way artistic and scientific creativity changes during one's career and how these patterns differ between different fields. Finally, skill[1] had been studied in the fields of human factors and engineering psychology, amongst other related fields, starting with the early work of Bryan and Harter (1899). Probably the best-known result in this field is Fitts's (1964) idea that the acquisition of perceptual and motor behavior consists of three stages: the *cognitive phase*, the *associative phase*, and the *autonomous phase*. In general, when one progresses through these phases, behavior moves from conscious effort (use of declarative rules and trial and error) to unconscious, automatic, and more efficient behavior. As behavior demands less attention in the last phase, it is then possible to perform several tasks in parallel.

While these approaches were almost certainly known to researchers of the classic approach to expertise, their influence was less than that of Dutch psychologist Adriaan de Groot, who pioneered some of the tools central to the classic approach, such as the recall task and the use of concurrent verbal protocols for studying problem solving, as discussed in the following section.

De Groot's Research on Chess

Just before World War II, De Groot collected data on chess players' thinking for his PhD thesis, which was later turned into a book (1946, 1965). Whereas the wisdom of

[1] Many authors use "skill" and "expertise" interchangeably, but others prefer using "skill" for the mastery of relatively simple, typically perceptual and motoric, behaviors and "expertise" for more complex, typically cognitive, behaviors. Following Chase and Simon's example who titled one of their papers "Skill in chess," this chapter will use both terms as synonyms.

the time was that there would be large differences between grandmasters and amateurs in the amount of search carried out (e.g., in measures such as the number of positions visited or the depth of search), he found that these variables hardly distinguished such players. Rather, the key differences were in perception and intuition (the speed with which grandmasters can identify the key features of a position) and in knowledge (e.g., typical positions and the playing methods to use in specific cases).

Numerous are De Groot's contributions to the classic approach to expertise.[2] The recall task he devised, originally aimed at understanding perception, became one of the central weapons in the experimental arsenal of the classic approach. So did his task of asking individuals of different skill levels to think aloud when trying to find a solution to a problem. This type of experiment was not totally new, of course. It was originally developed by Otto Selz (1922) to understand the processes underpinning productive thinking in (non-expert) individuals and was used by Bahle (1930) in a study of music composition. Selz's proposition that thinking consists of a linear chain of cognitive operations anticipated the idea of a production system (Newell & Simon, 1972) but was expressed only verbally and thus lacked precision.

De Groot's genius was to devise statistics to quantify the structure of problem solving, in particular using the mathematical idea of a tree. In doing so, not only was he able to study Selz's framework statistically but he also anticipated Newell and Simon's (1972) idea of problem-space states and many measures of search trees later used in computer science. With hindsight, this seems a rather obvious choice, as trees are a very natural way to describe how different variations and subvariations are explored when trying to solve a chess problem. However, applying this formalism to human thinking was ground-breaking.

De Groot's book is a gold mine of ideas and insights. While his experiments did not satisfy the canons of current methodology—and perhaps because of this!—he anticipated most of the key questions in the study of expertise: the roles of perception, memory and knowledge, the organization of expert problem solving, and the importance of both talent and practice.

Simon's Early Computer Models of Chess Cognition

Chase and Simon's three seminal papers on chess skill combined Chase's acumen in designing experiments with Simon's theoretical and methodological knowledge—a

[2] A good case could be made for including De Groot in the classic approach. After much deliberation, I decided against it, as this would destroy the classic Aristotelian unities in drama (action, time, and place) organizing this chapter. In addition, some of De Groot's views were clearly at variance with the classic approach, notably his defense of introspection as scientific method and his skepticism about the usefulness of computer simulations (for a discussion, see De Groot, Gobet, & Jongman, 1996).

very powerful mix! While Chase was a young associate professor when the papers were published, Simon had already received world fame for having created the fields of artificial intelligence and modern cognitive psychology.

Several previous lines of Simon's research are evident in these three papers, in particular when the discussion focuses on the mechanisms enabling expertise. The concept of bounded rationality (Simon, 1956), for which Simon would later earn the Nobel Prize in Economics, looms large in chunking theory, which strongly emphasizes the limits of short-term memory, the restricted span of attention, and the slow rate at which learning takes place. The idea of selective search is also used when Chase and Simon discuss problem solving.

Several computer programs developed by Simon and colleagues had considerable influence on chunking theory. Two programs implemented the idea of selective search and satisficing (the use of satisfactory rather than optimal solutions), a key concept in Simon's theory of bounded rationality: NSS (Newell, Shaw, & Simon, 1958a) and MATER (Baylor & Simon, 1966). A third program, PERCEIVER (Simon & Barenfeld, 1969), simulated a chess player's eye movements. The results showed that information-processing mechanisms similar to those used for problem solving can explain high-level perception.

Another important source of ideas was the EPAM model,[3] which was originally applied to verbal learning (Feigenbaum & Simon, 1962). Not only did EPAM implement several ideas linked to the notion of bounded rationality, such as limited short-term memory capacity and slow learning rates, but it also provided mechanisms crucial for chunking theory. It showed that it was possible to closely link perception, learning, and memory in a computational model; in addition, it provided mechanisms explaining how chunks can be incrementally learned as a function of the information provided by the environment. A chunk is essentially a node in an EPAM network. In later work, Simon (1990, p. 1) wrote that "human rational behavior is shaped by a scissors whose blades are the structure of task environments and the computational capabilities of the actor." EPAM, by incorporating both learning mechanisms exploiting the regularities of the environment and assumptions about cognitive limits, addresses the two blades of this analogy.

With respect to methodology, the classic expertise approach was clearly influenced by the views that Newell and Simon (1972) expounded in their opus magnum, *Human Problem Solving*. In particular, several researchers adopted their emphasis on processes, often focusing on the behavior of very few participants studied in great detail, rather than opting for a large number of participants and statistical power. There was also an emphasis on content, as is appropriate given that experts master a specific domain: being an expert in ornithology is not the same as being an expert in poker. Moreover, Newell and Simon stressed the importance of strategies and heuristics, even when they are specific to a single individual.

[3] EPAM stands for Elementary Perceiver and Memorizer.

A final methodological point made by Newell and Simon was of course about the best way to express theories. They advocated the use of computer programs and were ruthless in criticizing the inadequacies of informal, verbal theorizing (e.g., Newell, Shaw, & Simon, 1958b; Newell & Simon, 1961) (for a discussion, see Gobet & Lane, 2015). Computer modeling offers several important advantages: behavior can be decomposed into elementary processes, which makes it possible to provide mechanistic explanations; behavior can be studied as a process that evolves as a function of time, as opposed to using just one or at most a few snapshots, as is common in experimental psychology; and finally, models are *sufficient* in the sense that they offer explanations that can generate the behavior under study. For example, a computer model of chess memory can recall chess positions, ideally with the same errors as those committed by humans.

The methodology of verbal protocols, another important contribution of Simon, was not much used in the empirical work described in Chase and Simon's papers. However, it was important in other developments with the classic approach, most notably Chase and Ericsson's experiments on the digit span task (see the section "Digit span task").

Chase and Simon's Seminal Papers

"Perception in chess" (Chase & Simon, 1973b) reports an experiment with three chess players (one beginner, one class A player (good amateur), and one international master) and two tasks: De Groot's short-term recall task and a perception task, where players had to copy a position in plain view onto a second chessboard. In addition to the game positions used by De Groot (1946, 1965), Chase and Simon also used random positions. The originality of the paper consists in the converging methods developed for identifying chunks (configurations of pieces): glances at the stimulus position in the perception task, and, in both tasks, latencies between the placement of two pieces and pattern of chess relations between those pieces. According to the authors, the results strongly suggested that expertise in chess derives from the ability to encode positions as perceptual chunks.

"The mind's eye in chess" (Chase & Simon, 1973a) is a chapter in the *Proceedings of the Eighth Annual Carnegie Symposium on Cognition* (Chase, 1973), devoted to visual information processing. It starts by summarizing the data presented in the first paper, and then presents additional analyses and new experiments aiming at pinpointing the concept of a chunk. The impression left by the first part of the paper is that much can be learnt by clever experiments addressing well-specified hypotheses, even with a sample as small as three participants. The second part of the paper develops the narrow version of chunking theory, an information-processing theory of chess skill. The interest here is to explain skill differences in De Groot's recall task. Based on earlier computational models of chess perception and memory developed by Simon and colleagues (see previous section), the authors argue that chess skill is made possible by the acquisition of a large number of chunks. Attentional mechanisms detect salient pieces, to which eye movements are directed. Pointers to recognized chunks are then placed in short-term memory. The third part of the paper presents additional experiments, focusing on long-term memory of positions and games, and on immediate recall of moves. It also investigates the Knight's Tour, a task supposed to measure chess talent. The fourth and final part describes the full version of chunking theory, aimed not only at explaining behavior in memory tasks, but also how chess players select good moves. The theory combines the ideas of mental imagery, chunking, and production system (Newell & Simon, 1972).[4] With respect to mental imagery, the mind's eye not only stores visuo-spatial structures from external inputs and memory stores but can also manipulate them by mental operations. Its capacity is limited. Long-term memory (LTM) chunks, such as specific patterns of pieces on a chessboard, act as *conditions* to *actions*, which evoke possible moves or plans. Chunks are also linked to information allowing configurations of pieces to be manipulated in the mind's eye. Pattern recognition makes it possible for players to visualize the chessboard and to anticipate moves during look-ahead search. An important aspect of the theory is that pattern recognition occurs not only on the external board, but also with the boards imagined in the mind's eye. It is thus an important mechanism for explaining how players keep selecting reasonable moves during look-ahead search.

While the first two papers were mostly addressed to psychologists, the third paper, "Skill in chess" (Simon & Chase, 1973), wooed a broader audience and was published in a generalist journal, the *American Scientist*. The emphasis is on the computational models of search, perception, and memory developed by Simon and colleagues, and on the narrow and full versions of chunking theory. The summary of the empirical work described in the previous two papers is limited to the methods and data supporting the psychological reality of a chunk. A fair amount of discussion is devoted to the number of chunks (50,000) required to become a master, as estimated by Simon and Gilmartin (1973), and the time needed for becoming a class A player (from 1,000 to 5,000 hours) and a master (from 10,000 to 50,000 hours). In the final paragraphs of the article, Simon and Chase (p. 403) answer the question of how one becomes a master—"The answer is *practice*—thousands of hours of practice" and note that "clearly, practice also interacts with talent."

Key Empirical Work in the Classic Expertise Approach

Chase and Simon's paper spawned a flurry of experimental papers. Beyond replicating the skill effect in the memory recall task in several domains (sports, Allard & Starkes, 1980; bridge, Charness, 1979; electronics, Egan & Schwartz, 1979; computer programming,

[4] Productions are rules of the type (IF condition THEN action). For example: IF the traffic light is red, THEN stop.

Schneiderman, 1976), several papers aimed to test some of the theoretical assumptions of chunking theory. This section reviews some of the most influential papers, organizing them by domain of research (for a detailed discussion, see Ericsson, Charness, Feltovich, & Hoffman, 2006; Gobet, 2016; Gobet, de Voogt, & Retschitzki, 2004).

Games

Some studies tested the hypothesis that information is grouped as chunks in players' long-term memory. For example, Frey and Adesman (1976) incrementally displayed a chess position by groups of four pieces, each new group appearing every two seconds. Under a control condition, the position was displayed column by column. Players recalled the position better when it was presented by chunks than when it was presented by columns. An unexpected support for chunking theory was that recall was better under the chunk-by-chunk condition than when the position was shown in its totality for the same duration (12 seconds). It is possible that delineating chunks enables players to recognize them better than when they must extract them from a complete position.

Reitman (1976) replicated Chase and Simon's study with two Go players, one master and one beginner. Unlike the original experiment, the correspondence was poor between the chunks identified by glances in the perception task and those identified with latencies in the recall task. Reitman also performed a partitioning task, where participants were asked to draw the boundaries of the clusters of pieces they found meaningful. Contrary to Chase and Simon's hypothesis that chunks form a hierarchy, her master perceived the positions as overlapping groups, a result that was later replicated by Chi (1978) with chess.

Charness (1976) and Frey and Adesman (1976) tested the hypothesis that it takes a relatively long time to encode new information in LTM (8 seconds to create a new chunk). Together with the assumption that the capacity of short-term memory (STM) is limited to seven items, this hypothesis implies that recall should be much affected by interpolating a second task between the presentation of a chessboard and its recall. This is because the second task should wipe out the pointers to LTM chunks from STM, as is clearly the case in experiments using unfamiliar material (Kintsch, 1970). However, it turned out that this does not happen with familiar material such as chess, as only a small decrease in memory recall was observed (around 10 percent) (Charness, 1976; Frey & Adesman, 1976). Interference was minimal even when the interpolated task was of a similar nature as the main task, such as finding the best move in a chess position.

Several researchers argued that the chunks proposed by Chase and Simon are simply too small to be of any use, and that higher-level knowledge structures are employed rather. Some results speak in favor of this viewpoint. Showing participants the moves leading to a position produces better recall (Goldin, 1978). Similar results were obtained in an unexpected recall task by manipulating the level of semantic processing at which players study a position. Independently of skill level, recall was higher when the task was to evaluate a position and select a good move than to count the number of pieces placed on white and black squares (Lane & Robertson, 1979). However, the effect disappeared when participants were previously told that they would have to recall the position.

In an elegant experiment, Chi (1978) demonstrated that research on chess could have implications beyond the study of expertise—in her case, a challenge to leading theories in developmental psychology. A dominant theory in developmental psychology holds that cognitive development is in great part produced by neural changes affecting memory capacity, speed of information processing, and executive functions (e.g., Pascual-Leone, 1970). (Currently, this is probably the leading explanation in cognitive neuroscience.) Chi showed that this is a vast simplification, and that memory capacity interacts with knowledge, and chunking in particular. Children and adults were asked to recall digits and chess positions. While adults outperformed children in the digit span task, chess-playing children had a better recall in the chess task than adults who did not play chess. Thus, knowledge in a domain more than compensates (putative) differences due to neural development. Incidentally, age differences in the digit span can also, at least to some extent, be explained by the idea that, compared to children, adults have acquired more chunks about numbers, such as dates and prices (Gobet, 2016; Simon, 1974).

Physics

Several key papers were written on physics, often using protocol analysis to compare the way novices and experts solve problems. When trying to solve relatively simple problems, novices tend to begin from the goal, and move back to the givens of the problem (i.e., they *search backward*), while experts tend to do the opposite: they start from the current situation and move to the goal of the problem (i.e., they *search forward*) (Larkin, McDermott, Simon, & Simon, 1980a; Simon & Simon, 1978). However, experts revert to backward search with difficult problems. A similar pattern has later been identified in other, but by no means all domains of expertise.

Chi, Feltovich, and Glaser (1981) were interested in the types of representation used by experts in physics and the kind of knowledge that enables them. A sorting task, where participants had to categorize problems from an undergraduate textbook, yielded two main results: first, the categories used by the novices and experts did not have much overlap; and second, the kind of representations used by the two groups were vastly different: experts' problem representations employed fundamental principles of physics, such as the concept of force, whilst novices' representations were based on superficial aspects of the problems, such as the kind of device used (e.g., pulley vs inclined plane). These experimental results were supplemented by a theoretical analysis, where novice and expert protocols were coded as node–link structures (schemas) and production rules. Based on this very detailed analysis, the authors concluded that experts' schematic knowledge allows them to build efficient representations of a problem and is linked to likely solution methods.

The research on physics expertise had important implications for education, as it provided insight about the kind of representations and solving methods that students need to acquire in order to move from novices to experts. This does not mean that novices can become experts simply by searching forward and learning the fundamental principles of physics by rote. Experts' knowledge is encoded as productions, and acquiring them requires solving many problems—hence a considerable amount of practice. Learning occurs through doing. In this respect, Simon (1980) notes that textbooks in physics and other domains tend to underemphasize the condition part of conditions, and hence opportunities to acquire perceptual knowledge.

Writing

While the analysis of verbal protocols had historically been predominantly used to study problem solving, Flower and Hayes drew on this methodology for understanding the processes involved in expert writing (Flower & Hayes, 1981; Hayes & Flower, 1980). They used Newell and Simon's (1972) framework of task environment, long-term memory, and processes. The task environment includes many external aspects influencing writing, such as the topic to address, the targeted audience, and the writer's motivation. Long-term memory includes knowledge about the topic and the audience, as well as methods such as writing plans. Finally, Flower and Hayes assumed that there are three main writing processes: planning, translating, and reviewing, each of them being subdivided into several subprocesses. Contrasting with previous stage models of writing, their model assumes that the process of writing is not linear, but hierarchical in nature: processes form a complex hierarchy, where any process can be embedded within other processes. Writers follow two kind of goals: process goals, which are instructions writers give to themselves (e.g., "let's copy-edit this chapter"), and content goals, such as how a writer intends a section to affect the reader. Goals are dynamic: the initial goals are typically revised as writing progresses and new goals are created. Just like the study of expertise in physics, research into writing expertise ensured that the classic approach would have a high impact on the field of education. Flower and Hayes (1981) is a classic in the field, with more than 4,500 citations on Google Scholar.

Computer Programming

Although Chase and Simon's recall experiment had been replicated several times, several authors disputed their explanation, which is essentially based on the number of chunks in LTM. Rather, they argued that knowledge organization is the key factor. Using a statistical technique making it possible to infer tree structures from the order with which the information is recalled, McKeithen, Reitman, Rueter, and Hirtle (1981) found important skill differences in chunk organization. Compared to beginners and intermediates, expert programmers' organization relies less on simple mnemonics and common-language associations, and more on abstract and functional programming knowledge. Similar results were obtained by Adelson (1981), who found that novices' organization relied on syntax, while experts' organization was more hierarchical, semantic, and abstract in nature, and was related to the functions of the programs.

Music

Sloboda (1976b) applied Chase and Simon's recall experiment to music, comparing novices with experienced musicians in how they could memorize musical excerpts presented visually. After the brief presentation of notes, participants were required to reproduce them on an empty stave. To control for novices' lack of knowledge of musical notation, Sloboda used very simple stimuli (between one and six random notes). Presentation times ranged from 20 ms to 2 seconds.

Musicians were much better with the 2-second presentation, but not with the 20-ms presentation (both groups performed poorly in this case). However, when participants had to reproduce not the exact sequence of notes, but the contour of notes signaled by ups and downs, musicians outperformed novices even with 20 ms (Sloboda, 1978). Note that Sloboda found a skill effect with random sequences of notes, whilst there was no skilled difference with random positions in Chase and Simon's experiment.[5] Another interesting result, this time with an auditory presentation of the stimuli, was that recall was not affected when participants were asked to perform interfering tasks—even remembering a melody—when listening to musical stimuli. Encoding music is therefore an automated skill.

The hypothesis of automaticity, but also of the use of high-level schemas, was supported by a further experiment carried out by Sloboda (1976a) with competent pianists. Scores of classical music were modified by displacing one note by one scale step, creating dissonant sequences. The pianists sight read each piece twice. Whilst accuracy was very high (above 97 percent overall), about 40 percent of the misprints were incorrectly recalled, pianists substituting the incorrect note by the correct one. Interestingly, the number of mistakes increased the second time pianists played the piece, although they made fewer mistakes overall. Sloboda suggests that pianists built an overall representation of the structure of the piece, which led them to correct the misprints in the score.

[5] Later research showed that, in most domains of expertise, there is a skill effect even with randomized material (Sala & Gobet, 2017).

Digit Span Task

The research on extraordinary memory in the digit span (Chase & Ericsson, 1981; Ericsson, Chase, & Faloon, 1980) tested, and refuted, one of the key assumptions of chunking theory—that long-term memory encoding is slow. SF, a student with average STM capacity, was trained to improve his performance in the digit span task. Digits were dictated at the pace of one digit per second. After a fairly short but intensive practice (about 250 hours in 2 years), SF was able to recall eighty-four digits, which was equivalent to the performance of the best mnemonists at the time.[6] To do so, he developed a number of mnemonics, including recoding digits using his knowledge of race times and dates, and using a pre-learned hierarchical structure to store them. These results are hard to explain with chunking theory, which would have to make the implausible assumption that SF had learned very large chunks so that eighty-four digits could be encoded in STM. They led to the development of skilled memory theory, which will be discussed later.

Key Theoretical Work in the Classic Expertise Approach

The classic expertise approach also produced a large number of influential theoretical works. Some studies aimed to refine the idea of chunking, others to replace it with alternative mechanisms.

Computational Model of Chess Memory

MAPP (Memory-Aided Pattern Perceiver) (Simon & Gilmartin, 1973) is an application of EPAM (see earlier) to chess that implements a subset of chunking theory. It assumes that the human information-processing system is limited by constraints applying to all individuals, irrespective of their level of expertise. For example, STM is limited to seven items and it takes 8 seconds for creating a new chunk in LTM. Learning occurs by adding nodes (chunks) to a discrimination network, which can then be accessed through perceptual cues. MAPP simulates the recall task by directing its attention to salient pieces and trying to recognize chunks around them. If a chunk is recognized, a symbol pointing to it is placed in STM. During recall, MAPP uses the symbols in STM to access LTM chunks, and then unpacks the information contained in those chunks.

[6] The world record is currently held by Lance Tschirhart (USA) with 456 digits. Unlike SF but like most mnemonists, Tschirhart uses a mnemonic based on recoding numbers into words and creating vivid associations between these words.

In the computer simulations, MAPP could reproduce the performance of a good amateur, but not of a master. Extrapolating from the simulations, Simon and Gilmartin proposed that masters have learnt about 50,000 chunks.

Computational Models of Physics and Engineering Problem Solving

Larkin, McDermott, Simon, and Simon (1980b) built a production system (ABLE) explaining how novices become experts in physics; in particular, it accounts for change in search strategy—from backward search with novices to forward search with experts. ABLE can use declarative statements for solving problems and derive new results, and then reuse these results when solving new problems. Bhaskar and Simon (1977) modeled a more complex and semantically richer domain, thermodynamics as it is taught to students in engineering. A methodological contribution of this project was a computer program that codes verbal protocols semi-automatically.

Skilled Memory Theory

Skilled memory theory (Chase & Ericsson, 1981; Ericsson et al., 1980) was motivated by several anomalies in Chase and Simon's data (e.g., the fact that chunks were relatively small for their master) and also by the difficulty faced by chunking theory in explaining SF's extraordinary memory in the digit span task. Skilled memory theory consists of three principles. First, memory cues are used to link new information with previous long-term memory knowledge; second, retrieval structures allow rapid storage in LTM; and third, intensive practice leads to a decrease of LTM storage and retrieval times. Retrieval structures are assumed to be domain-specific: for example, a structure developed for memorizing briefly presented digits is of little help for memorizing colors. The main application of the theory was to explain mnemonists' memory and, later, mental calculation (Staszewski, 1988). An infelicity of the theory is that its principles belong to different levels of explanation. The first two are about memory mechanisms, while the third principle is a redescription of the empirical data.

High-Level Knowledge Theories

Chase and Simon's theory stressed the importance of acquiring a large number of chunks. Other theories emphasized the way knowledge is organized in LTM and qualitative differences between novices and experts. In addition, the knowledge structures postulated are more complex and abstract than chunks and often refer to schemas, which encode both static and variable information. A good example of this approach is the work of Chi and colleagues on physics (Chi et al., 1981), and some years later, the work of Patel and Groen (1986) on medical reasoning.

Outgrowths of the Classic Approach to Expertise

This section briefly discusses several more recent theories that were directly influenced by the classic approach to expertise.

Long-Term Working Memory

Long-term working memory is a theory of expertise and memory based on skilled memory (Ericsson & Kintsch, 1995). Schemas, retrieval structures, and associations with items and context enable rapid access to LTM, to the point that LTM can be used as an extension of working memory in domains where one has acquired expertise. More specifically, cognitive processing is considered as a succession of stable states representing end products, which can be encoded in LTM when individuals have acquired sufficient expertise in a domain. Beyond digit span and mental calculation, the theory has been applied to domains such as reading and problem solving in chess.

EPAM-IV

EPAM-IV (Richman, Staszewski, & Simon, 1995) is a computer model based on EPAM (see earlier). It simulated the behavior of DD, a mnemonists who was able to increase his memory for briefly presented digits up to 108 digits (Ericsson & Staszewski, 1989), using strategies similar to those employed by SF (see the earlier section "Digit Span Task"). Compared to earlier versions of EPAM, the model makes more detailed assumptions about the components of STM and LTM. However, the critical innovation is the assumption that DD uses retrieval structures to encode information in LTM quickly, in a matter of a few hundred milliseconds. These structures are fairly similar to those postulated by skilled memory theory; in particular, while using them is rapid, learning them takes a long time.

