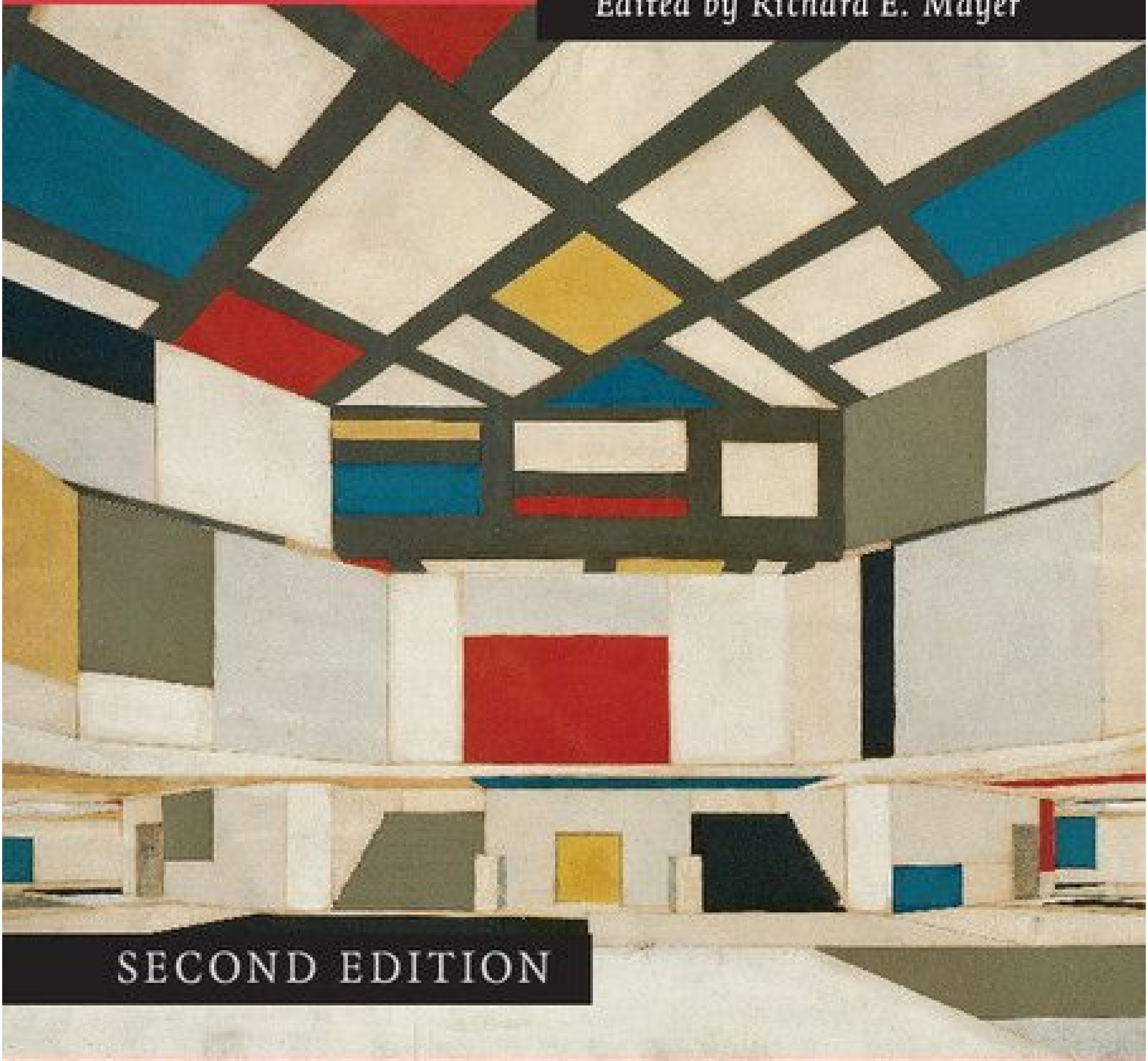
THE CAMBRIDGE HANDBOOK OF

MULTIMEDIA LEARNING

Edited by Richard E. Mayer



SECOND EDITION

maaf jika anda ingin full
kirim vulsa Rp.50.000
ke 087771519313

The Cambridge Handbook of Multimedia Learning

Second Edition

Multimedia learning is learning from words and pictures. The rationale for studying multimedia learning is that people can learn more deeply from words and pictures than from words alone. Multimedia environments, including online presentations, e-courses, interactive lessons, simulation games, slideshows, and even textbooks, play a crucial role in education. *The Cambridge Handbook of Multimedia Learning* is unique in offering a comprehensive, up-to-date analysis of research and theory in the field, with a focus on computer-based learning. Since the first edition appeared in 2005, it has shaped the field and become the primary reference work for multimedia learning. In recent years, multimedia learning has developed into a coherent discipline with a significant research base, which is reflected in the 34 chapters of this handbook. This second edition incorporates the latest developments in multimedia learning, including a sharp increase in the research base, the addition of seven new principles of multimedia learning, a broadening of contexts for studying multimedia learning, a better delineation of boundary conditions for principles, and refinements of theories of multimedia learning.

RICHARD E. MAYER is Professor of Psychology at the University of California, Santa Barbara. He served as President of the Division of Educational Psychology of the American Psychological Association (APA) and as Vice President of the American Educational Research Association (AERA) for the Learning and Instruction Division. He received the E. L. Thorndike Award for career achievement in educational psychology, the Sylvia Scribner Award for research contributions in learning and instruction, and the Distinguished Contributions of Applications of Psychology to Education and Training Award from the American Psychological Association. He has authored more than 400 publications, including 25 books, such as *Applying the Science of Learning*, *Multimedia Learning*, *Learning and Instruction*, and *e-Learning and the Science of Instruction* (with R. Clark).

The Cambridge Handbook of Multimedia Learning

Second Edition

Edited by

Richard E. Mayer

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

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

Cambridge University Press is part of the University of Cambridge.

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

www.cambridge.org
Information on this title: www.cambridge.org/9781107610316

First published 2005

Second edition published 2014

Printed in the United States of America

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

Library of Congress Cataloging in Publication data
The Cambridge handbook of multimedia learning / edited by Richard E. Mayer. – Second Edition.
pages cm – (Cambridge Handbooks in Psychology)
Includes index.
ISBN 978-1-107-03520-1 (Hardback) – ISBN 978-1-107-61031-6 (Paperback)
1. Computer-assisted instruction. 2. Interactive multimedia. I. Mayer, Richard E.
LB1028.5.C283 2014
371.33′4–dc23 2014002252

ISBN 978-1-107-03520-1 Hardback
ISBN 978-1-107-61031-6 Paperback

Contents

Contributors

Shaaron Ainsworth
University of Nottingham

Paul Ayres
University of New South Wales

Roger Azevedo
North Carolina State University

Benoit Bediou
University of Geneva

Anne Britt
Northern Illinois University

Kirsten R. Butcher
University of Utah

Fei Chen
State University of New York, Albany

Michelene T. H. Chi
Arizona State University

Richard E. Clark
University of Southern California

Ruth Colvin Clark
Clark Training & Consulting

Sharon J. Derry
University of North Carolina at Chapel Hill

David F. Feldon
Utah State University

Logan Fiorella
University of California, Santa Barbara

J. D. Fletcher
Institute for Defense Analyses

Arthur C. Graesser
University of Memphis

Mary Hegarty

University of California, Santa Barbara

Xiangen HU

University of Memphis

Allison J. Jaeger

University of Illinois at Chicago

Jeroen Janssen

Utrecht University

Cheryl I. Johnson

U.S. Army Research Institute for the Behavioral and Social Sciences

Ton De Jong

University of Twente

Slava Kalyuga

University of New South Wales

Liesbeth Kester

Open University of the Netherlands

Femke Kirschner

Utrecht University

Paul A. Kirschner

Open University of the Netherlands

Susanne P. Lajoie

McGill University

Ard W. Lazonder

University of Twente

Detlev Leutner

Duisburg-Essen University

Renae Low

University of New South Wales

Richard K. Lowe

Curtin University

Richard E. Mayer

University of California, Santa Barbara

Benjamin D. Nye

University of Memphis

Fred Paas

Erasmus University Rotterdam

Celeste Pilegard

University of California, Santa Barbara

Jan L. Plass

New York University

Heather A. Priest

U.S. Army Research Institute for the Behavioral and Social Sciences

Alexander Renkl

University of Freiburg

Jean-FranÇois Rouet

University of Poitiers

Christopher A. Sanchez

Oregon State University

Katharina Scheiter

Leibniz-Institut für Wissensmedien

Annett Schmeck

Duisburg-Essen University

Wolfgang Schnotz

University of Koblenz-Landau

Ruth N. Schwartz

New York University

Bruce L. Sherin

Northwestern University

Miriam Gamoran Sherin

Northwestern University

John Sweller

University of New South Wales

Sigmund Tobias

State University of New York, Albany

Tamara Van Gog

Erasmus University Rotterdam

Jeroen J. G. Van MerriËNboer

Maastricht University

Jennifer Wiley

University of Illinois at Chicago

Alexander P. Wind

State University of New York, Albany

Ruth Wylie

Arizona State University

Preface

As the first and only comprehensive research-based handbook on multimedia learning, *The Cambridge Handbook of Multimedia Learning* has helped define and shape the field and has become recognized as its major reference work. Since the publication of the first edition in 2005, the field of multimedia learning has grown as a coherent discipline with an accumulated research base worthy of being synthesized and organized in an updated handbook. Therefore, I am pleased to serve as editor of this second edition of *The Cambridge Handbook of Multimedia Learning*, which remains the most comprehensive and up-to-date volume summarizing research and theory in the field of multimedia learning.

This second edition of the *Handbook* constitutes the latest progress report from the world's leading multimedia researchers. As in the first edition, the focus of this volume is on how people learn from words and pictures, particularly in computer-based environments. For purposes of the *Handbook*, multimedia learning is defined as learning from words (e.g., spoken or printed text) and pictures (e.g., illustrations, photos, maps, graphs, animation, or video). Multimedia environments include online instructional presentations, interactive lessons, e-courses, simulation games, slideshows, and even textbooks. Overall, the major goal of this second edition of the *Handbook* is the same as that of the first edition – to establish what works (by systematically examining research-based principles of effective multimedia instruction) and to explain how it works (by grounding the research findings in cognitive theory).

There are many books providing advice on how to design multimedia learning environments, but they are based largely on the practical experience and wisdom of the authors. Similarly, there are books reporting on the development of online instructional programs and Web sites, but the development efforts are generally based on best practices and informal case studies. Until recently, the lack of scientific research evidence in many multimedia learning books could be justified on the grounds that a solid research base did not yet exist. However, the quantity and quality of scientific research – conducted by researchers around the world – have reached a level warranting a revision of the field's first comprehensive research-based handbook of multimedia learning.