Being stated formally, EPAM-IV makes very precise predictions, which in general are in line with the observed data. Another important contribution of this work is that the concept of retrieval structure was specified formally and with great precision, which is in contrast with skilled and long-term working memory theories.

Template Theory and CHREST

A first aim of CHREST (Chunk Hierarchy and REtrieval STructures; Gobet & Simon, 2000), the computational implementation of template theory (Gobet & Simon, 1996), was to develop a theory seamlessly integrating perception, learning, and memory. A second aim was to show how low-level knowledge structures (i.e., chunks) could lead to the acquisition of high-level, schema-like structures, called templates. Templates make it possible to encode information rapidly into LTM. In this respect, they are comparable to the retrieval structures of skilled memory and EPAM-IV. However, a key difference is that, with CHREST, their learning is assumed to be unconscious and based on the statistical properties of the input rather than on deliberate goals as with mnemonists. CHREST associates every cognitive process to a time parameter, which enables precise predictions about the timing of behavior.

The first application was chess, where CHREST corrected several limitations of MAPP (Simon & Gilmartin, 1973): the new model simulated eye movements in detail, selected the chunks to learn itself (this was done by the programmer with MAPP), and could simulate grandmaster recall (MAPP was stuck at expert level). Later applications covered aspects of expertise in physics, board games such as Awele and Go, and computer programming (Gobet, Lloyd-Kelly, & Lane, 2017). In all these cases, the program learns chunks and templates by processing naturalistic input. The model has also been applied to the acquisition of language (both syntactic structures and vocabulary). An exciting implication is that acquiring one's first language can be considered as acquiring a type of expertise (Gobet et al., 2001; Jones, Gobet, & Pine, 2000). This view is certainly at odds with standard linguistics, which considers that important aspects of language, in particular those related to syntax, are innate.

It is worth noting that the key insights of skilled memory, long-term working memory, and template theory—that LTM storage is rapid with experts—has recently obtained support from brain imaging research (Guida, Gobet, Tardieu, & Nicolas, 2012). If there is sufficient practice in a domain, then cerebral functional reorganization occurs allowing LTM structures to be used as virtual working memory.

Naturalistic Decision Making

The naturalistic decision-making approach studies expert decision making in real-world and high-stake situations such as fire-fighting and hospital intensive care units (Klein, 1998; Zsambok & Klein, 1997). In most of these cases, decisions are made under severe time pressure. The recognition-primed decision model used in this approach emphasizes mechanisms similar to those used by chunking theory: pattern recognition, intuition, selective search, and satisficing in expert decision making. As argued by Gobet (2016), the model might be less applicable for explaining expert decision making in cases where there is sufficient time for carrying out search when solving difficult problems.

Deliberate Practice

As noted earlier, Simon and Chase argued that it took from 10,000 to 50,000 hours of practice to become an expert in chess, which is roughly equivalent to 10 years of study and practice (Simon, 1969). The framework of deliberate practice (Ericsson, Krampe, & Tesch-Römer, 1993) has focused on the characteristics of the practice necessary for becoming an expert. For example, it emphasizes that goal-directed, highly structured activities aimed at correcting errors and weaknesses are essential for developing expertise. In addition, it is crucial to receive rapid and veridical feedback, preferably by a coach or a teacher. The repetitive nature of training means that it is not enjoyable per se. A key prediction of the framework is that performance improvement is directly related to the amount of deliberate practice. Note that, contrary to most of the theories discussed in this chapter, the empirical evidence is mostly correlational and the focus is on the characteristics of practice and not on the cognitive mechanisms underpinning expertise.

Characteristics of the Classic Expertise Approach

The classic approach displays several characteristics that are worth discussing. A first striking feature is its strong unity. There is of course the clear geographical concentration, with very complementary research being carried out at Carnegie Mellon University and its close neighbor the University of Pittsburgh. But there is also a strong theoretical unity, which can be traced back to the works of Adriaan de Groot and Herbert Simon. De Groot provided not only empirical methods (recall task and problem solving with quantitative analysis of verbal protocols) but also key theoretical insights such as the central role of perception in expertise. Simon provided powerful theoretical ideas, such as bounded rationality and satisficing. Finally, the three Chase and Simon papers also contributed to this unity by providing a unifying theory and elegant methods for operationalizing the notion of a chunk.

A second important characteristic is that the approach was very successful in cross-fertilizing with other fields of research. With cognitive psychology, there was a fruitful exchange of methods and theoretical ideas. The classic approach was particularly influential because it cut across traditional boundaries of cognitive psychology (perception, memory, and problem solving). With artificial intelligence (AI) and particularly the then nascent subfield of expert systems, theoretical exchanges focused on the notion of knowledge. How do experts represent knowledge? Can we use what we know about human expert knowledge to build artificial expert systems? Conversely, representations were imported from AI, such as the concept of a schema. Other fields were also influenced by the classic approach. For example, Chi's (1978) study is well known in developmental psychology and, just like the research on physics, Flower and Hayes's (1981) work on reading impacted on education.

Third, the classic approach had a very distinctive theoretical flavor, inspired by Newell and Simon's (1972) framework of information-processing psychology. It put a strong emphasis on mechanisms and the relative invariants of cognition, such as the limited capacity of STM, originally assumed to be seven items but later downscaled to five items (Simon, 1974), 8 seconds to learn a chunk, and 50,000 chunks to become an expert. Computer modeling played an essential role in laying out the theoretical foundations, although it had relatively few followers even within the classic approach. While terms such as expertise, talent, and genius are often seen as ill defined and hard to study scientifically, computer modeling imposed rigorous constraints on theorizing and enabled clear-cut predictions.

Finally, the classic approach is characterized empirically by the detailed analyses of the processes in play rather than statistical elegance and power. Several key papers had very few participants indeed: Chase and Simon (1973a, b) had three participants; Reitman (1976) had two, and Ericsson, Chase, and Faloon (1980) had only one.

Looking at the Future

The aim of this section is to argue that some of the characteristics that made the classic approach so successful could be used with considerable benefits by current and future expertise researchers. While these characteristics are met every so often in contemporary expertise research, it is clear that they do not describe most mainstream approaches (e.g., deliberate practice). In my view, the main weakness of research on expertise today is that the theories that are developed tend to be unspecified and thus cannot make clear-cut predictions and be testable.

I believe that one important reason behind the success of the classic approach is that it was anchored in rigorously specified theories. Two of the Chase and Simon papers (Chase & Simon, 1973a; Simon & Chase, 1973) devoted substantial space to discuss the computer models developed by Simon and colleagues. Computational modeling is currently a rare citizen in expertise research, although there are some exceptions, as noted earlier. My first advice is then to develop more rigorous theories, implemented as computer programs. Recent developments in AI—e.g., deep learning (LeCun, Bengio, & Hinton, 2015)—open exciting prospects, both for theory and applied research.

Today, perhaps an inevitable price for specialization, researchers aim to understand expertise per se. By contrast, a second conspicuous feature of the classic approach is that it aimed to address general questions of cognition, using expertise as a means. For example, it was interested in the micro-structure of cognition and sought to find its relative invariants: the time to create a chunk in memory, the capacity of STM, and the time necessary to become an expert. Some expertise researchers have criticized this goal as misguided, as practice in some cases improves encoding and retrieval times

(Ericsson & Kintsch, 1995). It could also be argued that the classic approach did not investigate the possibility of individual differences due to talent (e.g., creating a new chunk might take 7.9 seconds for one individual and 8.1 seconds for another, a small difference that snowballs with the learning of 50,000 chunks; Chassy & Gobet, 2010). However, the possibility of using strategies for speeding memory processes and the presence of individual differences do not mean that such (approximate) constants do not exist. Identifying these parameters and setting their values, even approximatively, would bring considerable benefits not only for our understanding of expertise but also for education and coaching. Such an endeavor would also make it easier to develop rigorous theories of expertise. Thus, my second advice is to go back to fundamental questions, including setting the value of the parameters of cognition.

Another characteristic of the classic approach is that it was interested in the content of expertise, including the strategies used. Think, for example, of Chi et al.'s (1981) research on physics. Here, the interest was not only on the relationship between inputs and outputs but in the way inputs are processed, using domain-specific strategies and other kinds of knowledge, to produce outputs. This emphasis on knowledge and strategies has been lost recently,[7] possibly because publishing such research is not easy in the current culture that emphasizes sophisticated statistical analyses. My third advice—carrying out detailed analyses of experts' knowledge and behavior—dovetails naturally with the need to express theories as computer models.

My final piece of advice is without any doubt the most controversial. If we are to understand in great detail the mechanisms, knowledge, and strategies enabling expertise, it will be necessary, at least in some cases, to use single-subject designs such as those used by Ericsson, Chase, and Faloon (1980). Such designs do not mean that theory development is impossible. In fact, the research carried out by Ericsson and colleagues is a beautiful example of how detailed analyses can help answer theoretical questions. However, unless the research is purely descriptive, it is essential to have clear-cut hypotheses. Thus, my final advice is to carry more research with $n = 1$, ideally with a large number of experimental tasks and the parallel development of a computational model accounting for the data (Gobet, 2017; Gobet & Ritter, 2000).

Conclusion

This chapter has presented the classic expertise approach, which started with Chase and Simon's three seminal papers in 1973. The classic approach has broadly focused on chunking theory, often to refute it and extend it. It has had considerable impact, for both theory and applied research (e.g., education). Several characteristics of the classic approach—strong emphasis on theory, use of computer models, emphasis on

[7] An exception is the research carried out in the naturalistic decision-making tradition.

the micro-structure of cognition and multidisciplinary approach—contributed to its success. The chapter recommends that current and future students of expertise incorporate these characteristics in their research.

References

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. *Memory and Cognition* 9, 422–433.

Allard, F., & Starkes, J. L. (1980). Perception in sport: volleyball. *Journal of Sport Psychology* 2, 22–33.

Bahle, J. (1930). *Zur Psychologie des musikalischen Gestaltens. Eine Untersuchung über das Komponieren auf experimenteller und historischer Grundlage*. Leipzig: Akademische Verlagsgesellschaft.

Baylor, G. W., & Simon, H. A. (1966). A chess mating combinations program. In *1966 Spring Joint Computer Conference* (Vol. 28, pp. 431–447). Boston.

Bhaskar, R., & Simon, H. A. (1977). Problem solving in semantically rich domains: An example from engineering thermodynamics. *Cognitive Science* 1, 193–215.

Binet, A. (1894a). La psychologie de la prestidigitation. *Revue des Deux Mondes* 25, 903–922.

Binet, A. (1894b). *Psychologie des grands calculateurs et joueurs d'échecs*. Paris: Hachette. (Re-edited by Slatkine Ressources, Paris, 1981.)

Binet, A. (1966). Mnemonic virtuosity: A study of chess players. *Genetic Psychology Monographs* 74, 127–162. (Translated from the *Revue des Deux Mondes* (1893), 1117, 1826–1859).

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language. The acquisition of a hierarchy of habits. *Psychological Review* 6, 345–375.

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory* 2, 641–653.

Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology* 33, 1–16.

Chase, W. G. (1973). *Visual information processing*. New York: Academic Press.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–189). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Simon, H. A. (1973a). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

Chase, W. G., & Simon, H. A. (1973b). Perception in chess. *Cognitive Psychology* 4, 55–81.

Chassy, P., & Gobet, F. (2010). Speed of expertise acquisition depends upon inherited factors. *Talent Development and Excellence* 2, 17–27.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Erlbaum.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science* 5, 121–152.

De Groot, A. D. (1946). *Het denken van den schaker*. Amsterdam: Noord Hollandsche.

De Groot, A. D. (1965). *Thought and choice in chess* (first Dutch edition in 1946). The Hague: Mouton Publishers.

De Groot, A. D., Gobet, F., & Jongman, R. W. (1996). *Perception and memory in chess: Heuristics of the professional eye*. Assen: Van Gorcum.

Djakow, I. N., Petrowski, N. W., & Rudik, P. A. (1927). *Psychologie des Schachspiels*. Berlin: de Gruyter.

Egan, D. E., & Schwartz, E. J. (1979). Chunking in recall of symbolic drawings. *Memory & Cognition 7*, 149–158.

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (2006). *The Cambridge handbook of expertise and expert performance*. New York: Cambridge University Press.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science 208*, 1181–1182.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review 102*, 211–245.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review 100*, 363–406.

Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovski (Eds), *Complex information processing: The impact of Herbert A. Simon*. Hillsdale, NJ: Erlbaum.

Feigenbaum, E. A., & Simon, H. A. (1962). A theory of the serial position effect. *British Journal of Psychology 53*, 307–320.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243–285). New York: Academic Press.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication 32*, 365–387.

Frey, P. W., & Adesman, P. (1976). Recall memory for visually presented chess positions. *Memory & Cognition 4*, 541–547.

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. London: MacMillan.

Gobet, F. (2016). *Understanding expertise: A multidisciplinary approach*. London: Palgrave.

Gobet, F. (2017). Allen Newell's program of research: The video game test. *Topics in Cognitive Science 9*, 522–532.

Gobet, F., de Voogt, A. J., & Retschitzki, J. (2004). *Moves in mind: The psychology of board games*. Hove, UK: Psychology Press.

Gobet, F., & Lane, P. C. R. (2015). Human problem solving—Beyond Newell et al.'s (1958): Elements of a theory of human problem solving. In M. W. Eysenck & D. Groome (Eds), *Cognitive psychology: Revisiting the classic studies*. Thousand Oaks, CA: Sage.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences 5*, 236–243.

Gobet, F., Lloyd-Kelly, M., & Lane, P. C. R. (2017). Computational models of expertise. In D. Z. Hambrick, G. Campitelli & B. N. Macnamara (Eds), *The science of expertise: behavioral, neural, and genetic approaches to complex skill* (pp. 347–364). New York: Psychology Press.

Gobet, F., & Ritter, F. E. (2000). Individual data analysis and Unified Theories of Cognition: A methodological proposal. In N. Taatgen & J. Aasman (Eds), *Proceedings of the Third International Conference on Cognitive Modelling* (pp. 150–157). Veenendaal, The Netherlands: Universal Press.

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology 31*, 1–40.

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science 24*, 651–682.

Goldin, S. E. (1978). Effects of orienting tasks on recognition of chess positions. *American Journal of Psychology 91*, 659–671.

Guida, A., Gobet, F., Tardieu, H., & Nicolas, S. (2012). How chunks, long-term working memory and templates offer a cognitive explanation for neuroimaging data on expertise acquisition: A two-stage framework. *Brain and Cognition 79*, 221–244.

Hayes, J. R., & Flower, L. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3–30). Hillsdale, NJ: Lawrence Erlbaum.

Jones, G., Gobet, F., & Pine, J. M. (2000). A process model of children's early verb use. In L. R. Gleitman & A. K. Joshi (Eds), *Proceedings of the Twenty Second Annual Meeting of the Cognitive Science Society* (pp. 723–728). Mahwah, NJ: Erlbaum.

Kintsch, W. (1970). *Learning, memory, and conceptual processes*. New York: John Wiley.

Klein, G. A. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Lane, D. M., & Robertson, L. (1979). The generality of the levels of processing hypothesis: An application to memory for chess positions. *Memory & Cognition 7*, 253–256.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980a). Expert and novice performance in solving physics problems. *Science 208*, 1335–1342.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980b). Models of competence in solving physics problems. *Cognitive Science 4*, 317–345.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444.

Lehman, H. C. (1953). *Age and achievements*. Princeton, NJ: Princeton University Press.

McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. (1981). Knowledge organisation and skill differences in computer programmers. *Cognitive Psychology 13*, 307–325.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.

Newell, A., Shaw, J. C., & Simon, H. A. (1958a). Chess-playing programs and the problem of complexity. *IBM Journal of Research and Development 2*, 320–335.

Newell, A., Shaw, J. C., & Simon, H. A. (1958b). Elements of a theory of human problem solving. *Psychological Review 65*, 151–166.

Newell, A., & Simon, H. A. (1961). Computer simulation of human thinking. *Science 134*, 2011–2017.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Pascual-Leone, J. A. (1970). A mathematical model for transition in Piaget's developmental stages. *Acta Psychologica 32*, 301–345.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science 10*, 91–116.

Reitman, J. S. (1976). Skilled perception in go: Deducing memory structures from inter-response times. *Cognitive Psychology 8*, 336–356.

Richman, H. B., Staszewski, J. J., & Simon, H. A. (1995). Simulation of expert memory with EPAM IV. *Psychological Review 102*, 305–330.

Sala, G., & Gobet, F. (2017). Experts' memory superiority for domain-specific random material generalizes across fields of expertise: A meta-analysis. *Memory & Cognition 45*, 183–193.

Schneiderman, B. (1976). Exploratory experiments in programmer behavior. *International Journal of Computer and Information Sciences 5*, 123–143.

Selz, O. (1922). *Zur Psychologie des produktiven Denkens und des Irrtums*. Bonn: Friedrich Cohen.

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 323–348). Hillsdale, NJ: Erlbaum.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review 63*, 129–138.

Simon, H. A. (1969). *The sciences of the artificial.* Cambridge, MA: MIT Press.

Simon, H. A. (1974). How big is a chunk? *Science 183*, 482–488.

Simon, H. A. (1980). Problem solving and education. In D. Tuma & F. Reif (Eds), *Problem solving and education* (pp. 81–96). Hillsdale, NJ: Erlbaum.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology 41*, 1–20.

Simon, H. A., & Barenfeld, M. (1969). Information processing analysis of perceptual processes in problem solving. *Psychological Review 7*, 473–483.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist 61*, 393–403.

Simon, H. A., & Gilmartin, K. J. (1973). A simulation of memory for chess positions. *Cognitive Psychology 5*, 29–46.

Sloboda, J. A. (1976a). Effect of item position on likelihood of identification by inference in prose reading and music reading. *Canadian Journal of Psychology 30*, 228–237.

Sloboda, J. A. (1976b). Visual perception of musical notation: Registering pitch symbols in memory. *Quarterly Journal of Experimental Psychology 28*, 1–16.

Sloboda, J. A. (1978). Perception of contour in music reading. *Perception 7*, 323–331.

Staszewski, J. J. (1988). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds), *The nature of expertise* (pp. 71–128). Hillsdale, NJ: Erlbaum.

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

CHAPTER 3

DOMAIN-GENERAL MODELS OF EXPERTISE

The Role of Cognitive Ability

DAVID Z. HAMBRICK,
ALEXANDER P. BURGOYNE,
AND FREDERICK L. OSWALD

Domain-General Models of Expertise: The Role of Cognitive Ability

Why do some people reach higher levels of expertise in complex real-world tasks than other people? There is no doubt that domain-specific knowledge and skills contribute substantially to individual differences in expertise, whether it be in vocational or avocational pursuits (see Ward, Belling, Petushek, & Ehrlinger, 2017, for a review). Here, while not denying the major importance of domain-specific factors, we consider the contribution of *domain-general cognitive ability factors*, reflecting the efficiency and effectiveness of basic mental processes.

Scope and Organization

In everyday life, people often rely on reputation to identify individuals with expertise—physicians, carpenters, auto mechanics, and so on. However, reputation does not ensure a high level of expertise (Ericsson, 2006). As a scientific concept, expertise is better thought of as a person's objective level of performance in a domain, as quantified by domain-relevant tasks (Ericsson & Smith, 1991) or proxy measures (e.g., performance-based rankings). For some domains, a single type of task may be sufficient to measure expertise. For example, given that playing good chess obviously depends on making good chess moves, chess expertise can be measured with *move-choice* tasks (de Groot, 1965/1978). For other domains, no single type of task captures expertise. For example, musical expertise comprises playing music from memory, sight-reading, and improvising, among other activities. Some musicians may be strong in all these activities; others may be strong in some but weak in others. Similarly, some auto mechanics may specialize in repairing diesel engines, others in transmissions, and still others in body repair. In short, expertise may be multidimensional.

Here, we review evidence for the role of cognitive ability in acquiring expertise. Along with limited space, there are two major reasons for this restricted focus. First, much of the controversy in contemporary research on expertise revolves around the question of whether, and to what extent, cognitive ability plays an important role in acquiring expertise (see, e.g., Detterman, 2014). Second, as industrial–organizational psychologists have demonstrated, measures of cognitive ability (along with other measures) are useful in organizational settings for selecting qualified job applicants, because they are consistently and positively correlated with job performance (Schmidt & Hunter, 1998, 2004). Similarly, scores on standardized cognitive tests such as the Graduate Record Examination (GRE), the Law School Admission Test (LSAT), and the Graduate Management Admission Test (GMAT) are useful and valid predictors of success in advanced academic studies (Kuncel & Hezlett, 2007).

Table 3.1 lists the cognitive ability constructs that we consider, with a definition of each construct and examples of assessments. Though often treated as if they are empirically and conceptually distinct, measures of these constructs correlate positively, and sometimes nearly 1.0 after correcting for measurement error variance (e.g., Engelhardt et al., 2016; McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). This implies that common mechanisms underlie individual differences in these constructs, which could include acquired factors such as general problem-solving strategies, neural factors such as the functional connectivity of different brain regions, and genetic factors (see Haier, 2016). Any (or all) of these factors could contribute, directly or indirectly, to associations between cognitive ability factors and expertise. Cognitive ability constructs are also sometimes described as being *innate*, but heritability (i.e., estimated genetic contribution) of any human characteristic is always less than 100 percent (Turkheimer, 2000), leaving room for a contribution by environmental factors. At the population level, heritability is typically around 50 percent for measures of cognitive ability, indicating roughly equal contributions of genetic and environmental factors to individual differences (Knopik, Neiderhiser, DeFries, & Plomin, 2016).

Table 3.1 Domain-general cognitive ability factors, with representative definitions and examples of assessments

Construct	Definition/tests
Intelligence	Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience. (Gottfredson, 1997, p. 13) - Wechsler Adult Intelligence Scale (full-scale IQ) - Raven's Progressive Matrices (fluid intelligence, or Gf) - Air Force Officer Qualifying Test (AFQT)
Executive functioning	Executive function can be thought of as the set of abilities required to effortfully guide behavior toward a goal, especially in nonroutine situations. (Banich, 2009, p. 89) - Wisconsin Card Sorting task - Tower of Hanoi - Trailmaking
Working memory capacity	[Working memory capacity refers to] the attentional processes that allow for goal-directed behavior by maintaining relevant information in an active, easily accessible state outside of conscious focus, or to retrieve that information from inactive memory, under conditions of interference, distraction, or conflict. (Kane et al., 2007, p. 23) - Operation span - n-back - Backward digit span
Attentional control	Attention control refers to the ability to protect items that are actively being maintained in working memory, to effectively select target representations for active maintenance, and to filter out irrelevant distractors and prevent them from gaining access to working memory. (Unsworth, Fukuda, Awh, & Vogel, 2015, p. 864) - Attention Network Task (ANT) - Stroop task - Flanker task
Speed of processing	Processing speed refers to the ability to quickly and efficiently carry out mental operations. (Tucker-Drob, 2011, p. 333) - Digit-symbol substitution - Letter/pattern comparison - Choice reaction time

Review of Evidence for Role of Cognitive Ability in Expertise

Classical theories of skill acquisition (e.g., Fitts & Posner, 1967) posit that domain-general processes impact performance early in training, after which procedural knowledge becomes the major determinant. Consistent with this assumption, there is ample evidence that cognitive ability predicts initial acquisition of knowledge/skill in complex domains. For example, measures of cognitive ability from test batteries such as the Armed Services Vocational Aptitude Battery (ASVAB) positively predict job training performance, with validity coefficients averaging around 0.50 (Schmidt & Hunter, 2004). It is less clear whether cognitive ability remains a valid predictor of performance differences after extensive training. This question is not only of theoretical interest to expertise researchers (e.g., the *circumvention-of-limits hypothesis;* Hambrick & Meinz, 2011), but one of applied interest: If a measure significantly predicts performance in a task, especially beyond the beginner level, then that measure might be used to help make decisions such as whom to select for a costly training program.

Next, we review evidence relevant to this question. We performed literature searches in Google Scholar and PsycINFO, using a wide range of search terms (e.g., "expertise" and "cognitive ability" with "sports," "chess," and "aviation"). We searched approximately 1,300 documents, identifying relevant studies in two primary literatures: the literature on expertise in domains such as music, chess, and sports; and the literature on job performance. Our review focuses on studies that tested cognitive ability–performance relations across different levels of expertise, or at least in non-beginners. (We excluded studies that measured specific aptitudes, such as music aptitude and mechanical aptitude.) The specific question we set out to address is whether expertise mitigates the effect of cognitive ability on domain-relevant performance. Throughout the chapter, we note correlations between domain-specific factors and domain-relevant performance for comparative purposes.