What distinguishes this book from some other books on distance learning or Web-based instruction is our commitment to taking a scientific, evidence-based approach. My goal as editor is to make sure the *Handbook* provides a comprehensive and focused overview of the state of scientific research on multimedia learning. Each chapter is based on empirical research and grounded in cognitive theory, rather than offering unsubstantiated recommendations, describing best practices, or summarizing software development accomplishments. The chapter authors are research leaders from around the world, who have records of research publication in multimedia learning. As the most comprehensive research-based handbook on multimedia learning, the second edition of *The Cambridge Handbook of Multimedia Learning* is intended to continue to define and shape the field for years to come.

As the editor, I asked leading multimedia researchers to author chapters in areas in which they have contributed to the empirical research base. Because the field is largely international, the chapter authors span the globe, with more than half the chapters written by authors outside the United States – including chapter authors from Australia, Canada, Germany, France, the Netherlands, and the United Kingdom.

The *Handbook* consists of 34 chapters organized into five parts. Each chapter focuses on a particular theory of multimedia learning (in Part I), a basic principle of multimedia learning (Part II), an advanced principle of multimedia learning (Part III), multimedia learning of cognitive processing (Part IV), or multimedia learning within an advanced computer-based context (Part V).

In order to provide a common structure among the chapters, I asked authors to organize their chapters around a common set of issues. In particular, I asked the authors of the theory chapters in Part I to provide a concise description of the theory or model with concrete examples, to summarize the theory's contributions to cognitive theory (i.e., to specify predictions that have been tested), to summarize the theory's contributions to instructional design (i.e., to specify recommendations for instruction), to describe any limitations of the theory, and to suggest future directions for research. I asked the authors of each of the other chapters to provide a clear definition and example of the principle or topic of the chapter, to review the relevant published research literature in sufficient detail, to assess the limitations of the research base, to summarize the implications for cognitive theory and for instructional design, and to suggest directions for future research.

I solicited chapters that were concise (i.e., containing no more than 40 double-spaced pages), focused (i.e., reviewing the research on the specified topic), well-referenced (i.e., containing a rich set of relevant references), evidence-based (i.e., providing an up-to-date review of the best empirical evidence), theory-based (i.e., relating the findings to testable predictions of theories when appropriate), and educationally relevant (i.e., drawing implications for educational practice when appropriate). In order to minimize confusion, I asked the authors to clearly define jargon terms in the text as well as in a glossary at the end of the chapter. Each chapter was reviewed and revised.

This book is for anyone interested in how people learn from words and pictures in computer-based environments. Although the *Handbook* summarizes the research base in multimedia learning, it is intended to be accessible to a general audience. On one hand, the *Handbook* is designed to support readers with practical interests in how to design or select multimedia learning environments that promote learning. On the other hand, it is designed to support readers who have academic interests in conducting or evaluating research in multimedia learning. The *Handbook* would be appropriate for courses related to cognitive science, educational psychology, instructional design, human factors, multimedia arts and technology, professional training, and interface design. It would also be useful for instructors interested in designing or improving multimedia lessons in school settings, job training contexts, and informal environments. In short, *The Cambridge Handbook of Multimedia Learning* belongs on the bookshelf of anyone who is interested in taking an evidence-based approach to Web-based learning, e-learning, hypermedia, multimedia, computer games, Web site design, distance learning, instructional technology, human–computer interaction, virtual environments, or applied cognitive psychology.

As the editor, I have tried to ensure that the *Handbook* reflects the values that I think are important for our field. In particular, I sought to produce a handbook with the following characteristics:

> *Research-based*: The *Handbook* is intended to summarize the empirical research on multimedia learning rather than describe untested best practices or software development projects. Although I have much respect for the craft knowledge of practitioners and designers, it is important to know if recommendations are supported by scientific evidence and under what conditions they are supported. Thus, I value a focus on scientific evidence as the key to progress in our field.
> *Theory-grounded*: The *Handbook* is intended to relate empirical research to cognitive theories of how people learn. My overriding premise is that multimedia learning environments should be designed in ways that are consistent with what is known about how people learn.
> *Educationally relevant*: The *Handbook* focuses on issues that are relevant to education, that is, to helping people learn. Thus, I sought chapters that offer research-based implications for instructional design.
> *Comprehensive*: The *Handbook* offers a broad view of the field, including contributions from multimedia researchers around the world. I value the perspectives of researchers who have devoted so much of their energy to understanding multimedia learning.
> *Timely*: The *Handbook* offers an up-to-date overview of the field. I value timelines because the scientific study of multimedia learning is maturing at a rapid pace, and so are the practical demands for building multimedia learning environments – ranging from e-courses to in-class simulations.
> *Readable* – In my role as editor I have tried to make sure that the chapters are clear and concise, with key terms defined and concrete examples provided. In a multidisciplinary field like this one, it is important that the chapters communicate what is known in a way that general readers can appreciate.

In short, my values motivated me to seek chapters based on empirical research and grounded in cognitive theory rather than chapters that mainly describe development efforts or best practices.

In order to prepare for the second edition, I solicited suggestions from 12 leading multimedia researchers concerning new chapters to add, old chapters to delete or reshape, and new authors to include. I also examined notes and comments I had received and made concerning the first edition, and I examined the current state of the field in terms of research activity. In light of this analysis, I sought to retain the *Handbook*'s basic goal and structure but to ask authors to update and revise their chapters.

This second edition of the *Handbook* begins (in Part I) with a look at four foundational theories of multimedia learning, each of which has been updated since the previous edition – Sweller's cognitive load theory, Mayer's cognitive theory of multimedia learning, Schnotz's integrated model of text and picture comprehension, and van Merriënboer's four-component instructional design theory.

As in the first edition, each core chapter (in Parts II and III) focuses on a well-established effect or principle that has been researched extensively. On the basis of developments in the field, I added a chapter on the signaling principle (i.e., highlighting parts of a graphic during instruction) to Part II. In Part III, I added chapters on the drawing principle (i.e., asking learners to draw during learning), feedback principle (i.e., giving explanations after learner responses), multiple representation principle (i.e., using different modes to present the information), animation principle (i.e., presenting graphics in dynamic form), learner control principle (i.e., allowing the learner to make choices about the pace and order of instruction), and working memory principle (i.e., the role of individual differences in working memory). I deleted chapters on aging, site maps, and navigation and incorporated much of the material into other chapters to better reflect the development of the research base during the past decade.

In ensuing chapters (in Parts IV and V), I asked the authors to examine the research base in specific contexts of multimedia learning such as teaching of metacognitive skills in a hypertext environment or teaching of cognitive skills using educational games. I reshaped Part IV to focus on multimedia learning of specialized content – cognitive skills, metacognitive strategies, and reasoning about complex systems – which has grown rapidly in the past 10 years, and to downplay multimedia learning in subject areas – deleting chapters on reading, mathematics, history, chemistry, meteorology, and second-language learning, which are better covered in other chapters. In Part V, I broke the chapter on simulations and games into two separate chapters to better reflect the growth of both of those areas, I substituted a chapter on multimedia learning with intelligent tutoring systems for chapters on multimedia learning with pedagogical agents and in virtual reality to also better reflect current research directions, I substituted a chapter on multimedia learning from multiple sources for one on hypermedia, and I added a chapter on learning with video to reflect the development of a solid research base.

Editing this book has been a treat for me, because I could commission chapters from the best researchers in the field and be the first to learn what they had to say. I am pleased to share the fruits of this enterprise with you in a timely fashion. My hope is that you will enjoy reading this *Handbook* as much as I have enjoyed editing it. I will consider it a success if it helps you to understand what is known about how people learn from words and pictures, gives you useful help in building or selecting effective multimedia learning environments, or encourages you to produce or investigate research that contributes to cognitive theory and educational practice. I hope that you will feel free to contact me at mayer@psych.ucsb.edu to share your comments about the *Handbook*.

Acknowledgments

Although my name is listed as the editor, this *Handbook* depended on the contributions of many people. In particular, I wish to thank the chapter authors for producing excellent chapters, for keeping this project on schedule, and for responding so well to the reviewers' comments. In particular, I wish to thank John Sweller and Jeroen van Merriënboer, who, in addition to providing outstanding chapters, gave me invaluable feedback on my own chapters. I also wish to thank David Repetto, Matthew Bennett, Hetty Marx, and the staff of Cambridge University Press for their many contributions to making this book a success. I am grateful to my many research collaborators who have worked with me over the years in the study of multimedia learning, including Deanne Adams, Richard B. Anderson, Robert Atkinson, Julie Campbell, Paul Chandler, Dorothy Chun, Ruth Clark, Krista DeLeeuw, Gayle Dow, Logan Fiorella, Joan Gallini, Shannon Harp, Mary Hegarty, Julie Heiser, Nabil Issa, Cheryl Johnson, Lewis Johnson, Claudia Leopold, James Lester, Detlev Leutner, Steven Lonn, Patricia Mautone, Sarah Mayer, Bruce McLaren, Roxana Moreno, Harold O'Neil, Jr., Celeste Pilegard, Jan Plass, Hector Ponce, Annett Schwamborn, Valerie Sims, Hiller Spires, Andy Stull, Eunmo Sung, and Ning Wang. I appreciate my home institution – the University of California, Santa Barbara – and numerous funding agencies, including the National Science Foundation, the Office of Naval Research, the U.S. Department of Education, and the Andrew Mellon Foundation, which have supported my research on multimedia learning. Finally, my deepest appreciation goes to my wife, Beverly; my children, Ken, David, and Sarah; my grandchildren, Jacob, Avery, James, Emma, and Caleb; and to the memory of my parents, James and Bernis Mayer.

1 Introduction to Multimedia Learning

Richard E. Mayer

University of California, Santa Barbara

Abstract

Multimedia learning is learning from words and pictures. The rationale for studying multimedia learning is that people can learn more deeply from words and pictures than from words alone. A goal of research on multimedia learning is to understand how to design multimedia learning environments that promote meaningful learning. The research base concerning multimedia learning is reflected in the 34 chapters of this handbook. What is new in this second edition is a sharp increase in the research base, the addition of seven new principles of multimedia learning, a broadening of contexts for studying multimedia learning, a better delineation of boundary conditions for principles, and refinements of theories of multimedia learning. The approach taken in this handbook is learner-centered rather than technology-centered, views learning as a constructive process rather than solely as a process of adding new information to memory or strengthening associations, seeks to foster meaningful learning rather than rote learning, and favors appropriate cognitive activity during learning rather than behavioral activity per se.

Introduction

People can learn more deeply from words and pictures than from words alone. This seemingly simple proposition – which can be called the *multimedia learning hypothesis* – is the main focus of this second edition of *The Cambridge Handbook of Multimedia Learning*.[1] Each of the 34 chapters examines an aspect of the multimedia learning hypothesis. In particular, multimedia researchers are interested in how people learn from words and pictures and in how to design multimedia learning environments that promote learning. In this chapter, I provide a definition of multimedia learning, offer a rationale for multimedia learning, outline the research base for multimedia learning, summarize changes since the first edition, and draw distinctions between two approaches to multimedia design, three metaphors of multimedia learning, three kinds of multimedia learning outcomes, and two kinds of active learning.

What Is Multimedia Learning?