Games

There is evidence that cognitive ability predicts acquisition of chess skill at the beginner level (e.g., de Bruin, Kok, Leppink, & Camp, 2014), but it is unclear what role it plays at higher levels of expertise. Evidence is mixed. For example, in two studies, Unterrainer and colleagues (Unterrainer, Kaller, Halsband, & Rahm, 2006; Unterrainer, Kaller, Leonhart, & Rahm, 2011) found near-zero correlations between IQ measures and chess rating in small samples of chess players ($N = 25$ and 26, respectively) with intermediate-level average chess ratings, whereas Grabner, Stern, and Neubauer (2007) found a correlation of 0.35 between IQ and chess rating using a larger sample ($N = 90$) with a slightly higher average rating. Even the latter finding is tentative because the confidence interval around a correlation of 0.35 with a sample of 90 is quite wide, ranging from 0.15 to 0.52.

To try to make sense of the conflicting evidence, Burgoyne et al. (2016) performed a meta-analysis of the relationship between cognitive ability and chess expertise. Across 19 studies, four cognitive abilities were measured: fluid intelligence, crystallized intelligence, short-term/working memory, and processing speed. The meta-analytic average of the correlations was 0.22 ($p < 0.001$). (Correlations between chess rating and domain-specific factors are typically much larger (e.g., Pfau & Murphy, 1988, $r = 0.68$).)

Burgoyne et al. also found that the correlation between fluid intelligence and expertise was stronger for less skilled (unranked) chess players than for more skilled (ranked) players (0.33 vs 0.10; see Burgoyne et al., 2018, for a correction to the originally reported values). However, it is important to note that expertise was highly confounded both with age (i.e., nearly all ranked chess players were adults, nearly all unranked chess players were youths) and with type of skill measure (i.e., Elo ratings for ranked players, chess test scores for unranked players).

In another meta-analysis, Sala et al. (2017) found that chess players are, on average, significantly higher in measured cognitive ability than non-chess players. As most of the chess samples included relatively highly skilled players, this could be because people high in cognitive ability are more likely to enjoy success in chess than those lower in cognitive ability, and are thus less likely to quit the game (i.e., performance effects). Alternatively, it could be that playing chess enhances cognitive ability (i.e., training effects) or because higher-ability individuals are more likely to take up chess than lower-ability individuals (i.e., selection effects).

Summarized, evidence is inconclusive on whether the importance of cognitive ability declines with chess expertise. The same is true for other games. In neuroimaging studies, Lee et al. (2010, $N = 16$) and Jung et al. (2013, $N = 17$) reported IQ data on small samples of elite Baduk (Korean for *Go*) players. Full-scale IQ was lower for the Baduk players than for a control group by 8 points in Lee et al. ($M = 93.2$ vs 101.2, $p = 0.052$) and 7.7 points in Jung et al. ($M = 93.1$ vs 100.8, $p = 0.06$). The fact that the Baduk group in each study had a *lower* average IQ than the control group is somewhat puzzling and may partly reflect the fact that the Baduk group had less education on average than the control group (by 1.3 years in Lee et al., $p < 0.05$; and by 1.1 years in Jung et al., $p = 0.19$).

A much larger study of Go was carried out by Masunaga and Horn (2001). Participants ($N = 263$) representing wide ranges of Go expertise completed tests of both domain-general and domain-specific factors. The domain-general battery included standard tests of fluid reasoning, short-term memory, and perceptual speed; the domain-specific battery included *Go-embedded* tests designed to measure the same abilities but with Go-specific content. The Go reasoning test was modeled after move-choice tasks in chess (de Groot, 1965/1978), and can be considered a measure of Go skill. On average, the domain-general measures correlated 0.18 with Go move-choice. The highest correlations were for fluid intelligence (avg. $r = 0.27$); group average r values (obtained from Takagi, 1997) were as follows: beginner (avg. $r = 0.21$, $p = 0.001$, $n = 62$), intermediate (avg. $r = 0.33$, $p < 0.001$, $n = 89$), expert (avg. $r = 0.27$, $p < 0.001$, $n = 92$), and professional (avg. $r = 0.18$, $p = 0.14$, $n = 20$). These correlations are not significantly different from each other (z statistics < 1). The average correlations between the fluid intelligence measures and Go rank were non-significant: beginner (avg. $r = -0.03$), intermediate (avg. $r = -0.06$), expert (avg. $r = 0.03$), and professional (avg. $r = -0.26$). It is somewhat surprising that fluid intelligence correlated with move-choice performance but not with Go rank, given the high correlation between the latter measures ($r = 0.71$) and that move-choice must be critical for success in Go tournaments. It could be that the move-choice task was somewhat artificial in that it presented the player with novel positions, whereas in actual Go games a skilled player can steer a game toward familiar territory and thus encounter more familiar positions. The average correlation of the domain-specific measures with Go move-choice was 0.46 and with Go rank was 0.47.

Word games have also been used to investigate the relationship between cognitive ability and expertise. Tuffiash, Roring, and Ericsson (2007) compared groups of elite, average, and novice Scrabble players on tests of various domain-specific and domain-general cognitive abilities. There were significant group differences (favoring higher expertise) in the domain-relevant tasks (e.g., anagramming; medium-to-large effect sizes), but not in domain-general perceptual speed (i.e., digit-symbol substitution). However, the rated players (average and elite groups) outperformed the novices on tests of vocabulary and reading comprehension (ds > 2). More recently, Toma, Halpern, and Berger (2014) found that Scrabble and crossword puzzle experts tended to outperform the control subjects on two tests of working memory capacity (avg. $d = 1.23$). As in chess, these skill group differences could reflect performance effects, training effects, and/or selection effects.

There have been a few studies of poker expertise. In a study of undergraduate students described as being *familiar with* Texas Hold 'em poker, Leonard and Williams (2015) found that scores on several subtests from the Stanford–Binet Intelligence Scales correlated non-significantly with performance on a test of poker skills. However, in a sample of 155 undergraduates representing a wider range of Texas Hold 'em experience, Meinz et al. (2012) found that working memory capacity explained a significant amount of variance (avg. $R^2 = 0.071$) in measures of Hold 'em component skills (e.g., hand evaluation), above and beyond poker knowledge (avg. $R^2 = 0.358$). Moreover, there was no evidence for Poker Knowledge × Working Memory Capacity interactions, indicating that effects of working memory capacity on performance were similar across levels of poker knowledge.

Finally, Ceci and Liker (1986) found that groups of *nonexperts* ($n = 16$) and *experts* ($n = 14$) in horserace handicapping were not only nearly identical in average IQ, but both near the population mean of 100 (Ms = 99.3 and 100.8, respectively). However, in a re-analysis, Detterman and Spry (1988) found that the correlation between IQ and a key measure of success (correct top horse) was positive in the expert group ($r = 0.35$, or 0.59 after correction for unreliability) but negative in the novice group ($r = -0.25$, or -0.42 after correction for unreliability), casting some doubt on the argument that IQ is unrelated to success in horserace handicapping. That said, these sample sizes were very small, and the result would obviously need to be replicated in a larger sample.

Music

It is also unclear what role cognitive ability plays in music expertise beyond the beginner level. Ruthsatz, Detterman, Griscom, and Cirullo (2008) found that scores on a test of fluid intelligence (Raven's Progressive Matrices) correlated positively and significantly with musical achievement in high school band members (r = 0.25, n = 178), but not in university music majors (r = 0.24, n = 19) and music institute students (r = 0.12, n = 64)—although statistical power obviously differed across the samples. Moreover, the correlations did not differ between the lower- and higher-skill groups (tests of differences in rs, z statistics < 1). Correlations with estimated amount of *deliberate practice* (Ericsson, Krampe, & Tesch-Römer, 1993) in the high school, university, and music institute samples were 0.34, 0.54, and 0.31, respectively (all significant).

Meinz and Hambrick (2010) had 57 pianists provide estimates of deliberate practice and perform tests of both working memory capacity and sight-reading. Deliberate practice accounted for 45 percent of the variance in sight-reading performance; working memory capacity accounted for an additional 7.4 percent. (The correlation between deliberate practice and working memory capacity was near zero.) Moreover, the Deliberate Practice × Working Memory Capacity interaction was non-significant, indicating that the effect of working memory capacity on performance was similar across levels of deliberate practice. By contrast, perceptual speed did not contribute significantly to the prediction of sight-reading performance.

Using a sample of 52 pianists with a more uniform level of skill (piano majors at a music university), Kopiez and Lee (2008) found that although correlations between sight-reading performance and fluid reasoning (r = 0.12) and reaction time (avg. r = -0.07) were non-significant, there was a significant correlation for a measure of perceptual speed (r = -0.44; faster processing, higher performance). The correlation between working memory and sight-reading performance did not reach significance (r = 0.26, p = 0.062). Correlations between measures of domain-relevant motoric speed (*trilling*) and sight-reading performance averaged 0.50; the correlation between deliberate practice and sight-reading performance was 0.50.

Other studies have compared musicians of varying levels of skill on measures of cognitive ability, as well as musicians to non-musicians. Schellenberg and colleagues have found that musically trained individuals tend to score higher in full-scale IQ than non-musically trained individuals (see Schellenberg & Weiss, 2013, for a review). As with chess, this difference could reflect performance effects, training effects, and/or selection effects.

Sports

Evidence for the role of cognitive ability in sports expertise is inconsistent, as well. For example, Lyons, Hoffman, and Michel (2009) found that scores on the Wonderlic IQ test correlated near zero (r = -0.04) with future NFL performance in a large sample of elite college football players (total N = 762; see also Berri & Simmons, 2011), whereas Vestberg, Gustafson, Maurex, Ingvar, and Petrovic (2012) found that a measure of executive functioning (design fluency from the D-KEFs) significantly predicted goals scored in elite Swedish soccer players ($r = 0.54$, $N = 25$), albeit in a much smaller sample.

In a meta-analysis of 42 studies, Mann, Williams, Ward, and Janelle (2007) compared nonexpert and expert athletes on performance measures from sports-specific perceptual-cognitive tasks (e.g., occlusion paradigms). Across measures, there was a statistically significant advantage for experts ($ds = 0.23$ to 0.35). Given evidence for the importance of training in acquiring skill in sports (e.g., Ward, Hodges, Starkes, & Williams, 2007), these differences likely reflect domain-specific factors, but they could also reflect domain-general factors as well (Ward et al., 2017).

In a subsequent meta-analysis of 20 studies, Voss, Kramer, Basak, Prakash, and Roberts (2010) found a significant advantage for athletes over non-athletes on processing speed (Hedges' $g = 0.67$) and *varied* attention tasks (Hedges' $g = 0.53$) but not attentional cueing (Hedges' $g = 0.17$). (Hedges' g is similar to Cohen's d.) These results lend some support to the possibility that domain-general factors contribute to sports expertise (i.e., performance effects), but as before could also reflect selection effects and/or training effects (Ward et al., 2017).

Science

The relationship between cognitive ability and scientific expertise has also been of interest to psychologists. Early studies of this relationship yielded mixed evidence. Bayer and Folger (1966) reported a correlation of -0.05 between IQ and number of citations (a proxy for scientific expertise) in a sample of 224 biochemists, and Folger, Astin, and Bayer (1970) found correlations ranging from 0.04 to 0.10 between cognitive ability in high school and number of citations in a sample of 6,300 PhDs. However, Creager and Harmon (1966) found that scores on the GRE predicted citation counts 8–12 years later (median $r = 0.28$; cited in Clark & Centra, 1982) in NSF predoctoral fellowship applicants (see also Kaufman, 1972).

More convincing results come from a meta-analysis of 6,589 correlations from 1,753 independent samples by Kuncel, Hezlett, and Ones (2001). After applying psychometric corrections for statistical artifacts such as range restriction and measurement unreliability in the criterion measures, Kuncel et al. found that estimated validity coefficients (ρs) in the population for the General GRE test were positive and significant not only for first-year GPA (avg. $\rho = 0.36$; avg. $r = 0.24$) and overall GPA (avg. $\rho = 0.34$; avg. $r = 0.23$), but also for publication citation counts (avg. $\rho = 0.20$; avg. $r = 0.15$), and were positive for research productivity (avg. $\rho = 0.10$; avg. $r = 0.08$). Validities for the Subject GRE test (reflecting domain-specific knowledge) were higher for all outcomes, including publication citation counts ($\rho = 0.24$; $r = 0.20$) and research productivity ($\rho = 0.21$; $r = 0.17$).

This evidence corroborates the results of the Study of Mathematically Precocious Youth (SMPY). As part of a planned 50-year study, the Scholastic Aptitude Test (now just called the SAT) was administered to a large national sample of gifted youth by age 13, and those scoring in the top 1 percent were tracked into adulthood ($N >$ 2,300). Analyses have since demonstrated that—even within this highly restricted range of ability—SAT scores are positively predictive of success in scientific fields. For example, Lubinski (2009) found that, compared with individuals in the 99.1 percentile, those in the 99.9 percentile were about 5 times more likely to have published in a STEM journal and about 3 times more likely to have been awarded a patent.

Thus, there is evidence that cognitive ability predicts general measures of scientific expertise. There is, however, some evidence that cognitive ability may become less important in specific scientific tasks. Hambrick et al. (2012) had a sample of 67 participants representing a wide range of knowledge and experience in geological fields perform a highly realistic *bedrock mapping* task in which the goal was to create a *field map* representing the geological structure of an area based on observable features (e.g., rock outcrops). There was a significant Geological Knowledge × Visuospatial Ability interaction, such that a composite measure of visuospatial ability positively predicted map accuracy, but only in those with lower levels of geological knowledge.

Surgery/Medicine

There is a growing literature on the role of cognitive ability in surgical expertise, but the results are no clearer than in other domains. In a study of 120 surgical residents (Schueneman, Pickleman, Hesslein, & Freeark, 1984), 4 of 5 measures of visuospatial ability correlated significantly with surgical performance (avg. $r = 0.28$), as evaluated by attending surgeons. Year of residency correlated 0.60 with surgical performance. Gibbons, Baker, and Skinner (1986) found that scores on a hidden figures test correlated significantly with surgical performance in small samples of surgical residents (rs = 0.55 and 0.60, Ns = 42 and 16), but Deary, Graham, and Maran (1992) found no significant positive correlations between expert ratings of surgical ability and intelligence test scores in trainee surgeons ($N = 22$).

Several studies have compared ability–performance correlations across different levels of surgical expertise. Wanzel et al. (2003) found that scores on two tests of "high-level" visuospatial ability (mental rotation and surface development) correlated significantly with expert ratings of surgical performance in dental students (*novices,* $n = 27$, avg. $r = 0.56$), but not in surgical residents (*intermediates,* $n = 12$) or staff surgeons (*experts,* $n = 8$). The correlations for the latter groups were not reported, but given the extremely small sample sizes here, they would not be significantly different from the novice correlation even if those correlations were assumed to be zero. Comparing groups of surgeons on a simulated videoscopic task, Keehner et al. (2004) found that a measure of visuospatial ability correlated significantly with mean skill rating in a low experience group ($r = 0.39$, $n = 48$), but not in a high experience group ($r = 0.02$, $n =$ 45). But, again, the correlations are not significantly different ($z = 1.83$, $p = 0.067$).

Gallagher, Cowie, Crothers, Jordan-Black, and Satava (2003) found that scores on a test of visuospatial ability in which participants recover three-dimensional structures from two-dimensional images correlated significantly and similarly with performance on a laparoscopic laboratory cutting task in two samples of novices (*r*s = 0.50 and 0.50, *n*s = 48 and 32) and in experienced surgeons (r = 0.54, n = 18). These correlations also do not differ across skill level. Enochsson et al. (2006) compared 18 resident and 11 expert surgeons in a simulated gastroscopy task, and found that correlations between scores on a test of visuospatial ability (card rotation test) and various metrics of performance in these very small samples were generally non-significant for both groups (avg. r = 0.06).

Murdoch, Bainbridge, Fisher, and Webster (1994) found that both manual dexterity and visuospatial ability correlated significantly with medical students' performance on microsurgical tasks (*r*s = -0.54 and 0.36, respectively). And in a sample of surgeons (N = 94), Risucci, Geiss, Gellman, Pinard, and Rosser (2001) found that measures of visuospatial ability correlated moderately (and 10/12 significantly) with performance on four surgical tasks (avg. r = -0.30; higher ability, faster performance); a measure of domain-specific experience correlated significantly with two of the performance measures (*r*s = 0.35 and 0.29), as did a measure of domain-specific knowledge (post-test examination; *r*s = 0.30 and 0.39). Groenier, Schraagen, Miedema, and Broeders (2014) examined the validity of tests of cognitive ability for predicting performance in a laparoscopic training simulator in medical students (N = 53) over 2 months. In univariate analyses, visuospatial ability, spatial memory, perceptual speed, and reasoning ability significantly predicted one performance measure (motion efficiency), whereas visuospatial ability and reasoning ability predicted another performance measure (duration). By contrast, in multivariate analyses, which controlled for correlations among the predictor variables, only one of the preceding effects was significant. The finding that univariate effects became non-significant in the multivariate analyses suggests that variance common to the ability measures (a *g* factor) may have been predictive of surgical performance.

More recently, Louridas and colleagues performed a meta-analysis of 52 studies on the relationship between various measures of cognitive ability and performance in laparoscopic, open, and endoscopic surgery (Louridas, Szasz, de Montbrun, Harris, & Grantcharov, 2016). Only a few cognitive ability measures positively predicted surgical performance across multiple studies, among them the mental rotation test, a pictorial surface orientation test, and the grooved pegboard test. Louridas et al. concluded that "no single test has been reported to reliably predict technical performance across the range of techniques and skills required of surgical trainees" (p. 689).

One other study fits in this category. In a sample (N = 428) that included professionals in exercise science-related jobs (e.g., physicians, trainers) as well as participants from the general population, Petushek, Cokely, Ward, and Myer (2015) found that two measures of cognitive ability had non-significant effects on performance in a task designed to assess risk of injury to the anterior cruciate ligament (ACL). By contrast, domain-specific factors (i.e., ACL knowledge and use of particular visual cues) were positive and statistically significant predictors of performance (r = 0.59 for ACL knowledge; Petushek, 2014).

Aviation

Several studies have tested for cognitive ability–performance correlations in aviation. In a sample of 86 pilots representing a wide range of experience and skill, along with 96 non-pilots, Morrow, Menard, Stine-Morrow, Teller, and Bryant (2001) found that a cognitive ability composite (working memory, perceptual speed, and visuospatial ability) positively predicted aviation-related performance (i.e., a composite reflecting accuracy in recalling and understanding air traffic control, ATC, commands), accounting for 29 percent of the variance. An expertise composite (ATC knowledge and flight hours) accounted for an additional 37 percent of the variance, but the Expertise × Cognitive Ability interaction was non-significant for all performance measures, indicating that the effect of cognitive ability on performance was similar across levels of expertise.

In a similar study of pilots (N = 91), Morrow et al. (2003) found that a cognitive ability composite accounted for an average of 22 percent of the variance in ATC tasks; an expertise composite accounted for an additional 28 percent of the variance, on average. (Expertise × Cognitive Ability interactions are not reported for this study.) Consistent with these findings, in a study of 97 licensed pilots with a wide range of flight experience, Taylor, O'Hara, Mumenthaler, Rosen, and Yesavage (2005) found that performance in an aviation communication task correlated significantly with working memory (r = 0.76), processing speed (r = 0.33), and interference control (r = 0.43), but interactions of expertise (flight rating) with these factors were all non-significant.

Using a sample with 25 novice and 25 expert pilots, Sohn and Doane (2003) found that working memory capacity predicted success in an aviation situational awareness task, but only in pilots who scored low on an aviation-specific test measuring skilled access to long-term memory (i.e., long-term working memory; Ericsson & Kintsch, 1995), as evidenced by a significant Long-Term Working Memory × Working Memory Capacity interaction. In a similar study, Sohn and Doane (2004) found that two measures of working memory capacity (spatial span and verbal span) correlated more strongly with situational awareness in 25 novice pilots (rs = 0.52, $p < 0.01$, and 0.30, respectively) than in 27 expert pilots (rs = 0.10 and 0.10, respectively). However, these correlations are not significantly different from each other across skill groups (z statistics < 1.7). Sohn and Doane (2004) did not test the Long-Term Working Memory × Working Memory Capacity interaction using the full sample (as in their earlier study), but instead tested it separately in each skill group, finding significance only in the expert group.

Finally, in a small sample of private pilots (N = 24), Causse, Dehais, and Pastor (2011) examined the relationship of broad cognitive abilities (reasoning and processing speed) and executive functions (working memory updating, set-shifting, and inhibition) to performance during a 45-minute flight simulator task. Reasoning ability correlated significantly with flight-path deviations ($rs = -0.63$); the other correlations were non-significant.

Job Performance

Measures of general cognitive ability positively predict job performance (Schmidt & Hunter, 2004), but do they remain valid predictors after extended job experience? This question has long been of interest to industrial–organizational psychologists. Using laboratory perceptual–motor tasks, Fleishman and colleagues demonstrated that general ability factors become less important with training, whereas task-specific factors become more important (e.g., Fleishman & Rich, 1963; see Hulin, Henry, & Noon, 1990, for other examples). However, the general finding from large-scale studies of actual work performance (as opposed to laboratory tasks) is that cognitive ability remains a significant predictor of job performance even after extensive job experience.

McDaniel (1986) investigated the impact of job experience on the validity of general cognitive ability using the General Aptitude Test Battery (GATB) database.[1] Compiled by the U.S. Employment Service in the 1970s, this database includes information on a large sample of civilian workers, including measures of job performance (i.e., supervisor ratings), job experience, and cognitive ability. McDaniel computed correlations between an "intelligence" score from the GATB (based on visuospatial, vocabulary, and arithmetic reasoning scores) and job performance across different levels of job experience. As shown in Figure 3.1, the correlations decrease somewhat as a function of job experience, but are still significant at the maximum amount of job experience (10+ years, $r = 0.20$, corrected $r = 0.29$).

Again using the GATB database, Farrell and McDaniel (2001) extended Ackerman's (1988) model of skill acquisition to job performance. Briefly, Ackerman hypothesized that involvement of different cognitive abilities in skill acquisition is moderated by the consistency of the task: When the demands of the task are consistent, meaning that the stimuli, rules, and sequences of action remain constant, automaticity can develop and the influence of general cognitive ability (reflecting attentional resources) on performance decreases with training. Meanwhile, the influence of perceptual speed increases and later decreases (i.e., an inverted U function), and the influence of psychomotor speed increases. To test this model, Farrell and McDaniel classified jobs as consistent or inconsistent using two different definitions of consistency: the complexity of the job (low complexity = consistent, high complexity = inconsistent) and tolerance for repetition required to perform the job (high tolerance for repetition = consistent, low tolerance for repetition = inconsistent). They then computed correlations between two cognitive composites (along with psychomotor speed) from the GATB (intelligence and perceptual speed) and job performance for different levels of job experience.

[1] We thank Michael McDaniel for sending us a copy of this study.

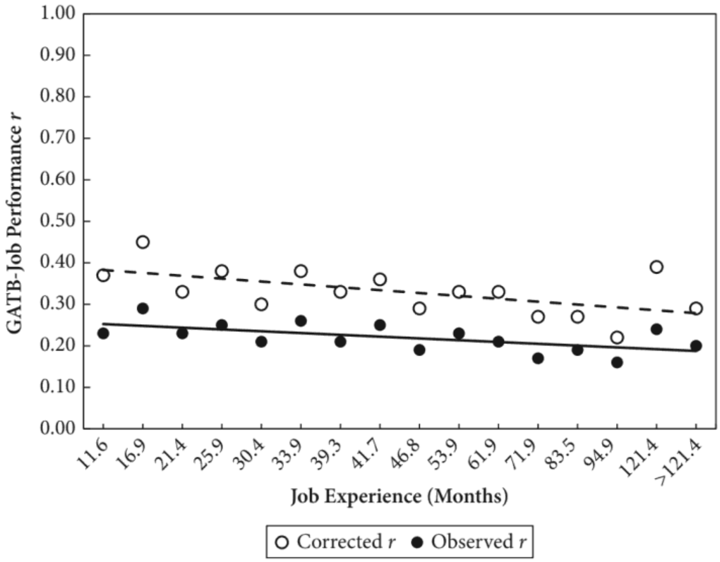


FIGURE 3.1 Correlations between GATB intelligence scores and job performance ratings as a function of job experience (total N = 16,058; across intervals, ns = 1,000 to 1,050, except for >121.4 months, n = 879). Solid circles represent observed (raw) correlations; open circles represent correlations after correction. Data from McDaniel, M. A., "The evaluation of a causal model of job performance: The interrelationships of general mental ability, job experience, and job performance," Tables 1 and 19, PhD thesis, George Washington University, Washington, D. C., 1986.