Table 1.1 summarizes definitions of multimedia, multimedia learning, and multimedia instruction.

Table 1.1. ***Definitions***

Term	Definition
Multimedia	Presenting words (such as printed text or spoken text) and pictures (such as illustrations, photos, animation, or video)
Multimedia	Building mental representations from words and pictures

learning	
Multimedia instruction	Presenting words and pictures that are intended to promote learning

Multimedia

The term *multimedia* conjures up a variety of meanings. You might think of watching a podcast on your smartphone or playing a strategy game on your tablet – that is, multimedia as a handheld experience. You might think of sitting in a room where images are presented on one or more screens and music or other sounds are presented via speakers – that is, multimedia as a "live" performance. Alternatively, you might think of sitting at a computer screen that presents graphics on the screen along with spoken words from the computer's speakers – that is, multimedia as an online lesson. Other possibilities include watching a video on a TV screen while listening to the corresponding words, music, and sounds or watching a PowerPoint presentation along with listening to the speaker's corresponding commentary. Low-tech examples of multimedia include a *chalk-and-talk* presentation, in which a speaker draws or writes on a blackboard (or uses an overhead projector) while presenting a lecture or a textbook lesson consisting of printed text and illustrations. In sum, most academic learning situations involve multimedia learning because students encounter words and graphics.

I define *multimedia* as presenting both words (such as spoken text or printed text) and pictures (such as illustrations, photos, animation, or video). By words, I mean that the material is presented in *verbal form*, such as printed text or spoken text. By pictures, I mean that the material is presented in *pictorial form*, such as static graphics, including illustrations, graphs, diagrams, maps, or photos, or dynamic graphics, including animation or video. This definition is broad enough to include all of the scenarios I described in the preceding paragraph – ranging from multimedia encyclopedias to online educational games to textbooks. For example, in a multimedia encyclopedia, words may be presented as narration and pictures may be presented as animation. In a textbook, words may be presented as printed text and pictures may be presented as illustrations. In an online educational game, on-screen characters may speak as they show you how to accomplish some task.

If multimedia involves presenting material in two or more forms, then an important issue concerns how to characterize a form of presentation. Three solutions to this problem are the delivery media view, the presentation modes view, and the sensory modalities view. According to the delivery media view, multimedia requires two or more delivery devices, such as a computer screen and amplified speakers or a projector and a lecturer's voice. According to the presentation modes view, multimedia requires verbal and pictorial representations, such as on-screen text and animation or printed text and illustrations. According to the sensory modalities view, multimedia requires auditory and visual senses, such as narration and animation or a lecture and slides.

I reject the delivery media view because it focuses on the technology rather than on the learner. Instead, I opt for the presentation modes view and, to some extent, the sensory modalities view. The presentation modes view allows for a clear definition of multimedia – presenting material in verbal and pictorial form – and is commonly used by multimedia researchers (Mayer, 2009). The presentation modes view is also the basis for Paivio's (1986, 2006) dual-code theory, as well as theories of multimedia learning presented in this handbook (Chapter 2, by Paas and Sweller; Chapter 3, by Mayer; Chapter 4, by Schnotz; and Chapter 5, by van Merriënboer and Kester). The sensory modalities view is also relevant because words can be presented as printed text (initially processed visually) or as spoken text (initially processed auditorily), whereas pictures are processed visually. In conclusion, as shown in Table 1.1, multimedia refers to using words and pictures.

Multimedia learning

Multimedia learning occurs when people build mental representations from words (such as spoken text or printed text) and pictures (such as illustrations, photos, animation, or video). As you can see from this definition, multimedia refers to the presentation of words and pictures, whereas multimedia learning refers to the learner's construction of knowledge from words and pictures. The process by which people build mental representations from words and pictures is the focus of Mayer's cognitive theory of multimedia learning (Mayer, 2009; see also Chapter 3), Sweller's cognitive load theory (Sweller, Ayres, & Kalyuga, 2011; see also Chapter 2), Schnotz's integrative model of text and picture comprehension (Schnotz & Bannert, 2003; see also Chapter 4), and, to some extent, van Merriënboer's four-component instructional design theory (van Merriënboer & Kirschner, 2007; see also Chapter 5).

Multimedia instruction

Multimedia instruction (or a multimedia learning environment) involves presenting words and pictures that are intended to promote learning. In short, multimedia instruction refers to designing multimedia learning environments in ways that help people build mental representations. The instructional design principles described in Parts II and III suggest ways of creating multimedia lessons intended to promote multimedia learning, and Parts IV and V offer examples of how the principles can be applied in a variety of advanced contexts ranging from educational games to intelligent tutoring systems.

What Is the Rationale for Multimedia Learning?

What is the value of adding pictures to words? Do students learn more deeply from words and pictures than from words alone? These questions are essential to the study of multimedia learning. For example, suppose I asked you to listen to a short explanation of how a bicycle tire pump works: "When the handle is pulled up, the piston moves up, the inlet valve opens, the outlet valve closes, and air enters the lower part of the cylinder. When the handle is pushed down, the piston moves down, the inlet valve closes, the outlet valve opens, and air moves out through the hose." Then I ask you to write down an explanation of how a bicycle tire pump works (i.e., retention test) and to write answers to problem-solving questions such as "Suppose you push down and pull up the handle of a pump several times but no air comes out. What could have gone wrong?" (i.e., transfer test). If you are like most of the students in our research studies (Mayer & Anderson, 1991, 1992), you remembered some of the words in the presentation (i.e., you did moderately well on retention) but you had difficulty using the material to answer problem-solving questions (i.e., you did poorly on transfer).

In contrast, suppose I showed you an animation of a bicycle tire pump that depicts the actions in the pump as the handle is pulled up and then as the handle is pushed down. Frames from the animation are shown in Figure 1.1. If you are like most students in our research studies (Mayer & Anderson, 1991, 1992), you would not do well on a retention test or on a transfer test.

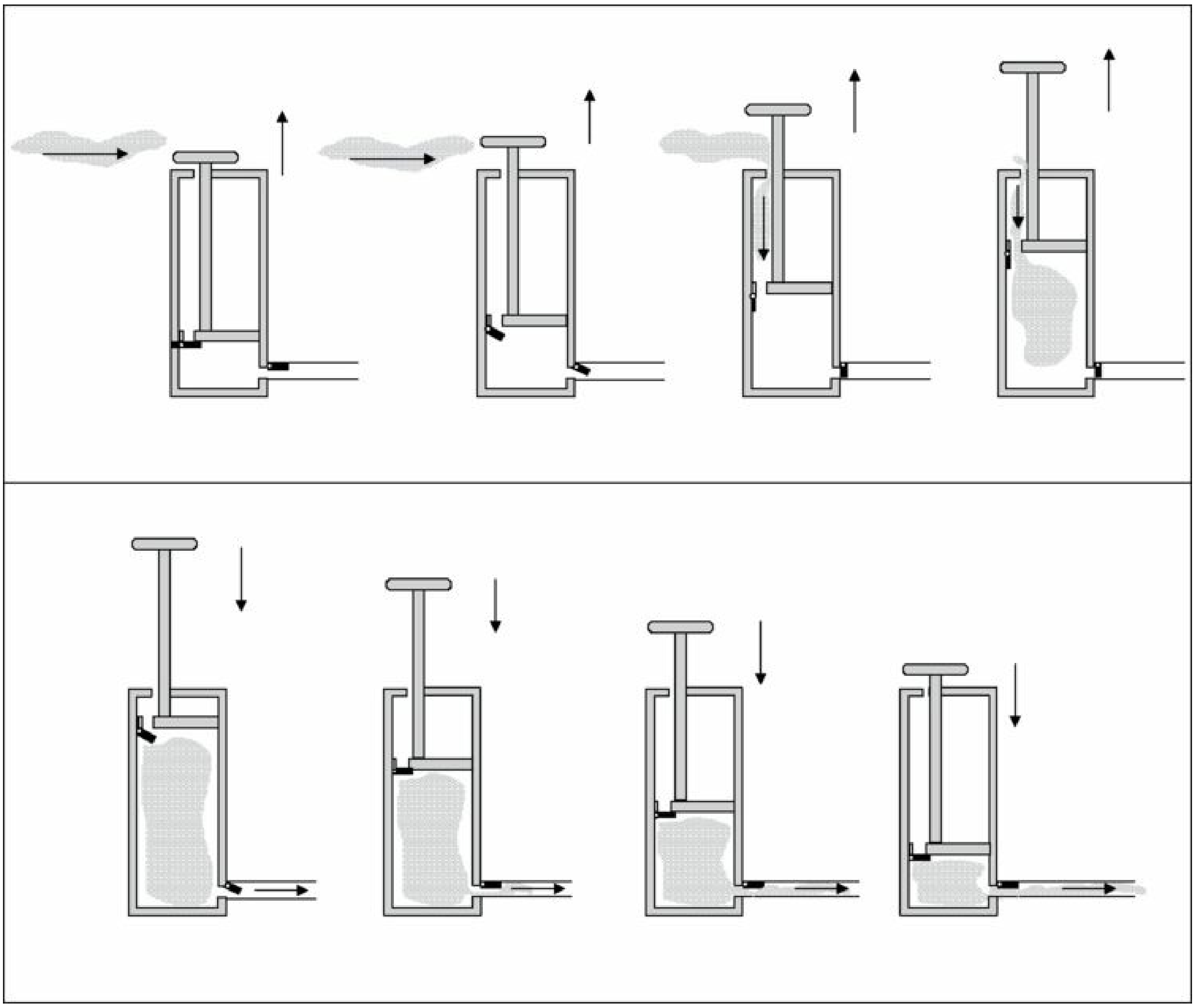
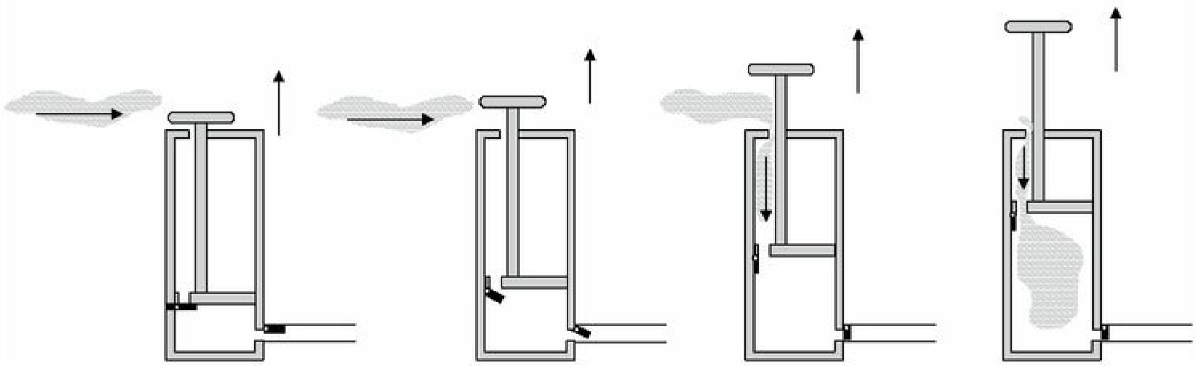
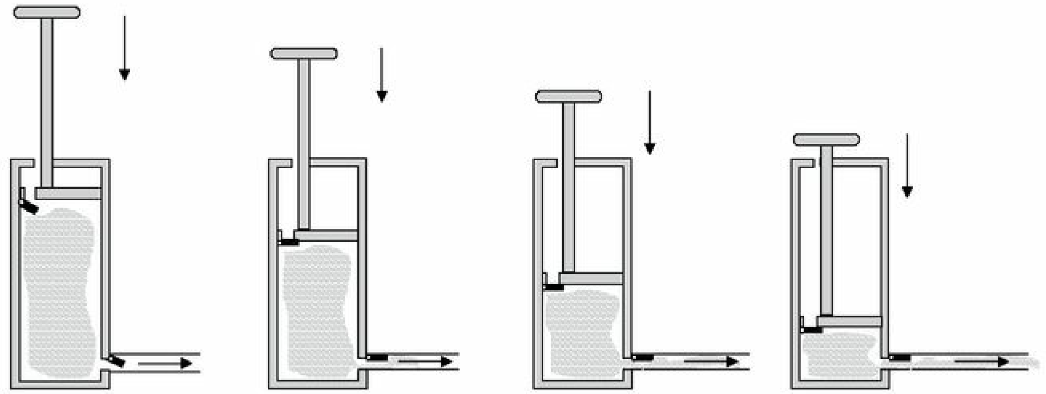