Support for Ackerman's model was mixed. For example, the intelligence correlations decreased as a function of job experience for low complexity jobs, but *increased* slightly for high-tolerance-for-repetition jobs. For the present discussion, the more important finding is simply that the cognitive ability factors significantly predicted job performance even at the maximum level of job experience: intelligence (avg. r = 0.25; avg. corrected r = 0.34) and perceptual speed (avg. r = 0.15; avg. corrected r = 0.20).

Studies of military personnel provide additional evidence that cognitive ability remains a significant predictor of job performance beyond initial training. Schmidt, Hunter, Outerbridge, and Goff (1988) tested for effects of cognitive ability and job experience on job performance in a sample of 1,474 soldiers in four jobs (armor repairman, armor crewman, supply specialist, and cook). Job performance was measured using work samples and supervisor ratings; cognitive ability was measured using the Armed Forces Qualification Test (AFQT) score from the ASVAB, which is based on Arithmetic Reasoning, Mathematics Knowledge, Paragraph Comprehension, and Word Knowledge subtests. (Job knowledge was also treated as a measure of job performance, though we think of it as a *predictor* of job performance.) Up to 5 years of job experience, correlations between AFQT scores and job performance were nearly constant. Across this span, correlations ranged from 0.38 to 0.42 for work samples and from 0.18 to 0.36 for supervisor ratings. Beyond 5 years of job experience (i.e., 61+ months), there was apparent convergence of ability groups for most measures, indicating a drop in validity beginning at 5 years. However, the average amount of job experience was actually much higher than 5 years in this group—from 9.5 years to 13 years, depending on the job. Moreover, only 1 of 12 AFQT × job experience interactions (work sample performance for armor crewman) was statistically significant, and it was not clearly interpretable as supporting convergence of the ability groups. Schmidt et al. concluded that "[a]t least out to 5 years, the validity of general mental-ability measures appears neither to decrease . . . nor to increase . . . . Instead, the validity remains relatively constant" (p. 56).

Wigdor and Green (1991) reported results of the Joint-Service Job Performance Measurement/Enlistment (JPM) Standards Project, a large study initiated in 1980 by the U.S. Department of Defense to develop measures of military job performance. Wigdor and Green reported that, across 23 jobs ($N = 7{,}093$ military personnel), the median correlation between AFQT scores and *hands-on* job performance (HOJP) was 0.26 (0.38 after correction for range restriction). They also reported mean hands-on performance for four AFQT categories (representing different levels of cognitive ability) as a function of job experience. As shown in Figure 3.2, mean differences among AFQT categories were largest at 0–12 months (about 10 points, or 1 *SD*), but still sizeable thereafter (5–6 points, or 0.50–0.60 *SD*). Wigdor and Green concluded that "the level of performance is positively related to AFQT score category at each of the four levels of job experience" (p. 163) and noted that "the lowest aptitude group never reaches the initial performance level of the highest aptitude group" (p. 163).

To further investigate cognitive ability–job performance relations, we obtained the JPM dataset.[2] The final dataset included 31 jobs and a total sample size of 10,088 military personnel. We performed three new analyses. First, we computed the AFQT–HOJP correlation across the job experience intervals used by Wigdor and Green (1991). As shown in Figure 3.3A, the correlations are as follows: 0–12 months ($r = 0.34$, $p < 0.001$, $n = 747$), 13–24 months ($r = 0.21$, $p < 0.001$, $n = 5{,}234$), 25–36 months ($r = 0.19$, $p < 0.001$, $n = 2{,}338$), and 37+ months ($r = 0.22$, $p < 0.001$, $n = 1{,}769$).[3] There is a statistically significant drop in the correlation from the first year of service to the second ($z = 3.60$, $p < 0.001$), but AFQT is still a statistically significant predictor of individual differences in HOJP after the first year of service. Second, capitalizing on the large data set, we broke the 37+-month group into additional experience intervals,

[2] We are grateful to Dr. Jane M. Arabian (Assistant Director, Accession Policy Office of the Under Secretary of Defense, The Pentagon, Washington, DC) for granting us permission to use the data, and to Dr. Rodney McCloy (Principal Scientist, Human Resources Research Organization, Louisville, KY) for sending us the data, with helpful notes.

[3] Prior to conducting statistical analyses, we screened the variables for values greater than $|3.5|$ *SD*s from the total sample mean (i.e., univariate outliers); 31 of the 30,264 values (0.1%) met this criterion, and we truncated these values to the $|3.5|$ *SD* cutoff value. There was one participant with zero months of job experience; prior to log transforming the job experience variable, we set this value to 0.03 months (1 day).

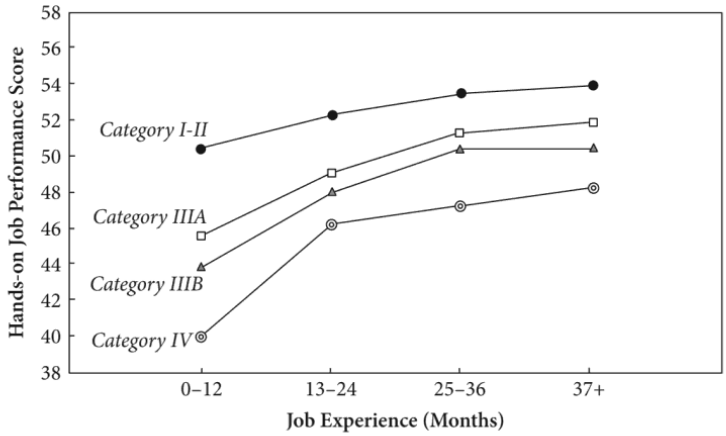


FIGURE 3.2 Mean Hands-on Job Performance Score by AFQT category (i.e., cognitive ability level). Percentile ranges for AFQT categories: I–II (65–99), IIIA (50–64), IIIB (31–49), and IV (10–30) (see Wigdor & Green, 1991, p. 53). Data from Wigdor, Alexandra K., and Green, Bert F., *Performance assessment for the workplace*, Volume 1, p. 53, Table 2.5, National Academy Press, 1991.

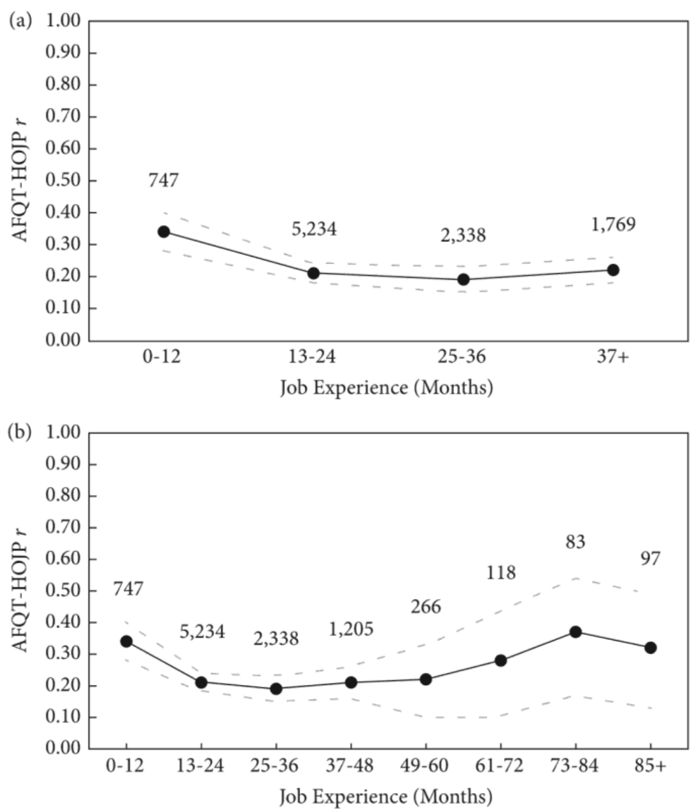


FIGURE 3.3 Correlations (with 95 percent confidence intervals) between AFQT scores and Hands-on Job Performance (HOJP) scores at 4 (A) and 8 (B) job experience intervals. Dashed lines are 95 percent confidence intervals; adjacent values are sample sizes. Data from Joint-Service Job Performance Measurement/Enlistment (JPM) Standards Project (N = 10,088).

to 85+ months (creating any more groups would result in small sample sizes, *ns* < 50). As shown in Figure 3.3B, the AFQT-HOJP correlation decreases from the first year to the second, stabilizes, and then increases—though the estimates become less precise as sample size decreases.

Finally, as the most statistically powerful analysis, we evaluated the Job Experience × AFQT interaction on HOJP using the entire data set via moderated multiple regression. (Prior to performing the regression analysis, we log-transformed job experience because it was non-normal, skewness = 2.40 and kurtosis = 9.56, and we mean-centered the predictors.) There were significant main effects of both AFQT ($\beta = 0.210$, $t = 21.92$, $p < 0.001$, part $r^2 = 0.044$) and log job experience ($\beta = 0.167$, $t = 17.37$, $p < 0.001$, part $r^2 = 0.028$) on HOJP. High levels of both AFQT and job experience were associated with higher HOJP. The AFQT × Log Job Experience interaction was also statistically significant and under-additive ($\beta = -0.023$, $t = -2.41$, $p = 0.016$, part $r^2 = 0.0005$), though the effect was virtually nil, indicating that AFQT was predictive of HOJP regardless of level of job experience (see Figure 3.4).

The overall picture to emerge from these large-scale employment studies is that cognitive ability remains a significant predictor of job performance, even after extensive job experience and even if validity drops initially. The question of how *far* beyond initial training cognitive ability predicts job performance is unanswered, but the results we have just reviewed indicate at least 5 years (Schmidt et al., 1986) to 10+ years (McDaniel, 1986). Reeve and Bonaccio (2011) reached a similar conclusion in their own review of the relationship between cognitive ability and job performance, noting

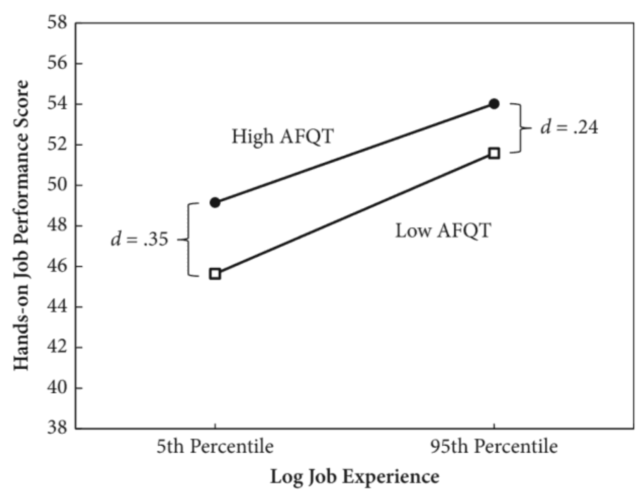


FIGURE 3.4 Predicted values for Hands-on Job Performance (HOJP) for low vs high AFQT (25th vs 75th percentile) at 5th vs 95th percentiles for Log Job Experience. Values generated using regression equation with AFQT score (mean-centered), Log Job Experience (mean-centered), and AFQT × Log Job Experience as predictors of HOJP score: HOJP score = 50.031 + 0.096(AFQT) + 8.119(Log Job Experience) + -0.052(AFQT × Log Job Experience). $N = 10{,}088$.

that "although validities might degrade somewhat over long intervals, we found no evidence to suggest that they degrade appreciably, thereby retaining practically useful levels of validity over very long intervals" (p. 269). Our analysis of the JPM data provide new support for this conclusion. Nevertheless, it remains possible that the validity of cognitive ability would drop to near zero over longer spans of time than have been examined in research (e.g., 20 years).

Discussion

What can be concluded about the role of cognitive ability in expertise? Table 3.2 summarizes findings from the expertise literature most directly relevant to the possibility of expertise-related mitigation of cognitive ability effects. These studies tested (or reported information to test) whether domain-specific factors mitigate effects of cognitive ability factors on domain-relevant performance, by either comparing ability-performance correlations across skill groups or testing interactions between domain-specific factors and cognitive ability factors. As shown, three studies provide evidence for expertise-related mitigation of cognitive ability effects and 10 studies do not; the results of two other studies are mixed or unclear. Based on sample size alone, the Burgoyne et al. (2016) chess meta-analysis might be seen as the best evidence for mitigation, but we reiterate that expertise level (i.e., ranked or unranked) was highly confounded with both age and type of skill measure. This evidence certainly does not warrant any strong conclusions about expertise-related mitigation of the effects of cognitive ability.

What can be made of these results? Unfortunately, not much, because studies in the expertise literature differ in methodological/design characteristics such as sample size, type of criterion task, tests used to measure cognitive ability factors, and use of single versus composite measures to index cognitive ability factors. Any (or all) of these differences could explain differences across studies in cognitive ability–performance relationships. With some of our own studies as examples (e.g., Meinz & Hambrick, 2010, $N = 57$), it is worth emphasizing that sample sizes in this literature are often very small for research on individual differences (see Table 3.2), leading not only to low statistical power but low precision. Consequently, it is not surprising when results do not replicate (it is, in fact, often more surprising when they do). Note also that correlations in the expertise literature are seldom corrected for measurement error and/or restriction of range, resulting in systematic underestimates of the true magnitude of underlying relationships (see McAbee & Oswald, 2017).

A more consistent picture emerges from large-scale studies of job performance. Though validity may drop somewhat initially, measures of cognitive ability significantly predict job performance well beyond initial training. Expertise research often focuses on a specific aspect or component of performance in a domain (e.g., flight path prediction, poker hand evaluation); job performance research more often uses global

Table 3.2 Summary of evidence for expertise-related mitigation of cognitive ability effects

Study	Domain	Sample Size		Cognitive Factor[b]	Evidence for Mitigation?	Empirical Test[c]
		N	Group n[a]			
Ceci & Liker (1986)	Handicapping	30	16/14	IQ	Unclear	Correlations
Masunaga and Horn (2001)	Go	263	62/89/92/23	Gf, Gc, PS	No	Correlations
Morrow et al. (2001)	Aviation	182	96/86	WMC, PS, VS	No	Interaction
Gallagher et al. (2003)	Surgery	98	48, 32/18	VS	No	Correlations
Sohn & Doane (2003)	Aviation	50	25/25	WMC	Yes	Interaction
Wanzel et al. (2003)	Surgery	47	27/12/8	VS	No	Correlations
Keehner et al. (2004)	Surgery	93	48/45	VS	No	Correlations
Sohn & Doane (2004)	Aviation	52	25/27	WMC	Mixed	Correlations
Taylor et al. (2005)	Aviation	97	25/53/19	WMC, PS, AC	No	Interaction
Enochsson et al. (2006)	Surgery	29	18/11	VS	No	Correlations
Ruthsatz et al. (2008)	Music	261	178/19/64	Gf	No	Correlations
Meinz & Hambrick (2010)	Music	57	NA	WMC	No	Interaction
Meinz et al. (2012)	Poker	155	NA	WMC	No	Interaction
Hambrick et al. (2012)	Geology	67	NA	VS	Yes	Interaction
Burgoyne et al. (2016)[d]	Chess	1,604	1,267/337	Gf	Yes	Correlations

Note. Studies are listed in chronological order.

[a] The skill group *n* values are listed in order of increasing expertise. NA for skill group *n* indicates that expertise was treated only as a continuous variable.

[b] Gf, fluid intelligence; Gc, crystallized intelligence; WMC, working memory capacity; PS, perceptual speed; VS, visuospatial ability; AC, attentional control.

[c] In the *interaction* test, mitigation is tested by evaluating the statistical interaction between a domain-specific factor and a cognitive ability factor. In the *correlations* test, mitigation is tested by testing for a difference in correlations between a cognitive ability factor and performance across groups representing different levels of a domain-specific factor.

[d] Meta-analysis.

measures of performance (e.g., overall supervisory ratings, total work sample scores). It could be that involvement of cognitive ability factors decreases as a function of skill in some components of a complex task or job but not in others (e.g., consistent but not variable components; Ackerman, 1992). This is one possible explanation for why correlations between cognitive ability and job performance may drop somewhat with job experience yet still remain statistically significant.

Before proceeding, we note that when cognitive ability and domain-specific factors are measured in the same study, the latter generally account for more variance in expertise than the former (see Ward et al., 2017, for examples). At the same time, cognitive ability and domain-specific knowledge cannot generally be assumed to be independent. For example, Schmidt, Hunter, and Outerbridge (1986) found a correlation of 0.46 between AFQT scores and job knowledge. One interpretation of this finding is that measures of cognitive ability (e.g., IQ, working memory capacity) capture basic mental processes involved in acquiring information in learning situations (Ackerman, 1996; Cattell, 1971; Jensen, 1998). Moreover, even if domain-specific factors explain far more of the variance in expertise than domain-general factors, this does not preclude the latter from being practically useful. We examine this issue next.

Potential Uses of Cognitive Ability Measures to Accelerate Acquisition of Expertise

There are two major ways that cognitive ability measures might be used in efforts to accelerate acquisition of expertise. The first is for personnel selection and classification. That is, cognitive ability measures might be used to make hiring decisions and to assign employees to jobs once hired. The second area of application is in the design of training programs. If a certain cognitive ability factor (e.g., attentional control) is found to be a significant predictor of performance in a domain, then designing training to augment or *bootstrap* that ability (e.g., prompts to direct attention to task-relevant information) might be particularly beneficial for individuals lower in the ability (though see Hoffman et al., 2014, for a cautionary note about removing *desirable difficulties* from training).

But how large must a validity coefficient for a cognitive ability test be to justify its use for these applications? What qualifies as a practically significant effect? Given real-world outcomes (vs outcomes that do not generalize easily to the real world), moderate correlations can prove to be very important. Moreover, although variance explained (r^2) may be of theoretical interest to researchers (e.g., Macnamara, Hambrick, & Oswald, 2014), it is r and not r^2 that is an index of the direct relationship or the utility of a measure in terms of prediction (see Schmidt, Hunter, McKenzie, & Muldrow, 1979). As Kuncel and Hezlett (2010) commented:

> Moderate relationships between predictors and criteria often are inappropriately discounted. For example, correlations of .30 have been dismissed as accounting for less than 10% of the variance in the criteria. However, this relationship is sufficiently large that hiring or admitting individuals who score better on the test can double the rate of successful performance. (p. 340)

This point was made nearly 80 years ago by Taylor and Russell (1939), who noted that interpreting the practical importance of correlation coefficients based on methods involving r^2

has led to some unwarranted pessimism on the part of many persons concerning the practical usefulness in an employment situation of validity coefficients in the range of those usually obtained. We believe that it may be of value to point out the very considerable improvement in selection efficiency which may be obtained with small correlation coefficients. (p. 571)

To that end, Taylor and Russell published a set of easy-to-use tables to determine the benefits of using selection tests of different validities in employment settings (see Law & Myors, 1993, for an automated approach). Three pieces of information are needed to use the tables: (3.1) the base rate of success in a job (i.e., the proportion of people who currently succeed in a job), (3.2) the selection ratio for the job (i.e., the ratio of applicants who are selected), and (3.3) the validity of the test. With these three pieces of information, one can consult a Taylor–Russell table and find the predicted improvement in using the test for selection versus not using it.

Table 3.3 gives an example where the base rate of success is 0.20. As shown, when the selection ratio is low, even a selection test with modest validity will lead to a substantial improvement in employee performance over the base rate. For example, if the selection ratio is 0.10, use of a test with validity of 0.20 to select applicants will lead to an 11 percent improvement over not using the test (or 17 percent for a test with validity of 0.30). However, if the selection ratio is high, even a test with high validity will yield little benefit. For example, if the selection ratio is 0.90, then even the use of a test with validity of 0.80 would lead to an improvement of only 2 percent. More sophisticated approaches to utility analysis have been developed since Taylor and Russell published their tables (see Hunter & Schmidt, 1996; Schmidt, Hunter, Outerbridge, & Trattner, 1986), but suffice it to say that use of a test with a moderate level of validity can be practically useful.

Table 3.3 Example of Taylor–Russell Utility Table

Base Rate of Success = 0.20

	Selection Ratio[a]								
Validity (*r*)	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0.00	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
0.10	0.25	0.24	0.23	0.23	0.22	0.22	0.21	0.21	0.21
0.20	0.31	0.28	0.27	0.26	0.25	0.24	0.23	0.22	0.21
0.30	0.37	0.33	0.30	0.28	0.27	0.25	0.24	0.23	0.21
0.40	0.44	0.38	0.34	0.31	0.29	0.27	0.25	0.23	0.22
0.50	0.52	0.44	0.38	0.35	0.31	0.29	0.26	0.24	0.22
0.60	0.60	0.50	0.43	0.38	0.34	0.30	0.27	0.24	0.22
0.70	0.69	0.56	0.48	0.41	0.36	0.31	0.28	0.25	0.22
0.80	0.79	0.64	0.53	0.45	0.38	0.33	0.28	0.25	0.22
0.90	0.91	0.75	0.60	0.48	0.40	0.33	0.29	0.25	0.22

Note. Validity (*r*): correlation between predictor variable and criterion variable.

[a] Selection ratio: proportion of applicants who are hired. Values in the cells of the table indicate incremental validity, i.e., expected rate of success as a result of using the selection test.

	Training Outcome		
Test Used to Select?	Fail	Pass	Total
No	60	40	100
Yes	40	60	100
Total	100	100	

FIGURE 3.5 Example of binomial effect size display (BESD) relevant to expertise research and application.

Rosenthal and Rubin's (1982) binomial effect size display (BESD) provides another way to indicate the practical significance of an effect size of a given magnitude (see Ward et al., 2007, for an example of how the BESD can be used in expertise research). Displaying the difference between two proportions (e.g., treatment vs no treatment; selection test vs no selection test), like the Taylor–Russell tables the BESD reveals that modest effect sizes can be practically important. For example, Rosenthal (2005) explained that "an *r* of 0.20 is said to account for 'only 4% of the variance', but the BESD shows that this proportion of variance accounted for is equivalent to increasing the success rate . . . from 40 to 60%." Figure 3.5 illustrates this point in terms of a hypothetical scenario where 100 individuals in an organization must be selected for a training program. In one case, a selection test with validity of 0.20 is used; in the other case, it is not used. As shown, using the selection test increases the chances that a trainee will pass the training program by 20 percent (i.e., 20 more people out of 100 pass), even though the scores on the test account for only 4 percent of the variance in the outcome.

Ethical Considerations

There are ethical issues associated with use of any psychological test to make decisions that affect people's lives (e.g., hiring decisions). Probably everyone would agree that it is unethical (not to mention legally unwise) to select individuals using a test with no demonstrated validity, but consider a situation where a test has modest validity—say, 0.30. One might argue that because the validity coefficient is far from perfect, it is unethical to use the test for selection because a considerable number of people with lower scores would be expected to succeed. However, one might also argue that *not* using the test for selection is unethical because lower-scoring individuals will be at a relatively high risk for failure, which may have adverse consequences for the individual (e.g., negative perceptions of other employees, lowered self-efficacy) and also the organization. Along with conducting a proper job analysis and validity study, any organization wishing to use a cognitive test for making personnel decisions must consider these sorts of ethical questions before putting the test into use (Landy & Conte, 2013).

Conclusions

Psychologists have long been interested in identifying traits that may help to explain individual differences in expertise (Hambrick, Campitelli, & Macamara, 2017). Here, we reviewed evidence for the contribution of cognitive ability. There is ample evidence that cognitive ability positively predicts individual differences in complex task performance early in training, but it is unclear whether it remains predictive after extensive practice or training. Evidence from research on traditional domains for expertise research (e.g., chess, music) is inconsistent. For some tasks, domain-specific factors may mitigate the effect of cognitive ability factors on performance (e.g., maintaining situational awareness in aviation), but for other tasks, this may not be the case (e.g., sight-reading music). Evidence from research on job performance is more consistent in indicating that measures of cognitive ability are predictive of job performance, well beyond initial training. In light of this evidence, we believe that at a broad level combining optimal procedures for training complex skills (Hoffman et al., 2014) with valid selection procedures holds tremendous promise for accelerating acquisition of expertise.

At a theoretical level, we believe that it is imperative for expertise researchers to develop and test formal models of expertise. Research on the involvement of cognitive ability factors in expertise has often proceeded somewhat haphazardly, with no general theory describing how mechanisms underlying performance differ across domains. There is no better illustration of this critical point than our own work. We have conducted a number of one-off studies—one in piano sight-reading (Meinz & Hambrick, 2010), another in geological bedrock mapping (Hambrick et al., 2012), another in Texas Hold 'em poker (Meinz et al., 2012)—with no theory to account for how results differ across these domains. Moving ahead, theories of expertise should draw on existing theoretical frameworks to identify potential predictors of expertise (e.g., Ackerman, 1996; Ericsson et al., 1993; Gagné, 2017). However, guided by both computational models (e.g., Altmann, Trafton, & Hambrick, 2014) and cognitive task analysis (Chipman, Schraagen, & Shalin, 2000), they must also specify the information processing mechanisms underlying performance in different types of tasks. Otherwise, there will continue to be no solid basis for comparing results across tasks, and evidence will remain fragmentary.

In the spirit of Hoffman et al.'s (2014) recommendations, we believe that it is also critical that expertise research expand beyond highly constrained activities such as chess, music, and sports, to messy real-world tasks in which the requirements of a job can change rapidly with technological developments and there is no well-circumscribed body of knowledge (as there is in, say, chess). We think that measures of cognitive ability factors hypothesized to underlie *adaptability* (e.g., attentional control, working memory capacity) may have particular promise for predicting performance in jobs such as these. These measures are also attractive because some research has suggested they may reduce group differences (e.g., by race/ethnicity) and resultant adverse impact in selection while still achieving high validity (Verive & McDaniel, 1996). More generally, we are optimistic that the scientific knowledge that will accumulate through programmatic research on individual differences in expertise has great potential to inform efforts to accelerate the acquisition of societally important skills.

Acknowledgements

We thank Paul Ward and Jan Maarten Schraagen for their comments on an earlier version of this chapter.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General 117*, 288–318.

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology 77*, 598–614.

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence 22*, 227–257.

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2014). Momentary interruptions can derail the train of thought. *Journal of Experimental Psychology: General 143*, 215–226.

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science 18*, 89–94.

Bayer, A. E., & Folger, J. (1966). Some correlates of a citation measure of productivity in science. *Sociology of Education 39*, 381–390.

Berri, D. J., & Simmons, R. (2011). Catching a draft: On the process of selecting quarterbacks in the National Football League amateur draft. *Journal of Productivity Analysis 35*, 37–49.

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2018). The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence 59*, 72–83.

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2018). Corrigendum: The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence*.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Oxford, UK: Houghton Mifflin.

Causse, M., Dehais, F., & Pastor, J. (2011). Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots. *International Journal of Aviation Psychology 21*, 217–234.

Ceci, S. J., & Liker, J. K. (1986). A day at the races: A study of IQ, expertise, and cognitive complexity. *Journal of Experimental Psychology: General 115*, 255–266.

Chipman, S. F., Schraagen, J. M., & Shalin, V. L. (2000). Introduction to cognitive task analysis. In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Eds), *Cognitive task analysis* (pp. 3–23). New York: Lawrence Erlbaum.

Clark, M. J., & Centra, J. A. (1982). *Conditions influencing the career accomplishments of Ph. D.s.* (GREB Research Report No. 76-2R). Princeton, NJ: Educational Testing Service.

Creager, J. A., & Harmon, L. R. (1966). *On-the-job validation of selection variables*. Office of Scientific Personnel, National Academy of Sciences-National Research Council.

Deary, I. J., Graham, K. S., & Maran, A. G. (1992). Relationships between surgical ability ratings and spatial abilities and personality. *Journal of the Royal College of Surgeons of Edinburgh 37*, 74–79.

de Bruin, A. B., Kok, E. M., Leppink, J., & Camp, G. (2014). Practice, intelligence, and enjoyment in novice chess players: A prospective study at the earliest stage of a chess career. *Intelligence 45*, 18–25.

de Groot, A.D. (1965/1978). *Thought and choice in chess*. Amsterdam: Amsterdam Academic Archive.

Detterman, D. K. (2014). Introduction to the intelligence special issue on the development of expertise: Is ability necessary? *Intelligence 45*, 1–5.

Detterman, D. K., & Spry, K. M. (1988). Is it smart to play the horses? Comment on "A day at the races: A study of IQ, expertise, and cognitive complexity" (Ceci & Liker, 1986). *Journal of Experimental Psychology: General 117*, 91–95.

Engelhardt, L. E., Mann, F. D., Briley, D. A., Church, J. A., Harden, K. P., & Tucker-Drob, E. M. (2016). Strong genetic overlap between executive functions and intelligence. *Journal of Experimental Psychology: General 145*, 1141–1159.

Enochsson, L., Westman, B., Ritter, E. M., Hedman, L., Kjellin, A., Wredmark, T., & Felländer-Tsai, L. (2006). Objective assessment of visuospatial and psychomotor ability and flow of residents and senior endoscopists in simulated gastroscopy. *Surgical Endoscopy and Other Interventional Techniques 20*, 895–899.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 683–703). New York: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review 102*, 211–245.

Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review 100*, 363–406.

Ericsson, K. A., & Smith, J. (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge, UK: Cambridge University Press.

Farrell, J. N., & McDaniel, M. A. (2001). The stability of validity coefficients over time: Ackerman's (1988) model and the General Aptitude Test Battery. *Journal of Applied Psychology 86*, 60–79.

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Oxford: Brooks/Cole.

Fleishman, E. A., & Rich, S. (1963). Role of kinesthetic and spatial-visual abilities in perceptual-motor learning. *Journal of Experimental Psychology 66*, 6–11.

Folger, J. K., Astin, H. S., & Bayer, A. E. (1970). *Human resources and higher education: Staff report on the commission on human resources and advanced education*. New York, NY: Russell Sage Foundation.

Gagné, F. (2017). Expertise development from an IMTD perspective. In D. Z. Hambrick, G. Campitelli, & B. N. Macnamara (Eds.), *The science of expertise: Behavioral, neural, and genetic approaches to complex skill* (pp. 307–327). New York: Routledge.

Gallagher, A. G., Cowie, R., Crothers, I., Jordan-Black, J. A., & Satava, R. M. (2003). PicSOr: an objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopopic performance. *Surgical Endoscopy 17*, 1468–1471.

Gibbons, R. D., Baker, R. J., & Skinner, D. B. (1986). Field articulation testing: A predictor of technical skills in surgical residents. *Journal of Surgical Research 41*, 53–57.
Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence 24*, 13–23.
Grabner, R. H., Stern, E., & Neubauer, A. C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica 124*, 398–420.
Groenier, M., Schraagen, J. M. C., Miedema, H. A., & Broeders, I. A. J. M. (2014). The role of cognitive abilities in laparoscopic simulator training. *Advances in Health Sciences Education 19*, 203–217.
Haier, R. J. (2016). *The neuroscience of intelligence*. Cambridge, UK: Cambridge University Press.
Hambrick, D. Z., & Campitelli, G., & Macamara, B. N. (2017). Introduction: A brief history of the science of expertise and overview of the book. In D. Z. Hambrick, G. Campitelli, & B. N. Macamara (Eds.), *The science of expertise: Behavioral, neural, and genetic approaches to complex skill* (pp. 1–10). New York: Routledge.
Hambrick, D. Z., Libarkin, J. C., Petcovic, H. L., Baker, K. M., Elkins, J., Callahan, C. N., . . . , & LaDue, N. D. (2012). A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology: General, 141*, 397–403.
Hambrick, D. Z., & Meinz, E. J. (2011). Limits on the predictive power of domain-specific experience and knowledge in skilled performance. *Current Directions in Psychological Science 20*, 275–279.
Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D. H. (2014). *Accelerated expertise: Training for high proficiency in a complex world*. New York: Psychology Press.
Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships. *Psychological Bulletin 107*, 328–340.
Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law 2*, 447–472.
Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Greenwood Publishing Group.
Jung, W. H., Kim, S. N., Lee, T. Y., Jang, J. H., Choi, C. H., Kang, D. H., & Kwon, J. S. (2013). Exploring the brains of Baduk (Go) experts: Gray matter morphometry, resting-state functional connectivity, and graph theoretical analysis. *Frontiers in Human Neuroscience 7*, 633.
Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York: Oxford University Press.
Kaufman, H. G. (1972). Relations of ability and interest to currency of professional knowledge among engineers. *Journal of Applied Psychology 56*, 495–499.
Keehner, M. M., Tendick, F., Meng, M. V., Anwar, H. P., Hegarty, M., Stoller, M. L., & Duh, Q. Y. (2004). Spatial ability, experience, and skill in laparoscopic surgery. *American Journal of Surgery 188*, 71–75.
Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2016). *Behavioral genetics* (7th ed). New York: Macmillan Higher Education.
Kopiez, R., & In Lee, J. (2008). Towards a general model of skills involved in sight reading music. *Music Education Research 10*, 41–62.

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science 315*(5815), 1080–1081.

Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science 19*, 339–345.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychological Bulletin 127*, 162–181.

Landy, F. J., & Conte, J. M. (2013). *Work in the 21st century: An introduction to industrial and organizational psychology* (4th ed). London: Wiley.

Law, K. S., & Myors, B. (1993). Cutoff scores that maximize the total utility of a selection program: Comment on Martin and Raju's (1992) procedure. *Journal of Applied Psychology 78*, 736–740.

Lee, B., Park, J. Y., Jung, W. H., Kim, H. S., Oh, J. S., Choi, C. H., . . . , & Kwon, J. S. (2010). White matter neuroplastic changes in long-term trained players of the game of "Baduk" (GO): A voxel-based diffusion-tensor imaging study. *Neuroimage 52*, 9–19.

Leonard, C. A., & Williams, R. J. (2015). Characteristics of good poker players. *Journal of Gambling Issues* 31, 45–68.

Louridas, M., Szasz, P., de Montbrun, S., Harris, K. A., & Grantcharov, T. P. (2016). Can we predict technical aptitude?: A systematic review. *Annals of Surgery 263*, 673–691.

Lubinski, D. (2009). Exceptional cognitive ability: The phenotype. *Behavior Genetics 39*, 350–358.

Lyons, B. D., Hoffman, B. J., & Michel, J. W. (2009). Not much more than g? An examination of the impact of intelligence on NFL performance. *Human Performance 22*, 225–245.

McAbee, S. T., & Oswald, F. L. (2017). Primer—Statistical methods in the study of expertise: Some key considerations. In D. Z. Hambrick, G. Campitelli, & B. N. Macnamara (Eds.), *The science of expertise: Behavioral, neural, and genetic approaches to complex skill* (pp. 13–30). New York: Routledge.

McCabe, D. P., Roediger III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology 24*, 222–243.

McDaniel, M. A. (1986). *The evaluation of a causal model of job performance: The interrelationships of general mental ability, job experience, and job performance* (Unpublished doctoral dissertation). George Washington University, Washington, DC.

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science 25*, 1608–1618.

Mann, D. T., Williams, A. M., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport and Exercise Psychology 29*, 457–478.

Masunaga, H., & Horn, J. (2001). Expertise and age-related changes in components of intelligence. *Psychology and Aging 16*, 293–311.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill the role of working memory capacity. *Psychological Science 21*, 914–919.

Meinz, E. J., Hambrick, D. Z., Hawkins, C. B., Gillings, A. K., Meyer, B. E., & Schneider, J. L. (2012). Roles of domain knowledge and working memory capacity in components of skill in Texas Hold'Em poker. *Journal of Applied Research in Memory and Cognition 1*, 34–40.

Morrow, D. G., Menard, W. E., Ridolfo, H. E., Stine-Morrow, E. A. L., Teller, T., & Bryant, D. (2003). Expertise, cognitive ability, and age effects on pilot communication. *International Journal of Aviation Psychology 13*, 345–371.

Morrow, D. G., Menard, W. E., Stine-Morrow, E. A., Teller, T., & Bryant, D. (2001). The influence of expertise and task factors on age differences in pilot communication. *Psychology and Aging 16*, 31–46.

Murdoch, J. R., Bainbridge, L. C., Fisher, S. G., & Webster, M. H. (1994). Can a simple test of visual-motor skill predict the performance of microsurgeons? *Journal of the Royal College of Surgeons of Edinburgh 39*, 150–152.

Petushek, E. J. (2014). *Development and validation of the anterior cruciate ligament injury-risk-estimation quiz (ACL-IQ)*. Retrieved from ProQuest Dissertations & Theses Global. (Accession No. 3643834).

Petushek, E. J., Cokely, E. T., Ward, P., & Myer, G. D. (2015). Injury risk estimation expertise: cognitive-perceptual mechanisms of ACL-IQ. *Journal of Sport and Exercise Psychology 37*, 291–304.

Pfau, H. D., & Murphy, M. D. (1988). Role of verbal knowledge in chess skill. *American Journal of Psychology 101*, 73–86.

Reeve, C. L., & Bonaccio, S. (2011). On the myth and the reality of the temporal validity degradation of general mental ability test scores. *Intelligence 39*, 255–272.

Risucci, D., Geiss, A., Gellman, L., Pinard, B., & Rosser, J. (2001). Surgeon-specific factors in the acquisition of laparoscopic surgical skills. *American Journal of Surgery 181*, 289–293.

Rosenthal, R. (2005). Binomial effect size display. *Encyclopedia of statistics in behavioral science*. Wiley StatsRef: Statistics Reference Online.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology 74*, 166–169.

Ruthsatz, J., Detterman, D., Griscom, W. S., & Cirullo, B. A. (2008). Becoming an expert in the musical domain: It takes more than just practice. *Intelligence 36*, 330–338.

Sala, G., Burgoyne, A. P., Macnamara, B. N., Hambrick, D. Z., Campitelli, G., & Gobet, F. (2017). Checking the "Academic Selection" argument. Chess players outperform non-chess-players in cognitive skills related to intelligence: A meta-analysis. *Intelligence 61*, 130–139.

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch (Ed.), *Psychology of music* (3rd ed, pp. 499–550). London: Academic Press.

Schmidt, F. L., Hunter, J. E., McKenzie, R. W., & Muldrow, T. W. (1979). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology* 64, 609–626.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin 124*, 262–274.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology 86*, 162–173.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology 71*, 432–439.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. *Journal of Applied Psychology 73*, 46–57.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Trattner, M. H. (1986). The economic impact of job selection methods on size, productivity, and payroll costs of the federal work force: An empirically based demonstration. *Personnel Psychology 39*, 1–29.

Schueneman, A. L., Pickleman, J., Hesslein, R., & Freeark, R. J. (1984). Neuropsychologic predictors of operative skill among general surgery residents. *Surgery 96*, 288–295.

Sohn, Y. W., & Doane, S. M. (2003). Roles of working memory capacity and long-term working memory skill in complex task performance. *Memory & Cognition 31*, 458–466.

Sohn, Y. W., & Doane, S. M. (2004). Memory processes of flight situation awareness: Interactive roles of working memory capacity, long-term working memory, and expertise. *Human Factors 46*, 461–475.

Takagi, H. (1997). *Cognitive aging: expertise and fluid intelligence* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 9733148).

Taylor, J. L., O'Hara, R., Mumenthaler, M. S., Rosen, A. C., & Yesavage, J. A. (2005). Cognitive ability, expertise, and age differences in following air-traffic control instructions. *Psychology and Aging 20*, 117–133.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology 23*, 565–578.

Toma, M., Halpern, D. F., & Berger, D. E. (2014). Cognitive abilities of elite nationally ranked SCRABBLE and crossword experts. *Applied Cognitive Psychology 28*, 727–737.

Tucker-Drob, E. M. (2011). Global and domain-specific changes in cognition throughout adulthood. *Developmental Psychology 47*, 331–343.

Tuffiash, M., Roring, R. W., & Ericsson, K. A. (2007). Expert performance in SCRABBLE: implications for the study of the structure and acquisition of complex skills. *Journal of Experimental Psychology: Applied 13*, 124–134.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science 9*, 160–164.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2015). Working memory delay activity predicts individual differences in cognitive abilities. *Journal of Cognitive Neuroscience 27*, 853–865.

Unterrainer, J. M., Kaller, C. P., Halsband, U., & Rahm, B. (2006). Planning abilities and chess: A comparison of chess and non-chess players on the Tower of London task. *British Journal of Psychology 97*, 299–311.

Unterrainer, J. M., Kaller, C. P., Leonhart, R., & Rahm, B. (2011). Revising superior planning performance in chess players: The impact of time restriction and motivation aspects. *American Journal of Psychology 124*, 213–225.

Verive, J. M., & McDaniel, M. A. (1996). Short-term memory tests in personnel selection: Low adverse impact and high validity. *Intelligence 23*, 15–32.

Vestberg, T., Gustafson, R., Maurex, L., Ingvar, M., & Petrovic, P. (2012). Executive functions predict the success of top-soccer players. *PloS One 7*, e34731.

Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S., & Roberts, B. (2010). Are expert athletes "expert" in the cognitive laboratory? A meta-analytic review of cognition and sport expertise. *Applied Cognitive Psychology 24*, 812–826.

Wanzel, K. R., Hamstra, S. J., Caminiti, M. F., Anastakis, D. J., Grober, E. D., & Reznick, R. K. (2003). Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. *Surgery 134*, 750–757.

Ward, P., Belling, P., Petushek, E., & Ehrlinger, J. (2017). Does talent exist? A re-evaluation of the nature–nurture debate. In J. Baker, S. Cobley, J. Schorer, & N. Wattie (Eds), *Routledge handbook of talent identification and development in sport* (pp. 19–34). New York: Routledge.

Ward, P., Hodges, N. J., Starkes, J. L., & Williams, M. A. (2007). The road to excellence: Deliberate practice and the development of expertise. *High Ability Studies 18*, 119–153.

Wigdor, A. K., & Green, B. F. (1991). *Performance assessment for the workplace*, Volumes I & II. Washington, DC: National Academy Press.

CHAPTER 4

STUDIES OF EXPERTISE AND EXPERIENCE

A Sociological Perspective on Expertise

HARRY COLLINS AND ROBERT EVANS

Introduction

Is expertise a status that is attributed to someone or a capacity a person possesses? There is merit in both perspectives but here we describe an approach known as studies of expertise and experience (SEE) that focuses on the second and treats expertise as real rather than relational (Collins & Evans, 2002, 2007). This realist conception of expertise differs from many psychological and philosophical approaches, in that it turns on acquiring expertise through socialization and can be used whether this socialization is ubiquitous—as in native language speaking—or esoteric—as in a scientific specialty. The idea that expertise is acquired through socialization makes SEE entirely compatible with social constructivist theories of knowledge. For example, a natural language is a social construct but speaking it is a deep and difficult expertise that can be acquired only through socialization.

The origins of SEE are to be found in the field of science and technology studies (STS), where science is seen as a *hard case* for the study of knowledge and expertise more generally. We therefore begin our exposition of the approach with a summary of the debates within STS that led to its development. We then examine how the ideas that inform SEE relate to wider debates about embodiment and tacit knowledge. Next we set out some novel theoretical and methodological consequences that follow from this approach: a new classification of expertise, summarized in a *Periodic Table of Expertises*, and the creation of a new research method known as the *Imitation Game*. We conclude by setting out the future questions for research using SEE and illustrate its utility and relevance through the examples of interdisciplinary working and the role of experts in a *post-truth* society.

Three Waves of STS

STS is a diverse field whose history can be told in many ways (see Felt, Fouche, Miller, & Smith-Doerr, 2017 for an illustration of this diversity). Here we simplify the history by distinguishing between *three waves* of STS that illustrate the key transitions in its understanding of science and, over time, of expertise more generally (Collins & Evans, 2002). The first wave of science studies is characterized by the idea that science is, without question, an epistemically superior form of knowledge-making that is different to other kinds of expert knowledge. To the extent that the social analysis of science is possible, its questions concern the institutional structures that support and facilitate science. To try to explain the content of valid scientific knowledge by social analysis makes no sense under wave one, though scientific mistakes can be explained this way.

Wave one dominated the sociology of science from the early part of the twentieth century with Robert Merton being its most important theorist (e.g., Merton, 1973). In the early 1970s, however, the sociology of science, like many other fields in the social sciences, began to take an increasingly constructivist approach. Rather than seeing science as passively *holding up a mirror to nature*, sociologists began treating scientific knowledge as actively created through agreements made by scientists about how to interpret what they saw in the many different mirrors they had made (Collins, 1975, 1985/1992; Latour & Woolgar, 1979). Wave two thus rejected the asymmetry of wave one, in which scientific mistakes alone are open to sociological explanation, and advocated a *symmetrical* approach in which both scientific truth and scientific error are treated as the same kind of social agreement (Barnes, Bloor, & Henry, 1996; Bloor, 1973, 1991).

Under this second wave of STS, expertise becomes a relational phenomenon in which being an expert amounts to being granted expert status by others. For sociologists, this means expertise is an *actors' category* in that it reflects the success or otherwise of particular social groups in attaining positions of credibility and authority (Collins, 2008). It also makes scientific expertise look a lot less special, as the difference between science and other forms of expertise is no longer an epistemic difference but a sociological one: scientific expertise is different because it is treated differently, not because it is made differently. This approach has been enormously successful, producing rich descriptions of how knowledge is created and shared, as well as revealing the ways in which elite institutions can use science to marginalize the concerns raised by others (Carr, 2010 provides a summary and overview).

The third wave of STS endorses the symmetrical description of science provided by wave two but questions the policy conclusions drawn by many of its advocates (e.g., Funtowicz & Ravetz, 1993; Jasanoff, 2003, 2013; Wynne, 2003). According to the relational view of wave two, the symmetrical analysis of science shows that it can never be neutral and that it invariably reflects the dominant interests in that society.

Improving policy-making under wave two thus means challenging the institutions and assumptions that grant authority to science and deny legitimacy to others. The result is that the policy recommendations of wave two always push in the same direction: democratize expertise by opening up esoteric questions to ever wider groups of participants. This works well when the problem of expertise is that of technocracy in which it is mistakenly thought that those who possess an esoteric expertise either exhaust what can be known about a topic or have special abilities when it comes to the exercise of ubiquitous expertise. It is, however, much more difficult to justify this stance when the expertise required is genuinely esoteric and the policy failure is more accurately described as technological populism (Collins & Evans, 2002). Two examples illustrate this, newer, phenomena. In the UK in the late 1990s, a scare over the MMR (measles, mumps, and rubella) vaccine led to a fall in vaccination rates as parents' local experiences were misleadingly labelled as *lay expertise* and treated as equivalent to the consensus medical researchers had reached through the epidemiological analysis of population data. More recently, and more widely, the corrosive consequences of widespread scepticism about experts for the idea that science—and expertise in general—should have some special status when technological decisions are being made are being seen in public and political debate across the Western world. SEE responds to this challenge by proposing a realist model of expertise that is consistent with wave two's description of science but which also argues that scientific values—more so than scientific facts—are an important part of any modern society (Collins & Evans, 2017). As explained in more detail in the section 'Making Boundaries: Expertise as Real', this enables SEE to argue for more participation in some cases and less in others.

Philosophical Roots: Expertise as a Form of Life

SEE starts with early work in the sociology of scientific knowledge (e.g., Bloor, 1973, 1991; Collins, 1974, 1975, 1985/1992) that draws on Ludwig Wittgenstein's concept of a *form of life* (Winch, 1958; Wittgenstein, 1953) as the fundamental unit of analysis. The link with social constructivism comes from the way a form of life, which captures the set of ideas, actions, traditions, and values that define a social group, gives its members the tools to make sense of the world and the standards by which to hold each other accountable. The concept applies to all scales and topics, so a form of life can be very large and include many people spread over wide geographical areas (e.g., English speakers or soccer players) or it can be very local and/or highly specialized (e.g., Basque language speakers in the Pyrenees or Maori kapa haka performers).

The notion of form of life can be used to understand the continuity of knowledge and what happens when new ideas emerge. Both aspects are illustrated in the following passage from Peter Winch's (1958) analysis of the relationship between philosophy and sociology. Note that, although the example given relates to changes in scientific knowledge, the insight is intended to generalize to all forms of social and cultural innovation:

> Imagine a biochemist making certain observations and experiments as a result of which he discovers a new germ which is responsible for a certain disease. In one sense we might say that the name he gives this new germ expresses a new idea, but I prefer to say in this context that he has made a discovery within the existing framework of ideas. I am assuming that the germ theory of disease is already well established in the scientific language he speaks. Now compare with this discovery the impact made by the first formulation of that theory, the first introduction of the concept of germ into the language of medicine. This was a much more radically new departure, involving not merely a new factual discovery within an existing way of looking at things, but a completely new way of looking at the whole problem of the causation of diseases, the adoption of new diagnostic techniques, the asking of new kinds of questions about illnesses, and so on. In short it involved the adoption of new ways of doing things by people involved, in one way or another, in medical practice. (Winch, 1958, pp. 121–122)

As explained in the first part of the passage, the *germ theory of disease* is the basis of a form of life within which medical scientists and practitioners carry out their day-to-day work. This is very similar to Kuhn's concept of a paradigm (Kuhn, 1996, originally published in 1962) as well as many other sociological concepts—collectivity, culture, subculture, microculture, etc.—that also describe the shared values and practices that define a social group. The idea expressed in the second part of the passage is close to the idea of a Kuhnian revolution and gives STS its more radical, constructivist, edge. Rather than describing the world, here we see the germ theory of disease creating and constituting that world, giving meaning to entities, institutions, and practices that did not—and could not—exist without the idea of germs.