Figure 1.1. Frames from a pumps animation.

Finally, consider the narrated animation summarized in Figure 1.2. In this situation, you hear the steps described in words as you see the steps depicted in the animation. When words and pictures are presented together as in a narrated animation, students perform well both on retention and on transfer tests (Mayer & Anderson, 1991, 1992). In particular, when we focus on tests of problem-solving transfer – which are designed to measure a student's understanding of the presented material – students perform much better with words and pictures than with words alone. My colleagues and I found this pattern in nine out of nine studies, yielding a median effect size of 1.50 (Mayer, 2009). I refer to this finding as the *multimedia principle*, and it is examined in detail by Butcher in Chapter 7.



“When the handle is pulled up, the piston moves up, the inlet valve opens, the outlet valve closes, and air enters the lower part of the cylinder.”



“When the handle is pushed down, the piston moves down, the inlet valve closes, the outlet valve opens, and air moves out through the hose.”

Figure 1.2. Frames from a pumps animation with corresponding narration.

The multimedia principle epitomizes the rationale for studying multimedia learning. There is reason to believe that, under certain circumstances, people learn more deeply from words and pictures than from words alone. For hundreds of years, the major format for instruction has been words, including lectures and books. In general, verbal modes of presentation have dominated the way we convey ideas to one another, and verbal learning has dominated education. Similarly, verbal learning has been the major focus of educational research.

With the recent advent of powerful computer graphics and visualization technologies, instructors have the ability to supplement verbal modes of instruction with pictorial modes of instruction. Advances in computer technology have led to an explosion in the availability of visual ways of presenting material, including large libraries of static images as well as compelling dynamic images in the form of animations and video. In light of the power of computer graphics, it may be useful to ask whether it is time to expand instructional messages beyond the purely verbal. What are the consequences of adding pictures to words? What happens when instructional messages involve both verbal and visual modes of learning? What affects the way that people learn from words and pictures? In short, how can multimedia presentations foster meaningful learning? These are the kinds of questions addressed in this handbook.

The case for multimedia learning is based on the idea that instructional messages should be designed in light of how the human mind works. Let’s assume that humans have two information processing systems – one for verbal material and one for visual material, as described more fully in Part I. Let’s also acknowledge that the major format for presenting instructional material is verbal. The rationale for multimedia presentation – that is, presenting material in words and pictures – is that it takes advantage of the full capacity of humans for processing information. When we present material only in the verbal mode, we are ignoring the potential contribution of our capacity to also process material in the visual mode.

Why might two channels be better than one? Two possible explanations are the quantitative rationale and the qualitative rationale. The quantitative rationale is that more material can be presented on two channels than on one channel – just as more traffic can travel on two lanes than on one lane. In the case of explaining how a bicycle tire pump works, for example, the steps in the process can be presented in words or can be depicted in illustrations. Presenting both is like presenting the material twice – giving the learner twice as much exposure to the explanation. While the quantitative rationale makes sense as far as it goes, I reject it mainly because it is incomplete. In particular, I take exception to the assumption that the verbal and visual channels are equivalent – that is, that words and pictures are simply two equivalent ways of presenting the same material.

In contrast, the qualitative rationale is that words and pictures, while qualitatively different, can complement one another and that human understanding is enhanced when learners are able to mentally integrate visual and verbal representations. As you can see, the qualitative rationale assumes that the two channels are not equivalent; words are useful for presenting certain kinds of material – perhaps representations that are more abstract and require more effort to translate – whereas pictures are more useful for presenting other kinds of material – perhaps more intuitive, more natural representations. In short, one picture is not necessarily the same as a thousand words (or any number of words).

The most intriguing aspect of the qualitative rationale is that understanding occurs when learners are able to build meaningful connections between pictorial and verbal representations – such as being able to see how the words "the inlet valve opens" relate to the forward motion of the inlet valve in the cylinder of the pump. In the process of trying to build connections between words and pictures, learners are able to create a deeper understanding than from words or pictures alone. This idea is at the heart of the theories of multimedia learning described in Part I.