The second distinctive feature of STS, and of SEE in particular, is the importance attached to tacit knowledge. Although Wittgenstein does not use the term, which is usually attributed to Michael Polanyi (e.g. Polanyi, 1966), the importance of knowledge that is not explicitly articulated is central to Wittgenstein's writings about the nature of rules and the role of a form of life in determining how rules are to be interpreted and applied. Using the analogy of a signpost, Wittgenstein writes:

> A rule stands there like a sign-post.—Does the sign-post leave no doubt open about the way I have to go? Does it shew [*sic*] which direction I am to take when I have passed it; whether along the road or the footpath or cross-country? But where is it said which way I am to follow it; whether in the direction of its finger or (e.g.) in the opposite one?—And if there were, not a single sign-post, but a chain of adjacent ones or of chalk marks on the ground—is there only *one* way of interpreting them? (Wittgenstein, 1953, para. 85)

It is possible to distinguish between several different types of tacit knowledge (Collins, 2001, 2010), with two—collective tacit knowledge and somatic-limit tacit knowledge—being especially important for understanding expertise. Collective tacit knowledge is the sort of knowledge implied in Wittgenstein's example of the sign-post: how do we know that the *pointy end* of the sign indicates which way to go or that we should follow the path even as it bends around a corner and we are no longer travelling in the *same* direction? The answer is not that these things are written down somewhere as explicit knowledge. The answer is that these, and a myriad other social practices, are given to us by our form of life and we learn how to follow signs, and other social rules, by being socialized into the collectivity that shares that particular form of life.

The other kind of tacit knowledge that matters in debates about expertise is somatic-limit tacit knowledge (Collins, 2010). Here the focus is on the individual expert and the bodily experience of performing physical tasks such as riding a bike or catching a ball. Again, the knowledge needed to perform the task is unexplicated and, even if rules could be discovered that would enable machines or other humans to perform the task, they would not necessarily describe the experience of being an expert practitioner (Collins, 2007). Indeed, in many such practical tasks a lack of self-consciousness is said to be the mark of the expert as in the Dreyfus five-stage model of expertise (e.g., Dreyfus, 2004).

Making Boundaries: Expertise as Real

In contrast to tacit knowledge and form of life, where there is much in common between the second and third waves of STS, the idea of boundaries reveals—perhaps ironically—the differences between the two approaches. In the case of wave two, expert status, and the demarcation this creates between experts and non-experts, is typically treated as the outcome of *boundary work* carried out by competing groups of social actors (see, e.g., Gieryn, 1999). As a result, the research questions addressed are largely descriptive: how were claims to expertise negotiated and expert status recognized in this or that particular case?

In practice, however, this descriptive approach contains an implicit normative agenda. Just as accepting the new germ theory of disease requires that new ways of doing medicine are developed, so the new description of science as a social practice, and not a source of authoritative and unbiased knowledge, suggests that new ways of relating to science and expertise are needed. For wave two, recognizing that social judgements are essential parts of science has led to the idea that science's institutions and practices should become more representative of the wider society (e.g. Irwin, 1995; Ottinger, 2010; Shapin, 2007; Wynne, 1992). The idea is that a more inclusive approach will allow a wider range of social interests and positions to be represented in the making of scientific and technical knowledge and a wider range of questions to be asked about its applications (for examples of these kinds of arguments see Callon, Lascoumes, & Barthe, 2011; Douglas, 2009; Fisher et al., 2015; Harding, 2006; Longino, 1990).

The third wave recognizes these problems but distinguishes between questions of framing and those of knowledge production. Whilst there can be nothing wrong with calls for more democracy in the setting of public policy, to solve problems of knowledge-production we must look to the inclusion of a wider range of *expertises*. The third wave endorses a widening of expert debate but this must mean including more of the appropriate expert communities, not diluting the notion of technical expertise to include the general public. The solution is better understanding of expertise in order to recognize a wider range of esoteric experts that includes the previously excluded experience-based experts found outside the elite institutions of science (Collins & Evans, 2002, 2017; Collins, Weinel, & Evans, 2010). Under this approach expertise is an *analysts' category* as well as an actors' category and this requires a theory that can explain what expertise consists of, the kinds of decisions for which it is relevant, and a way of telling who is and who is not an expert.

In many of the case studies used to develop wave two the danger of too much participation does not arise. Instead the pressing problem is invariably an overly technocratic form of decision-making in which elite experts frame the question and determine what will be allowed to count as a relevant policy option. In this scenario, sociological analysis exposes the hidden assumptions and values of experts and justifies a more inclusive approach that recognizes the value of local, indigenous, or other experiential knowledge. But what happens when it is not the assumptions and values of the mainstream elite that are suspect but those of the groups that want to reject the consensus view? We have already mentioned the case of MMR vaccinations, where limited and partial data were overinterpreted by parents, whose fears were encouraged and exacerbated by press reporting, with the result that vaccination rates fell, herd immunity was lost, and the health of vulnerable people was put at unnecessary risk. A similar diagnosis can be made of Thabo Mbeki's decision, when he was President of South Africa, not to permit the use of the anti-retroviral drug AZT (azidothymidine) to reduce the risk of mother-to-child transmission of HIV in South Africa. The decision, which led to more children being infected than would otherwise have been the case, was based, at least in part, on Mbeki's belief that the mainstream medical theories about HIV and AIDS were wrong, though in this case the misinformation came not from the media but from Internet pages maintained by scientists whose work had long been dismissed within the scientific community (Weinel, 2010). In a similar way, Oreskes and Conway (2010) show that the political controversy about anthropogenic climate warming is being deliberately kept alive by trained and qualified scientists who are supported by the oil industry to use their knowledge to create the appearance of uncertainty even though mainstream scientific opinion is clear. Wave two of science studies has no conceptual apparatus to explain why these practices are wrong since everything is a matter of social construction. SEE's approach, with its emphasis on the norms and values of a scientific form of life, does distinguish this kind of activity from science proper and is, therefore, able to say who is and is not a legitimate contributor to an expert debate turning on technical expertise.

Not all technical debates turn on technical expertise, however, and the policy consequences of any technical consensus are a matter of politics not science. Thus, Mbeki, even if he had accepted the safety and efficacy of anti-retroviral drugs, could still have taken the political decision that South Africa was not going to allow itself to be presented as a disease-ridden society and robbed by the money-gouging pharmaceutical companies. Even the acceptance of anthropogenic climate warming still leaves open the question of what to do about it. This distinction between the *technical* and *political* is a central element of SEE that is radically different to the fact–value distinction that was refuted by wave two. For wave three, the *technical* and *political* are two distinct forms of life turning on different expertises and different sets of values—in other words, the untenable fact–value distinction is replaced with sociologically informed value–value distinction. For political decision-making, democratic norms set the standard and the rights of citizens to participate in such decisions are guaranteed by law; for technical decision-making, scientific values provide the moral framework, with the participants selected using the typology of expertises set out in the next section.

Theoretical and Methodological Innovations

The Periodic Table of Expertises

The classification of expertises developed by SEE is summarized in the Periodic Table of Expertises (see Table 4.1) first published in Collins and Evans (2007).

Table 4.1 The Periodic Table of Expertises

Ubiquitous Expertises					
Dispositions				Interactive ability / Reflective ability	
Specialist Expertises	Ubiquitous			Specialist	
	Beer-mat knowledge	Popular understanding	Primary source knowledge	Interactional expertise	Contributory expertise
				Polimorphic / Mimeomorphic	
Meta-expertises	External		Internal		
	Ubiquitous discrimination	Local discrimination	Technical connoisseurship	Downward discrimination	Referred expertise
Meta-criteria	Credentials	Experience		Track-record	

Reproduced from Collins, Harry and Evans, Robert, *Rethinking Expertise*, p. 14, Table 1, © 2007, University of Chicago Press.

The principle behind the table is that different kinds of experiences give rise to different kinds of expertise. Working from the top, the first row acknowledges the society-wide *ubiquitous expertises* (e.g., cultural norms, speaking a natural language) derived from general socialization that are needed for a person to function within a given society. The second row identifies two individual *dispositions* that can facilitate the socialization needed to acquire the other kinds of expertise listed in the table but which are not, strictly speaking, necessary for socialization to take place.

The next row—*specialist expertises*—are the expertises associated with specific social groups such as scientists, car drivers, and plumbers. It is important to note that the difference between a specialist expertise and a ubiquitous expertise is sociological—access to the former is restricted in some way within a society, whereas access to the latter is not. Otherwise, they are both the product of successful socialization into a form of life.

Moving across the row of specialist expertises, there is a progression from very basic knowledge to complete mastery of the domain. The first three categories—beer-mat knowledge, popular understanding, and primary source knowledge—denote the kinds of understanding that can be achieved by using pre-packaged resources such as webpages, online videos, magazines, books, and academic journals. Acquisition of these expertises depends only on the deployment of ubiquitous expertises and, as they involve no direct interaction with specialist communities of practitioners, no socialization into the tacit knowledge of esoteric communities is possible.

In contrast, the last two categories of specialist expertise—interactional expertise and contributory expertise—can be attained only through sustained immersion in the relevant community. If this is successful, then expert performance can be attained as the learner will have acquired the specialist tacit knowledge needed to act independently but in ways that other members of the group would endorse as correct (for a recent assessment of contributory and interactional expertises see Collins & Evans, 2015; Collins, Evans, & Weinel, 2016).

Because high-level expertise is hard to acquire—for example, it is suggested that it takes thousands of hours of deliberate practice in the case of esoteric expertises (Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson & Pool, 2016)—modern societies require citizens to make judgements about experts without having the relevant specialist expertise themselves. The meta-expertises row lists some of the ways in which this can be done. These include *external* social judgements based on discrimination and which make little or no reference to the content of what is being judged and *internal* judgements that require some familiarity with the specialist expertise in question. The final row of the table identifies criteria that might be used for identifying experts, with relevant experience being the best of the three options as it encompasses a wider range of possibilities and skills.

Using the Periodic Table of Expertises, it is possible to talk about experts and expertise in a more nuanced way. For example, the traditional idea of an expert is that of a skilled practitioner who has mastered the practice of a domain. Using Table 4.1, this usage corresponds to a contributory expert and the importance attached to acquiring tacit knowledge explains why achieving this type of expertise requires extensive immersion in the domain of practice. This is most obviously true for skills like speaking a language or driving a car, but the same is also true for other, more cognitive, domains of expertise including scientific specialties. Significantly, and despite the popularity of online learning, not even the most sophisticated descriptions or multimedia demonstrations will be enough to produce contributory expertise (or interactional expertise) in an esoteric domain, as participation in the community of practitioners is essential.

Unlike contributory expertise, which adds clarity to an old idea, the category of interactional expertise is novel and poses new questions about the relationship between language and practice. Interactional expertise is defined as fluency in the language that a community uses to describe its practices; hereafter its *practice language* (Collins, 2011). All contributory experts are also interactional experts but the novel claim is that it is possible to acquire interactional expertise to the level of that possessed by a contributory expert without mastering or even experiencing the physical practices that define the domain of expertise (Collins, 2004b, 2016b; Collins, Evans, & Weinel, 2017a).

The two new insights that interactional expertise provides are (1) that language itself is a social practice that cannot be reduced to written or recorded text and (2) that it is not necessary to perform a physical practice in order to acquire the practice language that describes it and which is used to make expert judgments about it. For these reasons, sustained immersion is always needed to achieve the necessary linguistic socialization but actually performing the practice, despite being an efficient way of entering a community and learning its practice language, is not necessary. As a result, although interactional expertise was initially seen as a relatively rare commodity—the property of the social scientist, specialist journalist, technically skilled manager, or peer reviewer, for example—we now see that it is far more widespread (Collins & Evans, 2015; Collins et al., 2017). Not only are contributory experts also interactional experts, interactional expertise may be shared wherever members of two different social groups interact on a regular basis and for a prolonged period of time.

The Imitation Game

The Imitation Game is a new research method that has been developed to explore the content and distribution of interactional expertise. It is a more rigorous version of the parlour game that inspired Alan Turing's famous *Turing Test* for the intelligence of computers (Turing, 1950). A basic Imitation Game consists of three players. One player, drawn from the *target group*, acts as the *Judge/Interrogator* and creates questions that are sent to the other two players. One of these, the *Non-Pretender*, is always drawn from the target group and answers naturally. The other, the *Pretender*, is recruited from a different group and asked to answer as if they were a member of the target group. The Judge/Interrogator then compares the answers and tries to work out which comes from the Pretender and which from the Non-Pretender. The hypothesis is that, where the Pretender has interactional expertise, the Judge/Interrogator will be unable to distinguish between the two sets of answers, no matter how demanding the questions they set. In many Imitation Games the players, who must be hidden from each other and may be in remote locations, interact via computers using specialist software but the game can also be played much more simply, over email via a *postman* who conceals the identities, or even with paper and pencil and some screens. (For more details of Imitation Game research see Collins et al., 2015, 2006; Collins & Evans, 2014; Evans & Collins, 2010.)

There are now several different variations of the Imitation Game method which have been used over a wide range of topics. The *classic*, three-player version is well suited to in-depth studies that probe the interactional expertise of individuals or small groups. This approach informed much of the early work using the method (Collins et al., 2006; Giles, 2006), where Collins demonstrated that his extended sociological fieldwork had enabled him to develop interactional expertise in gravitational wave physics. Other, more recent, work has used variants of this small-scale approach. Wehrens has used small-scale Imitation Games to prompt group discussions as part of a larger study investigating the extent to which therapists in an eating disorder clinic can take the perspective of their patients (Wehrens, 2014), whilst Evans and Crocker (2013) used the method to demonstrate that dieticians are able to take the perspective of patients with celiac disease and articulate the lived experience of the condition in ways that go beyond everyday understandings. Kubiak and Weinel (2016) have used Imitation Games to explore the change in East German identities following the reunification of Germany, showing that younger East Germans tend to see East Germany in geographical rather than cultural or institutional terms, and Collins (2016a) has used a variation on the method to explore how successful different kinds of Judges are at distinguishing between answers provided by experts and non-experts. This latter experiment used gravitational wave physics as the case study and showed that the expertise needed to make good judgements was restricted to contributory and interactional experts, with lay participants completely unable to distinguish between answers provided by experts and non-experts.

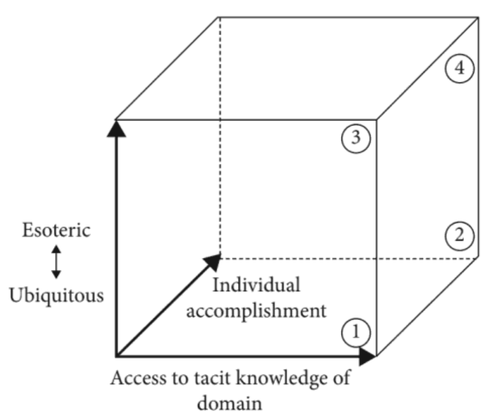
Larger-scale research has focused on more general topics such as religion, gender, and sexuality, with a significant amount of methodological work done to refine and develop protocols. Early reports of this research are published in Collins and Evans (2014), whilst Collins et al. (2015) include examples of content analysis and an analysis of the effect of running Imitation Games with the three roles played by small groups, rather than individuals. These early results show that Pretenders find it harder to succeed when questions are set by groups of Judges rather than individuals and that, even where groups are very diverse—as would be the case for gender—it is still possible for Judges to correctly identify the Pretender and Non-Pretender.

Three-Dimensional Model of Expertise

The sociological model of expertise put forward in this chapter can be summed up by saying that it operationalizes expertise along three dimensions (Collins, 2013b):

1. *Individual accomplishment*: this is similar to the model proposed by the philosopher Dreyfus (e.g., Dreyfus, 2004), as well as many analysts drawn from the psychological tradition, and captures the proficiency of the individual with respect to the domain of expertise in question.
2. *Esotericity*: this refers to ease with which the social collectivity that holds the expertise can be accessed, with ubiquitous expertises being at the most open end of this scale.
3. *Exposure to tacit knowledge*: this describes the way in which the learner has accessed the domain, ranging from published sources that add no specialist tacit knowledge through linguistic socialization in the practice language to full participation in the form of life.

Combining these gives the three-dimensional expertise space shown in Figure 4.1. As with the Periodic Table of Expertises, disaggregating expertise into different elements allows a much richer description of expertise and the ways in which it develops and spreads over time. For example, it is possible to see how an expertise such as natural language speaking can be ubiquitous, whilst other domains of expertise, such as the sciences, are more esoteric and, in addition, to see how this might change over time (e.g., car driving might once have been esoteric but it is now nearly ubiquitous). It also enables individual accomplishment and practice to be set within a broader scheme that suggests what kind of practice is needed—thus, if full mastery means reaching the right-hand edge of the back wall of the *expertise space diagram*, then practice must involve social interaction with the relevant community as, without this, the tacit knowledge of the domain can never be attained.



1: 5-year old language speaker	**2:** Accomplished adult language speaker
3: Novice sociologist	**4:** Poet (virtuoso)

FIGURE 4.1 Three dimensions of expertise.

Applications and Future Challenges

The SEE programme has an ambitious research agenda that includes a number of elements. The Periodic Table of Expertises has provided a useful starting point for research, with the category of interactional expertise being particularly important in terms of generating new insights. There is, however, still much work to do in understanding the nature of meta-expertises. For example, just in terms of the Periodic Table of Expertises, there is more work to be done on the potential overlap between the internal meta-expertises, which require some knowledge of the domain, and the higher level specialist expertises in which this knowledge is complete. Looking more broadly, we can also ask how meta-expertise categories like discrimination relate to psychological work on heuristics or decision-making under uncertainty. There is also much more that can be done with the Imitation Game, in terms of developing the protocols needed for comparative and longitudinal research and exploiting its potential for the in-depth analysis of particular domains of expertise. Finally, and perhaps most importantly given the concerns about policy-making that led to the development of SEE, there is a clear need for a more thorough engagement with political theory in order to better understand the relationship between expertise and democracy. As we cannot cover all these issues here, we use the remainder of this section to examine what SEE can contribute to the understanding of two contemporary issues in which expertise is central.

Interdisciplinary Working

Interdisciplinary working requires collaboration and co-operation between different groups of experts. A number of different models exist for understanding how this can happen, with trading zones (Galison, 1997) and boundary objects (Star & Griesemer, 1989) being two of the most common in the STS literature. The ideas associated with SEE can contribute to this literature in two different ways, with the idea of interactional expertise being particularly important. First, the typology of expertises provided by the Periodic Table of Expertises can be used to identify common themes and differences between different approaches and so provides an overarching framework within which different forms and modes of collaboration can be discussed and compared (Collins, Evans, & Gorman, 2007; Evans & Marvin, 2006; Fisher et al., 2015).

Second, the idea of interactional expertise can be used to design and facilitate interdisciplinary collaborations. Emphasizing socialization and the importance of tacit knowledge suggests that successful collaborations will require at least one group to learn new ways of seeing and doing in order to understand what the other is seeking to achieve and why (Galison, 1997; Gorman, 2002). As acquiring interactional expertise is not easy, time and space for this to occur will need to be created (Fisher et al., 2015).

Quite what form of interaction is most suitable for promoting and supporting collaboration needs much further research. Working within the SEE framework, methods for promoting interactional expertise are of most concern and build on the recognition that interactional expertise is more widespread than first thought and might, therefore, play a more significant role in the division of labour than previously imagined. This is particularly clear in Collins' (2004a, 2013a, 2017) studies of gravitational wave physics, where a collaboration of specialists was able to work because each was able to understand what their colleagues in other subfields were doing. The novel insight that comes from SEE is that the scientists did not need to be practitioners (i.e., contributory experts) in those other subfields in order to work together; what they needed to be was interactional experts who could understand the discourse of those subfields.

This finding has implications for professional development and training in many areas. In medical contexts, for example, the Imitation Game might be used to examine how well different kinds of medical practitioners are able to transcend the medical model and develop the interactional expertise needed to bridge the gap between biomedicine and the lived experience of patients (Evans & Crocker, 2013; Wehrens, 2014). This approach also has some resonances with the idea of *T-shaped* expertise (Conley, Foley, Gorman, Denham, & Coleman, 2017; Glushko, 2008) in which the vertical bar is the person's contributory expertise and the horizontal bar represents their growing interactional expertise in the domains where this is to be applied. It also suggests novel forms of teaching, not least that of students on multidisciplinary courses learning subjects like thermodynamics as a *second language* rather than *mathematical practice*, that is, as an interactional expertise rather than a contributory expertise (Berardy, Seager, & Selinger, 2011).

Role of Experts in a *Post-truth* Society

When the SEE program was first proposed, the problems affecting technological decision-making in the public domain were mostly associated with technocracy. Although there were some instances of what could be called excessive populism, such as the MMR controversy in the UK (Boyce, 2007), Thabo Mkeki's refusal to approve the use of AZT in South Africa (Weinel, 2010), and the deliberate attempts to create the impression of a controversy within climate change research (Oreskes & Conway, 2010) mentioned earlier, these were relatively rare. In 2002, there was no sense that the anti-establishment populism that characterizes contemporary political discourse in the USA and, to a lesser extent, Europe was only a decade or so away. For a more wide-ranging analysis of this *war on expertise* see Chapter 49, "The 'War' on Expertise", by Klein et al. (this volume).

The contribution of STS to these events is difficult to pin down. Despite the claims made by so-called *science warriors* (Gross & Levitt, 1998; Koertge, 2000) there is little evidence of any causal connection between STS and the rise of anti-science sentiments, not least because the vast majority of STS is very *scientific* in its approach (Labinger & Collins, 2001). That said, it is also clear that wave two STS struggles to respond to these events because the relational model of expertise it adopts provides no means to criticize the attributions of expertise made by others (for an example of the difficulties this creates see Sismondo, 2017; for a reply see Collins, Evans, & Weinel, 2017b).

In contrast, SEE provides a clear way to articulate concerns about the miss-application of expertise. The key ideas are the distinction between the technical and political elements of a decision (Collins & Evans, 2002; Collins et al., 2010), which determines when esoteric and ubiquitous expertises respectively are needed, and the normative claim that a scientific approach provides a morally superior way of answering the propositional questions that characterize the technical phase. The argument is set out in detail in *Why Democracies Need Science* (Collins & Evans, 2017), but the basic principle is that people who know what they are talking about should be preferred to those who do not in technical matters. More specifically, preferring someone who knows what they are talking about means preferring someone with substantial relevant experience who adheres to the norms of science such as being willing to admit they are wrong, listening to counter-claims and criticism irrespective of their source, striving to corroborate or falsify results, and doing their best to put aside vested interests. Crucially, the model turns on moral arguments about the kind of society that we want to live in as opposed to instrumental or utilitarian arguments about the probability of experts being correct. This matters because it is only in this way that SEE can provide a justification for preferring science even when it is sometimes obvious that science will be unable to provide certainty in the short term.

In summary, the studies of expertise and experience approach provides a theory of expertise that includes both a technical understanding of the nature of knowledge and a normative agenda that aims to shape how expertise is used in democratic societies. The core of the technical element is the idea of a form of life, with expertise understood as the outcome of successful socialization that can be demonstrated by social fluency. The political element relates to the use of expertise in democratic societies, where the novel contribution of SEE is the distinction between the technical and political elements of a decision, treating them as different domains of expertise. By making this distinction, SEE avoids both the reification of science that characterizes wave one and the collapse of expertise into nothing but attributions of status that paralyses wave two. Putting both the technical and political dimensions of SEE together provides an account of science that is faithful to its practices and which sees it as complementing—not usurping—the overarching role of democratic institutions.

References

Barnes, B., Bloor, D., & Henry, J. (1996). *Scientific knowledge: a sociological analysis*. Chicago: University of Chicago Press.

Berardy, A., Seager, T. P., & Selinger, E. (2011). Developing a pedagogy of interactional expertise for sustainability education. In *IEEE International Symposium on Sustainable*

Systems and Technology (ISSST) (pp. 1–4). Chicago: IEEE. https://doi.org/10.1109/ISSST.2011.5936891

Bloor, D. (1973). Wittgenstein and Mannheim on the sociology of mathematics. *Studies in History and Philosophy of Science Part A 4*(2), 173–191. https://doi.org/10.1016/0039-3681(73)90003-4

Bloor, D. (1991). *Knowledge and social imagery*, 2nd edn. Chicago: University of Chicago Press.

Boyce, T. (2007). *Health, risk and news: The MMR vaccine and the media*. New York: Peter Lang.

Callon, M., Lascoumes, P., & Barthe, Y. (2011). *Acting in an uncertain world: an essay on technical democracy*, trans. G. Burchell. Cambridge, MA: MIT Press.

Carr, E. S. (2010). Enactments of Expertise. *Annual Review of Anthropology 39*(1), 17–32. https://doi.org/10.1146/annurev.anthro.012809.104948

Collins, H. M. (1974). The TEA set: Tacit knowledge and scientific networks. *Science Studies 4*(2), 165–185.

Collins, H. M. (1975). The seven sexes: A study in the sociology of a phenomenon, or the peplication of experiments in physics. *Sociology 9*(2), 205–224. https://doi.org/10.1177/003803857500900202

Collins, H. M. (1985/1992). *Changing order: Replication and induction in scientific practice*. Chicago: University of Chicago Press.

Collins, H. M. (1996). In praise of futile gestures: How scientific is the sociology of scientific knowledge? *Social Studies of Science 26*(2), 229–244. https://doi.org/10.1177/030631296026002002

Collins, H. M. (2001). Tacit knowledge, trust and the Q of sapphire. *Social Studies of Science 31*(1), 71–85. https://doi.org/10.1177/030631201031001004

Collins, H. M. (2004a). *Gravity's shadow: The search for gravitational waves*. Chicago: University of Chicago Press.

Collins, H. M. (2004b). Interactional expertise as a third kind of knowledge. *Phenomenology and the Cognitive Sciences 3*(2), 125–143. https://doi.org/10.1023/B:PHEN.0000040824.89221.1a

Collins, H. M. (2007). Bicycling on the moon: Collective tacit knowledge and somatic-limit tacit knowledge. *Organization Studies 28*(2), 257–262. https://doi.org/10.1177/0170840606073759

Collins, H. M. (2008). Actors' and analysts' categories in the social analysis of science. In P. Meusburger, M. Welker, & E. Wunder (Eds), *Clashes of knowledge* (Vol. 1, pp. 101–110). Netherlands: Springer. Retrieved from http://dx.doi.org/10.1007/978-1-4020-5555-3_4

Collins, H. M. (2010). *Tacit and explicit knowledge*. Chicago; London: University of Chicago Press.

Collins, H. M. (2011). Language and practice. *Social Studies of Science 41*(2), 271–300. https://doi.org/10.1177/0306312711399665

Collins, H. M. (2013a). *Gravity's ghost and big dog: Scientific discovery and social analysis in the twenty-first century* (enlarged edn). Chicago: University of Chicago Press.

Collins, H. M. (2013b). Three dimensions of expertise. *Phenomenology and the Cognitive Sciences 12*(2), 253–273. https://doi.org/10.1007/s11097-011-9203-5

Collins, H. M. (2016a). An Imitation Game concerning gravitational wave physics. *arXiv:1607.07373 [Physics]*. Retrieved from http://arxiv.org/abs/1607.07373

Collins, H. M. (2016b). Interactional expertise and embodiment. In J. Sandberg, L. Rouleau, A. Langley, & H. Tsoukas (Eds), *Skillful performance: Enacting expertise, competence, and capabilities in organisations* (Vol. 7). Oxford: Oxford University Press.

Collins, H. M. (2017). *Gravity's kiss: The detection of gravitational waves*. Cambridge, MA: MIT Press.

Collins, H. M., & Evans, R. (2002). The Third Wave of science studies: Studies of expertise and experience. *Social Studies of Science 32*(2), 235–296. https://doi.org/10.1177/0306312702032002003

Collins, H. M., & Evans, R. (2007). *Rethinking expertise*. Chicago: University of Chicago Press.

Collins, H. M., & Evans, R. (2014). Quantifying the tacit: The Imitation Game and social fluency. *Sociology 48*(1), 3–19. https://doi.org/10.1177/0038038512455735

Collins, H. M., & Evans, R. (2015). Expertise revisited, Part I—Interactional expertise. *Studies in History and Philosophy of Science Part A 54*, 113–123. https://doi.org/10.1016/j.shpsa.2015.07.004

Collins, H. M., & Evans, R. (2017). *Why democracies need science*. Cambridge, UK; Malden, MA: Polity Press.

Collins, H. M., Evans, R., & Gorman, M. (2007). Trading zones and interactional expertise. *Studies in History and Philosophy of Science Part A 38*(4), 657–666. https://doi.org/10.1016/j.shpsa.2007.09.003

Collins, H. M., Evans, R., Ribeiro, R., & Hall, M. (2006). Experiments with interactional expertise. *Studies in History and Philosophy of Science Part A 37*(4), 656–674. https://doi.org/10.1016/j.shpsa.2006.09.005

Collins, H. M., Evans, R., & Weinel, M. (2016). Expertise revisited, Part II: Contributory expertise. *Studies in History and Philosophy of Science Part A 56*, 103–110. https://doi.org/10.1016/j.shpsa.2015.07.003

Collins, H. M., Evans, R., & Weinel, M. (2017a). Interactional expertise. In C. Miller, L. Smith-Doerr, U. Felt, & R. Fouche (Eds), *Handbook of Science and Technology Studies*, 4th edn. Cambridge, MA: MIT Press.

Collins, H. M., Evans, R., & Weinel, M. (2017b). STS as science or politics? *Social Studies of Science 47*(4), 580–586. https://doi.org/10.1177/0306312717710131

Collins, H. M., Evans, R., Weinel, M., Lyttleton-Smith, J., Bartlett, A., & Hall, M. (2015). The Imitation Game and the nature of mixed methods. *Journal of Mixed Methods Research 11*(4), 510–527. https://doi.org/10.1177/1558689815619824

Collins, H. M., Weinel, M., & Evans, R. (2010). The politics and policy of the Third Wave: new technologies and society. *Critical Policy Studies 4*(2), 185–201. https://doi.org/10.1080/19460171.2010.490642

Conley, S. N., Foley, R. W., Gorman, M. E., Denham, J., & Coleman, K. (2017). Acquisition of T-shaped expertise: an exploratory study. *Social Epistemology 31*(2), 165–183. https://doi.org/10.1080/02691728.2016.1249435

Douglas, H. E. (2009). *Science, policy, and the value-free ideal*. Pittsburgh, PA: University of Pittsburgh Press.

Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science, Technology & Society 24*(3), 177–181. https://doi.org/10.1177/0270467604264992

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review 100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Ericsson, K. A., & Pool, R. (2016). *Peak: secrets from the new science of expertise*. Boston: Houghton Mifflin Harcourt.

Evans, R. (2008). The sociology of expertise: The distribution of social fluency. *Sociology Compass* 2(1), 281–298. https://doi.org/10.1111/j.1751-9020.2007.00062.x

Evans, R., & Collins, H. (2010). Interactional expertise and the Imitation Game. In M. E. Gorman (Ed.), *Trading zones and interactional expertise creating new kinds of collaboration* (pp. 53–70). Cambridge, MA: MIT Press.

Evans, R., & Crocker, H. (2013). The Imitation Game as a method for exploring knowledge(s) of chronic illness. *Methodological Innovations Online* 8(1), 34–52. https://doi.org/10.4256/mio.2013.003

Evans, R., & Marvin, S. (2006). Researching the sustainable city: Three modes of interdisciplinarity. *Environment and Planning A* 38(6), 1009–1028. https://doi.org/10.1068/a37317

Felt, U., Fouche, R., Miller, C., & Smith-Doerr, L. (Eds). (2017). *The handbook of science and technology studies*, 4th edn. Cambridge, MA: The MIT Press.

Fisher, E., O'Rourke, M., Evans, R., Kennedy, E. B., Gorman, M. E., & Seager, T. P. (2015). Mapping the integrative field: Taking stock of socio-technical collaborations. *Journal of Responsible Innovation* 2(1), 39–61. https://doi.org/10.1080/23299460.2014.1001671

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures* 25(7), 739–755. https://doi.org/10.1016/0016-3287(93)90022-L

Galison, P. L. (1997). *Image and logic a material culture of microphysics*. Chicago: University of Chicago Press.

Gieryn, T. F. (1999). *Cultural boundaries of science: Credibility on the line*. Chicago: University of Chicago Press.

Giles, J. (2006). Sociologist fools physics judges. *Nature* 442(7098), 8. https://doi.org/10.1038/442008a

Glushko, R. J. (2008). Designing a service science discipline with discipline. *IBM Systems Journal* 47(1), 15–27. https://doi.org/10.1147/sj.471.0015

Gorman, M. E. (2002). Levels of expertise and trading zones: A framework for multidisciplinary collaboration. *Social Studies of Science* 32(5–6), 933–938. https://doi.org/10.1177/030631270203200511

Gross, P. R., & Levitt, N. (1998). *Higher superstition: The academic left and its quarrels with science*. Baltimore: Johns Hopkins University Press.

Harding, S. G. (2006). *Science and social inequality: Feminist and postcolonial issues*. Urbana: University of Illinois Press.

Irwin, A. (1995). *Citizen science: a study of people, expertise, and sustainable development*. London; New York: Routledge.

Jasanoff, S. (2003). Breaking the waves in science studies: Comment on H.M. Collins and Robert Evans, "The Third Wave of science studies." *Social Studies of Science* 33(3), 389–400. https://doi.org/10.1177/03063127030333004

Jasanoff, S. (2013). Fields and fallows: A political history of STS. In A. Barry & G. Born (Eds), *Interdisciplinarity: Reconfigurations of the natural and social sciences* (pp. 99–118). London: New York: Routledge.

Koertge, N. (Ed.). (2000). *A house built on sand: exposing postmodernist myths about science*. New York: Oxford University Press.

Kubiak, D., & Weinel, M. (2016). DDR-Generationen revisited—Gibt es einen Generationszusammenhang der "Wendekinder"? In A. Lettrari, C. Nestler, & N. Troi-Boeck (Eds), *Die Generation der Wendekinder* (pp. 107–129). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-11480-0_8

Kuhn, T. S. (1996). *The structure of scientific revolutions*, 2nd edn, enlarged, Vol. 14. Chicago: University of Chicago Press.

Labinger, J. A., & Collins, H. M. (Eds). (2001). *The one culture?* Chicago: University of Chicago Press. Retrieved from http://www.press.uchicago.edu/ucp/books/book/chicago/O/bo3634845.html

Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Beverly Hills: Sage Publications.

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.

Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. New York: Bloomsbury Press.

Ottinger, G. (2010). Buckets of resistance: Standards and the effectiveness of citizen science. *Science, Technology & Human Values 35*(2), 244–270. https://doi.org/10.1177/0162243909337121

Polanyi, M. (1966). *The tacit dimension*. Chicago; London: University of Chicago Press.

Shapin, S. (2007). *A social history of truth: Civility and science in seventeenth-century England*. Chicago: University of Chicago Press.

Sismondo, S. (2010). *An introduction to science and technology studies*, 2nd edn. Chichester, UK; Malden, MA: Wiley-Blackwell.

Sismondo, S. (2017). Post-truth? *Social Studies of Science 47*(1), 3–6. https://doi.org/10.1177/0306312717692076

Star, S. L., and Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science 19*(3), 387–420.

Turing, A. (1950). Computing machinery and intelligence. *Mind 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Wehrens, R. (2014). The potential of the Imitation Game method in exploring healthcare professionals' understanding of the lived experiences and practical challenges of chronically ill patients. *Health Care Analysis 23*(3), 253–271. https://doi.org/10.1007/s10728-014-0273-8

Weinel, M. (2010). *Technological decision-making under scientific uncertainty: Preventing mother-to-child transmission of HIV in South Africa*, PhD dissertation. Cardiff University. Retrieved from http://orca.cf.ac.uk/55502/

Winch, P. (1958). *The idea of a social science and its relation to philosophy*. London; New York: Routledge & Kegan Paul.

Wittgenstein, L. (1953). *Philosophical investigations*, trans. G. E. M. Anscombe. Oxford: Blackwell.

Wynne, B. (1992). Misunderstood misunderstanding: Social identities and public uptake of science. *Public Understanding of Science 1*(3), 281–304. https://doi.org/10.1088/0963-6625/1/3/004

Wynne, B. (2003). Seasick on the Third Wave? Subverting the hegemony of propositionalism: Response to Collins & Evans (2002). *Social Studies of Science 33*(3), 401–417. https://doi.org/10.1177/03063127030333005

CHAPTER 5

GIFTEDNESS AND TALENT DEVELOPMENT IN CHILDREN AND YOUTH

STEVEN I. PFEIFFER

Introduction

This chapter provides an overview on recent thinking about gifted and talented children and youth. The chapter briefly discusses a history of gifted education and then tackles some big picture issues and future possibilities. The chapter addresses a number of questions, including the following: Who are the gifted? How are gifted identified? Is giftedness domain-specific or domain-general? How malleable is giftedness? Does giftedness represent a qualitative or quantitative difference? How does the concept of expertise fit into gifted education? The chapter concludes with future directions and challenges facing the gifted field. Because of space limitations, the chapter does *not* tackle the topic of the social and emotional development of the gifted; the interested reader is directed to resources by Neihart, Pfeiffer, & Cross (2016), and Pfeiffer, Shaunessy-Dedrick, & Foley-Nicpon (2017).

Who Are the Gifted and Talented?

To many, it may seem trivial to ask the question of exactly who are the gifted. If you work in schools, at least in the USA, then you know exactly who they are. The gifted are those students who meet the eligibility criteria that your school district or state stipulate to qualify as gifted students. And if you are a psychologist, then you also know who the gifted are. They are those youngsters who obtain an IQ test score that exceeds a certain threshold, according to what you learned demarcates giftedness—typically, two standard deviations above the mean in most places (Pfeiffer, 2015). These are the views of a great many practitioners, according to a recent national survey (McClain & Pfeiffer, 2012). This chapter departs from a traditional and, some might argue, antiquated view and provides a more recent and nuanced perspective on what we mean by the gifted and talented.[1] For example, recent conceptions of giftedness reflect an appreciation for domain-specific talent development, consideration of contextual and sociocultural factors, a stronger orientation toward processes that moderate and mediate the unfolding of gifts, and talent development over the lifespan. These are recent changes shared by leading researchers in the gifted and talent development field internationally (Stoeger, Balestrini, & Ziegler, 2017).

High IQ equals gifted dominated twentieth-century thinking. But we are now in a new millennium, and research in developmental psychology and the cognitive neurosciences has informed our thinking about high ability students—the gifted (Pfeiffer, 2002, 2003, 2013b, 2015; Ziegler, Stoeger, & Vialle, 2012). Shortly, the chapter will examine a few different models of giftedness that lead to different ways for conceptualizing and identifying giftedness. Most readers would agree that the young child who is reading at age three, excelling at competitive chess by age six, writing computer software programs at age eight, or performing piano in an orchestra by age ten is gifted. These four examples are indicative of children who are developmentally advanced, one hallmark of giftedness (Pfeiffer, 2001, 2002, 2009, 2012, 2013b).

Most gifted authorities, even today, agree that the gifted are those in the upper 3–10% compared to their same-age peers in general intellectual ability, distinguished performance in one or more academic domains, and/or evidence of creative work (Gagné, 2009; Pfeiffer, 2012, 2015; Wai, 2014). Not surprisingly, research confirms that there is a genetic influence in the expression of giftedness, at least at the high end of the IQ continuum (Plomin & Spinath, 2004). For example, the fields of music and mathematics are particularly rich with examples of child prodigies. Evidence also comes from the emergence of eminence among young children from impoverished environments (Nisbett, 2009). However, most authorities today also agree that the full unfolding of gifts and talents requires a nurturing and supportive environment, available resources, certain personality characteristics, and even good fortune (Foley-Nicpon & Pfeiffer, 2011; Pfeiffer, 2012, 2013b, 2015). This is a recent departure from traditional, twentieth-century views of giftedness seen as a fixed and immutable state of the young child.

The definition of gifted adopted for this chapter is based on the *tripartite model of giftedness*, which will be explained shortly. The definition of gifted is the following: "The gifted child demonstrates a greater likelihood, when compared to other students of the same age, experience and opportunity, to achieve extraordinary accomplishments in one or more culturally valued domains" (Pfeiffer, 2013b). Based on this definition, a child's gifts can be in any culturally valued domain, such as science,

[1] Whenever the chapter refers to "gifted," the meaning is "gifted and talented," terms often used interchangeably in the gifted literature.

mathematics, athletics, the performing arts, leadership, and even community volunteerism. The list of gifts is almost inexhaustible, limited only by what the culture values. As the child gets older, in most cultures and societies, there is increased opportunity for exposure to a growing number of different domains in which they can excel and gain expertise and eminence (Pfeiffer, 2015). For example, the young child who demonstrates precocious mathematical abilities at age seven has the potential to explore a wide variety of academic trajectories and become distinguished in a number of careers as an adult.

The gifted student's academic needs—in the USA and globally—are frequently *not* substantially met in the classroom or school, and quite often they require specialized programs or services not ordinarily provided in the regular classroom (Pfeiffer, 2015). This is the primary justification for gifted assessment in the schools—to determine whether a bright student has extraordinary intellectual abilities or evidence of potential for outstanding performance, frequently indicative of a need for special educational programs or resources not presently available in the regular classroom (Pfeiffer, 2018a).

Brief History of Gifted Education

There has been much written about the history of giftedness and gifted assessment (e.g., Mönks, Heller, & Passow, 2000; Pfeiffer, 2008, 2018a; Robinson & Clinkenbeard, 2008; Tannenbaum, 1983, 2000). As far back as Confucius in China and Plato in Greece, philosophers wrote about *heavenly* children. Their writings theorized on what high ability constituted, and also provided suggestions for how society should find and nurture these special young citizens (Mönks, Heller, & Passow, 2000).

In US and Western culture, we trace the early roots of gifted education to Lewis Terman's research at Stanford University. Terman conducted a large longitudinal study which followed a cohort of students tested with IQ scores at or above 140. Terman collected considerable data on these students over the course of fifty years. He stated that the "twofold purpose of the project was . . . to find what traits characterize children of high IQ, and . . . to follow them for as many years as possible to see what kind of adults they might become" (Terman, 1925, 223; Terman & Oden, 1951, 21). Terman concluded that children of high IQ (140 or higher) are healthier, better-adjusted, and higher achievers than unselected children (Robinson & Clinkenbeard, 2008).

There are other early Western studies and writings on the gifted, such as Galton's *Hereditary Genius* (1869) and Cattell's *A Statistical Study of American Men of Science* (1906–1910) (Whipple, 1924). However, nothing quite captured the imagination of the public as did Terman's *Genetic Studies of Genius* (Mönks, Heller, & Passow, 2000). Terman's work played a pivotal role in equating giftedness and high IQ in Western thinking. Almost one hundred years later, Terman's influence in Western culture remains pronounced. This traditional *gifted child focus* emphasizes general intelligence and assumes that the gifted reflect a clearly demarcated and fixed category of exceptional individuals who differ in quantitative *and* qualitative ways from their non-gifted peers. The *gifted child focus* dominated twentieth-century Western cultural thinking. It has been the major zeitgeist in gifted education through the twentieth century in Western societies.

Numerous publications offer insights into non-Western sociocultural conceptions of giftedness (cf., for an overview, Phillipson & McCann, 2007) and intelligence (cf., for an overview, Niu & Brass, 2011). Clearly, there have been, and still remain, large-scale East–West differences in giftedness conceptions. Perhaps the most extensive research on non-Western conceptions of giftedness has focused on East Asia (Stoeger, Balestrini, & Ziegler, 2017; Van Tassel-Baska, 2013). Typical of East-Asian conceptions of giftedness is less emphasis on entity theories of giftedness and greater focus on educability and Confucian holistic outlooks (Phillipson, 2013), as opposed to Western society, mind–body dualism. This translates in the East into more malleable views of learning and intelligence (and giftedness), and greater emphasis on effort and motivation (Li & Fischer, 2004; Pfeiffer, 2013a; Yeo & Pfeiffer, 2018). These Eastern views are consistent with the newly emerging talent development approach to the education of gifted students taking root in the West.

This new focus that is emerging and challenging the dominant *gifted child focus* in Western societies is a shift aligned more with Eastern philosophy and thinking. It is labeled a *talent development perspective* (Dai, 2010; Pfeiffer, 2013b, 2015; Worrell, Subotnik, & Olszewski-Kubilius, 2017). Major proponents of the talent development perspective advocate that this new zeitgeist in gifted education lends itself logically to *a path toward eminence* (Worrell, Subotnik, & Olszewski-Kubilius, 2017). The talent development approach doesn't necessarily define giftedness differently, but it clearly embraces a non-static, ecological, transactional, and developmental viewpoint in explaining the unfolding of gifts and talents over the individual's life. It elegantly links the education of the gifted, at an early age, with career planning and ultimate career eminence, at later ages. But before we get too far ahead of ourselves, let's return to some of the early figures in the history of gifted education.

Another early, influential figure was James Gallagher.[2] In 1960 Gallagher submitted to the Illinois legislature a report whose purpose was "to review and summarize all of the information now available relating to the education of gifted children" (Gallagher, 1960, 3). Gallagher's report, *Analysis of Research on the Education of Gifted Children*, concluded that "special programming for gifted children requires additional personnel and services" (131). Gallagher pointed out that only two cents out of every $100 spent on K–12 education in the USA supports the gifted, and that existing programs for the gifted do not reach nearly enough gifted students in America's schools (Pfeiffer, 2013b). He added that special programs for the gifted are a low priority, the federal role in services to the gifted is all but nonexistent, and that "gifted students have been relatively ignored in educational programs such as No Child Left Behind (and the more recent federal legislation, Reach for the Top)" (Gallagher, 2008, 7). In 2006, for

[2] One of my mentors in graduate school at the University of North Carolina—Chapel Hill.

example, the US Department of Education spent nearly $84 billion. The only program specifically funded to address the education of the gifted got $9.6 million, one-hundredth of 1% of federal education expenditures. Few countries, in fact, allocate anything but very nominal funds earmarked for the gifted (VanTassel-Baska, 2013).

A number of other individuals have influenced the history of gifted education. Leta Hollingsworth (1886–1939), for example, played an important early role with her case studies into high IQ students in the New York City schools. Hollingsworth was a psychologist who practiced in New York City about the same time as Terman was a professor at Stanford. Hollingsworth is author of the first textbook on gifted education, *Gifted Children: Their Nature and Nurture* (1926).

Most countries have been slow to respond to the educational needs of students of high ability (Pfeiffer, 2015). Many authorities believe that the ambivalence and disinclination of governments to address the unique needs of high ability students is the result of society's perception that they are already a privileged group, and will succeed without special funding or services (Stephens, 2008, 2011, 2018). There is also a sense that *equity* trumps *excellence* in driving educational policy. And yet the US National Science Board (2010) recognizes this dilemma. It stated in a report, "the opportunity for excellence is a fundamental American value and should be afforded to all" (5).

One final point bears mentioning. A recent survey indicates that changes in definitions and categories of giftedness are occurring, both in the USA and globally. For example, across the USA, states vary considerably in how they identify gifted students. A majority of states still adhere to Terman's view that giftedness equates to high IQ, although they don't use quite as high a threshold or cut-score for demarcating giftedness. States frequently endorse a multiple cutoff or averaging approach to gifted decision-making (McClain & Pfeiffer, 2012). Most states continue to embrace a *gifted child focus*, and have not considered an alternative *talent development perspective* that de-emphasizes measured IQ and emphasizes domain-specific definitions of gifted (Dai, 2010; Pfeiffer, 2015). A recent survey in Europe reported similar findings (Tourón & Freeman, 2017).

Alternative Ways of Conceptualizing Giftedness

There are many different ways to conceptualize giftedness. Sternberg & Davidson (2005) suggest at least twenty different ways to view giftedness. Many different authors have proposed models of giftedness (Pfeiffer, 2008, 2018a). Most of the widely cited models fit into one of four popular models that are described next. These different models imply different ways to define, identify, and nurture gifts.

The different models vary in their level of detail and in how easily they can be translated into assessment protocols and intervention programs. They also vary in terms of their relative emphasis on the role of individual differences, developmental antecedents, genetics, family, and the environment (Ackerman, 2013; Pfeiffer, 2015; Simonton, 2014; Wai, 2014). The four models discussed next are *traditional psychometric views, talent development models, expert performance perspectives*, and *multiple intelligences*. These models are not contradictory to the fifth model, the *tripartite model* (Pfeiffer, 2013b, 2015), which will also be introduced.

Julian Stanley's mathematically and verbally precocious talent search model, for example, reflects thinking that cuts across two models: traditional psychometric views and talent development models (Stanley, 1976, 1990, 2000). Francoys Gagné's developmental, differentiated model of giftedness and talent (Gagné, 2005), Joseph Renzulli's three-ring conception of giftedness (Renzulli, 1978, 2005, 2011), and Rena Subotnik's developmental model (Subotnik, 2003; Subotnik, Olszewski-Kubilius, & Worrell, 2011) all are *talent development models*. K. Anders Ericsson's work epitomizes a somewhat unique way of thinking about giftedness with a strong emphasis on the environment, from an *expert performance perspective* (Ericsson, 1996, 2014; Ericsson & Charness, 1995).

There is growing consensus among most gifted authorities that giftedness is best viewed as *specific*, that the expression of giftedness occurs within a particular domain (Mayer, 2005). I agree with this viewpoint, at least when we consider students of high ability beginning around the third or fourth grade (Pfeiffer, 2013b, 2015). In the earlier grades, one could make a compelling argument that giftedness—or rather the prediction of giftedness—is not necessarily yet specific to one particular domain but rather a reflection of general intellectual ability and potential to excel (Pfeiffer, 2015). For example, most would agree that a three year old who is reading at a second-grade level is gifted.

Traditional Psychometric Views

Some readers are probably familiar with the psychometric view of giftedness, which conceptualizes high IQ as the defining feature of the construct giftedness. The traditional psychometric model views high IQ and giftedness as synonymous (Pfeiffer, 2002, 2012, 2013b, 2015), although it is unclear just how high constitutes giftedness. Many of the earliest researchers investigated the scientific basis of giftedness from a domain-general perspective, using the terms gifted, genius, and talented interchangeably. Francis Galton's book *Hereditary Genius* (1869) introduced the notion of intellectual genius to the public. Galton carefully and scientifically analyzed the family lineage of distinguished men and concluded that genius is genetically inherited. His estimations of genius were subjective, not based on psychometric measures, but nonetheless set the stage for the scientific study of giftedness (Ackerman, 2013; Kaufman & Sternberg, 2008; Pfeiffer, 2015).

Galton's work was followed by Charles Spearman (1904), who used the newly developed statistical tool of factor analysis to demonstrate that there was a significant amount of shared variance across most cognitive tests. He called this ubiquitous shared ability *g*, or general intelligence (now called *psychometric g*). The analyses that he ran also uncovered specific abilities unique to one or two of the tests, labeled specific abilities *s*. At around the same time, Alfred Binet and Theodore Simon (1916) developed a mental scale to identify students struggling in the Paris schools who might need alternative education. Binet and Simon's scale was the first test to include assessment of higher-level cognitive skills.

Lewis Terman adapted Binet and Simon's scale and created the Stanford–Binet Intelligence Scale, one of the first tests to actually identify gifted students (Terman, 1916). Terman's scale yielded a global score that viewed giftedness from a domain-general perspective and intelligence as a single entity. He proposed a classification system in which a youngster who obtained an IQ score of 135 or above was labeled *moderately gifted* (Terman, 1925), above 150 as *exceptionally gifted*, and above 180 as *severely and/or profoundly gifted* (Kaufman & Sternberg, 2008; Webb, Meckstroth, & Tolan, 1982). This classification system is still popular today.

Psychometric *g* or IQ continues to be *the* leading index for identifying the gifted both in the USA and internationally (McClain & Pfeiffer, 2012; Pfeiffer, 2013a; Sternberg, Jarvin, & Grigorenko, 2011). The IQ enjoys wide popularity because it provides a seemingly precise, impartial, objective, and quantifiable integer representing human intelligence. There also is a large research literature supporting the validity of the IQ score. There is arguably more published research testing the validity of the IQ construct than any other psychological construct (Neisser et al., 1996; Nisbett, 2009). IQ does predict school performance moderately well, and also predicts many other important life outcomes, including attainment of expertise (Simonton, 2017). However, there are certainly other psychological constructs that have also shown great promise in predicting school performance and life success (Simonton, 2014; Sternberg, Jarvin, & Grigorenko, 2011).

There are many within the psychometric camp who don't endorse a domain-specific model of intelligence. Louis Thurstone (1938) was the first researcher who challenged the prevailing domain-general model and proposed the notion of specific abilities as an alternative way of conceptualizing intelligence. Thurstone used a different method of factor analysis and identified seven primary and independent mental abilities. A growing body of studies supports hierarchical factor models of intelligence. The hierarchical models have general ability at the very top of the apex, more general higher cognitive or intellectual abilities at the next level, and various more specific cognitive or intellectual skills lower in the hierarchy. The hierarchical model that has gained greatest acceptance in the psychometric community is Carroll's (1993) three-stratum theory. In Carroll's model, Stratum I consists of highly specialized cognitive skills, Stratum II somewhat less specific and more broad domains of intellectual abilities, and Stratum III, at the apex, only one ability, the *g* factor.

Recently, Carroll's model and a second hierarchical model, the Horn and Cattell (1966) model of fluid and crystallized intelligence, were synthesized into the Cattell–Horn–Carroll (CHC) theory (Flanagan & Harrison, 2012). Although the CHC model includes *g* at the apex, its main emphasis is on the measurement of those factors and cognitive abilities at the middle stratum (Kaufman & Sternberg, 2008). The CHC theory has influenced the development and revision of a number of IQ tests used in gifted identification, including the fifth edition of the Stanford–Binet (Roid, 2003), the second edition of the Kaufman Assessment Battery for Children (KABC-II; Kaufman & Kaufman, 2004), and the third edition of the Woodcock–Johnson Cognitive Abilities Assessment (WJ III; Mather, Wendling, & Woodcock, 2001).

Talent Development Models

The dominant and most familiar conceptualization of giftedness remains the traditional psychometric view. However, there has been growing interest in a group of new models, called *talent development models*. Two talent development models will be described to illustrate the richness and variety of talent developmental models within the gifted field, and how they depart from psychometric views. Because of space limitations, important talent development models will not be covered, most notably the *three-ring conception of giftedness*, proposed by Renzulli (Renzulli, 1984, 2005, 2009; Renzulli & Reis, 2017). Renzulli's writings have had a huge impact on the instructional pedagogy and the differentiated curriculum for gifted learners globally (Pfeiffer, 2013b; Reis & Renzulli, 2009; Renzulli & Reis, 2017).

The Differentiated Model of Giftedness and Talent

Professor Francoys Gagné conceptualizes giftedness as natural abilities which are transformed through learning and experience into high-level skills, in particular occupational fields. In this regard, he views gifts as residing within the child, the result of favorable genetics, prenatal environment, and neurobiological status. Gagné's conceptualization is particularly unique in distinguishing gifts from talents (Gagné, 2005, 2009, 2017). His model, which he calls the differentiated model of giftedness and talent (DMGT), proposes four broad aptitude domains: intellectual, creative, socioaffective, and sensorimotor. Each of these aptitude domains can be subdivided. Gagné acknowledges that many different and competing classification systems exist at this next level (e.g., the intellectual domain could be subdivided into verbal and nonverbal intelligence, fluid and crystallized intelligence, Caroll's (1993) three-level system, Sternberg's (1985; 2001) triarchic and expertise theories, and many other views). The same is true for the other three broad aptitude domains in the DMGT model.

Gagné proposes that talents progressively emerge from the *systematic transformation* of aptitudes—in the case of gifted, high aptitudes, into well-developed skills characteristic of a particular field or domain. This view reflects an appreciation for development over time, viewing "the talent development process consisting of transforming specific natural abilities into the skills that define competence or expertise in a given occupational field" (Gagné, 2005, 103). In this regard, his model is compatible with the ideas proposed by other theorists supportive of a developmental view, such as

Subotnik and her colleagues (Subotnik, 2003; Subotnik, Olszewski-Kubilius, & Worrell, 2011), and my own thinking (Pfeiffer, 2013b, 2015).

Gagné's DMGT conceptualization posits a five-level gifted classification system. He sets the first threshold at 10%, which he labels mildly gifted. This equates to an IQ of approximately 120 with, on average, one in ten students considered mildly gifted. Gagné sets the second threshold at 1%, which he labels moderately gifted: students with IQ scores of 135. The next three levels are for the highly gifted (145 IQ), exceptionally gifted (155 IQ), and extremely gifted (165 IQ) (Gagné, 2017). Gagné's elegant writings have influenced many, including my own thinking. It is one of the first developmental models formulated in response to the field's early emphasis on wholly genetic determinants of giftedness based on high IQ (Kaufman & Sternberg, 2008).

Subotnik's Developmental Transitions in Giftedness and Talent

Rena Subotnik's innovative ideas on talent development parallel the DMGT model proposed by Gagné—they both discuss factors that mediate the full unfolding of gifts. Subotnik's work has had a profound impact on my own thinking. During my tenure as director of Duke's gifted program (Duke TIP), I invited Dr. Subotnik to campus to share her provocative ideas on how a talent development model explains how general and specific abilities transform into competencies, then expertise and ultimately outstanding performance (Subotnik, 2009; Subotnik & Rickoff, 2010). What was particularly compelling was hearing Dr. Subotnik's first-hand experience observing gifted young performing artists, and how these experiences formulated her thinking on talent development and attaining expertise in dance. It parallels my own *learning on the sidelines* experience, observing and working with highly gifted young female soccer players, who reached the highest levels of expertise in their sport *on the pitch* (Pfeiffer, 2015).

Subotnik eloquently observes that giftedness is a dynamic construct and that eminence and real-world creativity develop over time. Her developmental model posits that gifted children transition first from broad educational experiences in the early years to more narrowly focused domains in college, institutes, and conservatories. And if these same highly able learners continue on a trajectory of talent development, they will engage in experiences and opportunities that afford them *the pursuit of scholarly productivity, innovation*, or *artistry*. In other words, her model views "talent development as the transformation of abilities into competencies, competencies into expertise, and expertise into outstanding performance or seminal ideas" (Subotnik, 2009, 155).

Subotnik's developmental model is similar to Gagné's conceptualization. One notable difference is that Subotnik's model extends the vision of gifted education and takes a long view in articulating the ultimate goal of our efforts with high ability kids. She emphasizes that the goal of gifted education should be recruiting and providing a large number of bright kids of high ability with a range of facilitative opportunities and experiences over childhood, adolescence, and even young adult life to maximize the likelihood that as many as possible ultimately reach the highest levels of expertise, creativity, or eminence in different fields. Subotnik envisions talent development as consisting of a series of transitions and stages, with environmental factors and psychosocial variables including motivation, persistence, drive, the will to overcome obstacles, and high interest in a field as all playing central roles in propelling the child along the talent development trajectory. At each stage of development, different factors come into play. For example, in Stage 1 which is a *transition from ability to competency*, high levels of intrinsic motivation, persistence, responsiveness to external rewards, and teachability are critical factors (Subotnik, 2009; Subotnik, Olszewski-Kubilius, & Worrell, 2017; Worrell, Subotnik, & Olszewski-Kubilius, 2017).

Subotnik recognizes that not every domain or field follows exactly the same path, and that future research may illuminate age and gender differences across various domain trajectories. For example, there are likely significant differences in the age on onset and relative influence of facilitative factors that promote expertise in the fields of ballet, architecture, psychotherapy, aerospace physics, and surgical medicine. Subotnik expects that future research may help develop algorithms that predict the relative role that family, school, mentors, psychosocial variables, personality, and community play in the unfolding of talents across different domains. She also believes that definitions of giftedness must change over the course of a child's development and path toward eminence (Subotnik, Olszewski-Kubilius, & Worrell, 2017). Similar to my own thinking, she contends that giftedness should be defined in terms of actual accomplishment (Subotnik, 2003; Worrell, Subotnik, & Olszewski-Kubilius, 2013). This remains a contentious idea in the gifted field.

Stanley's Talent Search Model

As mentioned earlier, Julian Stanley's model incorporates features of both the traditional psychometric view and a talent development model. During my tenure at Duke, I had the good fortune of visiting with Professor Stanley at Johns Hopkins University and became acquainted with the talent search model that he created (Stanley, 1976). Stanley's model is based on an *above-level* testing protocol that is both ingenious and elegant. Stanley was familiar with Hollingsworth's use of above-level testing, in which a student is given a test designed for older students—in other words, above-level (Stanley, 1990). Stanley piloted his model with math prodigies who were given the mathematics section of the Scholastic Aptitude Test in the seventh and eighth grades; he later expanded his talent search beyond math prodigies (Assouline & Lupkowski-Shoplik, 2012).

Stanley's talent search model is predicated on the principle of administering an above-level test (i.e., a test designed for older students) to already-identified bright students (in the top 3 to 5% on grade-level standardized tests); the above-level (also called *out-of-level*) test protocol provides a much higher ceiling to help further differentiate the range of abilities among extraordinarily bright youngsters. Using above-level testing, he was able to *cherry-pick* the very brightest from among an already-select group of high ability students (Park, Lubinski, & Benbow, 2008).

Stanley recognized that discovering exceptionally high ability among the very brightest was not enough. He provided these uniquely gifted youngsters with a different type of intensive (i.e., *high-powered*), highly challenging, and accelerated curriculum and educational experiences on the campus of Johns Hopkins University. At this writing, Stanley's talent search model has expanded exponentially with literally hundreds of summer programs—and weekend, home study, and online educational programs—offered on campuses around the globe for gifted students identified through regional talent searches.

The talent search model is one of the most well-researched and empirically supported models of talent development (Subotnik, Olszewski-Kubilius, & Worrell, 2011). Many talent search students complete one or more years of mathematics in a 3-week summer program (Brody & Benbow, 1987; Kolitch & Brody, 1992; Stanley, 2000). There is considerable empirical support for the predictive validity of this domain-specific gifted identification system used by talent search programs (Olszewski-Kubilius, 2004; Park, et al., 2008; Pfeiffer, 2015). Youths identified before age 13 as demonstrating profound mathematical or verbal reasoning abilities have been tracked longitudinally for nearly three decades. And their outcomes, as a group, have been impressive (Kell, Lubinski, & Benbow, 2013). The Florida Governor's School for Science and Space Technology pilot program, which I co-directed on the campus of Florida State University, in collaboration with NASA and Kennedy Space Center, was designed incorporating Stanley's talent search model. The pilot program considered general measures of intellectual ability in its admissions process, but the admissions review process also put considerable weight on evidence of each applicant's specific abilities and accomplishments in science and math. In selecting finalists for the pilot summer academy affiliated with Kennedy Space Center/NASA, the program also considered each applicant's level of motivation, persistence, and passion for learning—added psychosocial elements that the designers of the program believed sweetened the recipe predicting who would benefit most from our gifted summer academy (Pfeiffer, 2013b).

Expert Performance Perspective

Professor Anders Ericsson has enjoyed a highly successful career investigating the concept of expertise and expert performance and how it is accomplished. His research has focused on identifying and "specifying the mediating mechanisms that can be assessed by process-tracing and experimental studies" (Ericsson, Roring, & Nandagopal, 2007, 13). Ericsson does not believe that IQ tests or intellectual ability, for that matter, play a particularly useful role in predicting performance domains of expertise. He advocates for the power of environmental variables, including what he labels as the importance of *deliberate practice* in explaining extraordinary accomplishments. His writings downplay the relevance of innate ability, heritability, and individual differences in predicting gifted adults (Ericsson, 2014). Critics of his model label his position "the most extreme exemplar of the environmentalist viewpoint" (Ackerman, 2013, 1; 2017).

In a widely cited study of elite chess players, Nobel prizewinner Herbert Simon and William Chase (with whom Ericsson studied) proposed a *ten-year rule*, based on their observations that it took more or less a decade of intensive study and practice to reach the top ranks of chess (see Simon & Chase, 1973). Even Bobby Fisher was no exception (Colvin, 2008). Ericsson agrees with Simon and Chase and has conducted numerous studies that corroborate the idea that *deliberate practice* makes all the difference between expert performers and average adults across almost all domains. Deliberate practice is characterized by several components: it is activity designed specifically to improve performance, often under the watchful eye and close supervision of an instructor, mentor, or coach; it includes a good deal of specific and continuous feedback; it must be repeated a lot; it is highly demanding mentally (Colvin, 2008; Coyle, 2009). Deliberate practice is meant to stretch the individual beyond their comfort zone and beyond their current skills level; it requires that the learner and/or coach identify and isolate very specific elements of performance that need to be learned or improved upon to further development toward expertise (Ericsson, et al., 1993; Syed, 2010). Deliberate practice is effortful, it requires feedback to improve, and as Ericsson and his colleagues remind us, it is not inherently enjoyable (Ericsson, 1996, 2014).

Ericsson's work obviously deserves mention and other contributors to this handbook have prominently represented his important work on expertise. However, the reader may be wondering why a chapter on giftedness includes Ericsson's work on deliberate practice and the acquisition of expert performance. This is a fair question. Some authorities in the gifted field have embraced Ericsson's ideas as relevant to giftedness and talent development. This is despite his spirited opposition to the importance of natural abilities. For example, Ericsson et al. (2007) write, "With the exception of fixed genetic factors determining body size and height, we were unable to find evidence for innate constraints to the attainment of elite achievement for healthy individuals" (3). Some in the gifted field find this conclusion misguided and even heresy. Although I personally am not upset with Ericsson's de-emphasis of individual differences and his extreme environmental position, I don't agree with many of his foundational ideas. They run counter to over forty years of experience working first-hand with many young high ability children and youth. I think that an extreme environmental view ignores considerable research supporting the importance of natural ability, individual differences, early experiences, and critical periods (Ackerman, 2013, 2017; Ackerman & Lakin, 2018; Park, et al., 2008; Pfeiffer, 2015). Irrespective of one's thoughts and opinions about Ericsson's strong environmental position, his emphasis on the importance of deliberate practice, sustained effort in the face of frustration, and years of effortful practice to reach a high level of expertise, eminence, and creativity are important lessons for the gifted field. And as a colleague on the same campus at Florida State, I value his provocative ideas.

Multiple Intelligences Model

The *multiple intelligences model* proposed by Howard Gardner has enjoyed wide appeal by the lay public (Robertson, Pfeiffer, & Taylor, 2011). Gardner is a very recently retired professor at Harvard University and caught the attention of the public in 1983 with the pioneering publication of his highly popular and eminently readable book, *Frames of Mind* (Gardner, 1983), in which he proposed the idea of multiple intelligences. In his model, multiple intelligences are perceived as independent cognitive systems, *not* hierarchically nested under one general ability factor (Gardner, 1983, 1993; Pfeiffer, 2015). Gardner's theory of human intelligences was formulated by a selective analysis of the research literature, not psychometric techniques such as factor analysis. His review and synthesis of a wide-ranging literature in support of his multiple intelligences theory included studies of patients with brain damage, idiot savants, prodigies, evolutionary history, and research in psychometric and experimental psychology. Gardner concluded that there was compelling evidence for at least eight separate intelligences: linguistic, logical-mathematical, spatial, musical, bodily kinesthetic, interpersonal, intrapersonal, and naturalist. He has most recently added a ninth intelligence, existential intelligence, to his list (Dai, 2010; Kaufman & Sternberg, 2008).

Gardner's theory of multiple intelligences has had a significant impact on the educational field internationally, although much less so in the gifted field. His ideas have played a substantial role in greatly expanding educators' views on intelligence. In my international travel, his multiple intelligences theory is often among the first topics that educators want to discuss. Among the most significant aspects of Gardner's theory is the thesis that intelligence is not a single, unitary construct. This single idea, and his highly engaging writing style, helped make Gardner akin to a rock star for some in gifted education. Many intervention programs have been published by followers of the multiple intelligences model. However, there isn't a lot of hard empirical data supporting his theoretical model. The public and lay media have been infatuated with the model, particularly since many incorrectly interpret the multiple intelligences model as implying that everyone is gifted in something.

Hence, Gardner's theory is not without criticism. There is no published research that has tested the multiple intelligences theory. The multiple intelligences that Gardner proposes are based on highly selective reviews of the literature. He omits a considerable amount of the psychometric literature on intelligence, which arguably should be included in any unifying theory of intelligence (Sternberg, 2017). Finally, there exists only a handful of measures of his different intelligences, and most suffer from less-than-adequate psychometric rigor (Kaufman & Sternberg, 2008; Pfeiffer, 2013b).

Theory of Successful Intelligence: WICS

Robert Sternberg recently proposed an alternative way of looking at intelligence. His ideas aren't quite a well-developed model. But his *theory of successful intelligence* is provocative, as many of Sternberg's ideas are, and worth mentioning. Sternberg's theory of successful intelligence emphasizes how three components of intelligence work harmoniously. The components are creativity, intelligence (both academic and practical), and wisdom. Sternberg writes, "Successfully intelligent people balance adaptation to, shaping of, and selection of environments by capitalizing on strengths and compensating for or correcting weaknesses" (Sternberg, Jarvin, & Grigorenko, 2011, 43).

His theory of successful intelligence is referred to as WICS, representing the three components wisdom, intelligence, creativity, synthesized. He believes that giftedness involves both skills and attitudes; the skills are developing competencies and expertise, similar to Subotnik's and Ericsson's ideas. The attitudes are how the gifted individual employs the skills that they have developed. His proposal contends that gifted individuals do not necessarily excel at everything. He believes that gifted individuals are well aware of their strengths and limitations and make the most of their strengths and find ways to compensate for their weaknesses (Sternberg, 2017). This is a cheeky and seductive but yet untested hypothesis.

By creativity, Sternberg means the skills and attitudes needed to generate relatively novel, high quality, and appropriate ideas and products (Pfeiffer, 2015). Sternberg views intelligence as consisting of both those skills and attitudes that we think of when we consider conventional intelligence (or psychometric *g*), and practical intelligence as the skills and attitudes that individuals rely on to solve everyday problems. Sternberg contends that academic skills and attitudes are important for giftedness since gifted individuals need to be able to retrieve, remember, analyze, synthesize, and evaluate information. However, he also argues that practical intelligence is important; gifted individuals need to be able to adapt to their environment, change the environment to suit their needs, or seek a different, more facilitative environment. There is a flavor of Piaget in this notion of practical intelligence. Sternberg contends that ideally gifted individuals need to be high in practical as well as academic intelligence. "Their creativity may help them generate wonderful ideas, but it will not ensure that (gifted individuals) can implement the ideas or convince others to follow the ideas. Many creative individuals have ended up frustrated because they have been unable to convince others to follow their ideas" (Sternberg, Jarvin, & Grigorenko, 2011, 44).

Wisdom is the thorniest and most difficult component to define in his theory of successful intelligence. He believes that individuals are wise to the extent that they use their successful intelligence, creativity, and knowledge to pursue ethical values, balance their own and others' interests, and seek to reach common good. However, it is unclear how to operationalize or measure wisdom or the synthesis of intelligence, creativity, and wisdom. Without a doubt, wisdom is a highly valued character strength, but should not necessarily be viewed as a component of giftedness per se (Pfeiffer, 2013a, 2015). Many view wisdom as a highly valued character strength, virtue, or strength of the heart, but not a component of intelligence, that develops over time and experience (Pfeiffer, 2018b).

Parenthetically, in a seminar on intelligence that I first taught at Duke University, and now teach at Florida State University, as an *ice breaker* I routinely ask my students to rank order from a long list of adjectives those terms that they think most characterize an ideally intelligent, creative, and wise individual. I have used this classroom activity to illustrate the concept of implicit theories of intelligence, creativity, and wisdom, based on Sternberg's early work (Sternberg, 1985). When I have examined the ranking of adjectives over the years, fairly consistent findings exist, suggesting distinct implicit conceptions of how at least college students view the ideally intelligent, ideally creative, and ideally wise person! For example, seven characteristics are consistently selected most frequently as illustrative of an intelligent person: good problem-solving ability, inquisitive, reasons clearly, good at distinguishing between correct and incorrect answers, huge store of information, thinks quickly, and perceptive. The creative person, on the other hand, is characteristically described by college students as imaginative, unorthodox, takes chances, and is emotional, intuitive, and a free spirit. Finally, the wise person is viewed as reflecting four *signature* characteristics: a good listener, thoughtful, listens to all sides of an issue, and considers advice. Although falling short of carefully designed, rigorous research, this classroom activity lends anecdotal support to Sternberg's proposal for three components of successful intelligence.

Tripartite Model of Giftedness

There are many different ways to conceptualize giftedness. No one conceptualization is correct. They are all simply different ways to view children and youth who are in some way special. Each of the different models has implications for who are gifted, how to identify them, and what we should do to educate them and actualize their gifts and talents. As a result of many years' work with highly gifted youth during my tenure at Duke University, I proposed a *tripartite model of giftedness* (Pfeiffer, 2013b, 2015). The tripartite model provides three different ways of viewing students with high ability or extraordinary potential. The tripartite model offers three different, but complementary ways to conceptualize, identify, and program for gifted learners. The three distinct lenses through which high ability students can be viewed within this model are as follows:

- Giftedness through the lens of high intelligence;
- Giftedness through the lens of outstanding accomplishments; and
- Giftedness through the lens of potential to excel.

The first perspective, the *high intelligence* view, is by now familiar. Through this first lens, an IQ test or its proxy can be used to identify students functioning at a certain level considerably above average intellectually. Other tests can supplement the IQ test, but the criterion for high intelligence giftedness is based on compelling evidence that the child is advanced intellectually when compared to his or her same-age peers. This first gifted perspective can follow a general (*g*) or multidimensional view (e.g., C-H-C, structure of intellect, multiple intelligences) of intelligence. It can even be based on a neuroanatomical model of giftedness; recent work, for example, has postulated that