In summary, the rationale for the study of multimedia learning is that students may learn more deeply from words and pictures than from words alone. Thus, a primary purpose of this handbook is to explore the proposal that adding pictures to words may promote greater understanding than simply presenting words alone. However, not all pictures are equally effective. It is important to understand how best to incorporate pictures with words. Just because technologies are available that allow for state-of-the-art visualizations, this does not mean that instructors are well advised to use them. What is needed is a research-based understanding of how people learn from words and pictures and how to design multimedia instruction that promotes learning.

What Is the Research Base for Multimedia Learning?

Although research on verbal learning has a long and fruitful history in psychology and education, corresponding research on multimedia learning is just beginning to flourish. This second edition of *The Cambridge Handbook of Multimedia Learning* remains the world's first and most comprehensive summary of research on multimedia learning. In an attempt to organize the research base in multimedia learning, it is divided into five parts.

Part I – "Theoretical Foundations" – contains chapters that describe theories that are relevant to multimedia learning and that have had the greatest impact on research: Sweller's cognitive load theory (Chapter 2), Mayer's cognitive theory of multimedia learning (Chapter 3), Schnotz's integrative model of text and picture comprehension (Chapter 4), and van Merriënboer's four-component instructional design model for multimedia learning (Chapter 5).

Part II – "Basic Principles of Multimedia Learning" – begins with a chapter documenting questionable principles of multimedia learning, that is, principles that are commonly accepted but for which supporting evidence is lacking (Chapter 6, by Clark and Feldon). The remaining chapters explore the research evidence concerning basic principles for designing multimedia learning environments:

Multimedia principle:	People learn better from words and pictures than from words alone (Chapter 7, by Butcher).
Split-attention	People learn better when words and pictures are physically and temporally

principle:	integrated (Chapter 8, by Ayres and Sweller), similar to Mayer’s spatial contiguity and temporal contiguity principles (Chapter 12).
Modality principle:	People learn better from graphics and narration than from graphics and printed text (Chapter 9, by Low and Sweller), similar to Mayer’s modality principle (Chapter 13).
Redundancy principle:	People learn better when the same information is not presented in more than one format (Chapter 10, by Kalyuga and Sweller), similar to Mayer’s redundancy principle (Chapter 12).
Signaling principle:	People learn better when cues are added that highlight the key information and its organization (Chapter 11, by van Gog), similar to Mayer’s signaling principle (Chapter 12).
Coherence, signaling, spatial contiguity, temporal contiguity, and redundancy principles:	People learn better when extraneous material is excluded rather than included, when cues are added that highlight the organization of the essential material, and when corresponding words and pictures are presented near rather than far from each other on the screen or page or in time, and people learn better from graphics and narration than from graphics, narration, and on-screen text (Chapter 12, by Mayer and Fiorella).
Segmenting, pre-training, and modality principles:	People learn better when a multimedia message is presented in learner-paced segments rather than as a continuous unit, people learn better from a multimedia message when they know the names and characteristics of the main concepts, and people learn better from a multimedia message when the words are spoken rather than written (Chapter 13, by Mayer and Pilegard).
Personalization, voice, embodiment, and image principles:	People learn better when the words of a multimedia presentation are in conversational style rather than formal style, when the words are spoken in a standard-accented human voice rather than a machine voice or foreign-accented human voice, and when on-screen agents display humanlike gestures and movements; but people do not necessarily learn better when the speaker’s image is on the screen (Chapter 14, by Mayer).

Part III – “Advanced Principles of Multimedia Learning” – contains chapters that explore the research evidence for advanced principles of multimedia learning:

Guided discovery principle:	People learn better when guidance is incorporated into discovery-based multimedia environments (Chapter 15, by de Jong and Lazonder).
Worked examples principle:	People learn better when they receive worked examples in initial skill learning (Chapter 16, by Renkl).
Self-explanation principle:	People learn better when they are encouraged to generate self-explanations during learning (Chapter 17, by Wylie and Chi).
Drawing principle:	People learn better when they create drawings as they read explanative text (Chapter 18, by Leutner and Schmeck).
Feedback principle:	People learn better from multimedia lessons when they receive explanative feedback on their performance (Chapter 19, by Johnson and Priest).
Multiple representation principle:	There are circumstances under which people learn better from multiple representations (Chapter 20, by Ainsworth).
Learner control	People do not necessarily learn better when they have more control of the

principle:	selection and organization of the material (Chapter 21, by Scheiter).
Animation principle:	People do not necessarily learn better from animation than from static diagrams (Chapter 22, by Lowe and Schnotz).
Collaboration principle:	People can learn better with collaborative online learning activities (Chapter 23, by Kirschner, Kirschner, and Janssen).
Prior knowledge principle:	Instructional design principles that enhance multimedia learning for novices may hinder multimedia learning for more expert learners (Chapter 24, by Kalyuga).
Working memory principle:	The effectiveness of instructional design principles depends on the learner's working memory capacity (Chapter 25, by Wiley, Sanchez, and Jaeger).

Part IV – "Multimedia Learning of Cognitive Processes" – takes a somewhat different cut by examining research on how to design multimedia learning to support higher-level cognition. The chapters summarize research on multimedia learning of cognitive skills (Chapter 26, by Lajoie), metacognitive strategies (Chapter 27, by Azevedo), and reasoning about complex systems (Chapter 28, by Hegarty).

Finally, the chapters in Part V – "Multimedia Learning in Advanced Computer-Based Contexts" – examine multimedia learning research involving emerging technologies. The chapters summarize research on multimedia learning with advanced technologies that have generated the most research, such as intelligent tutoring systems (Chapter 29, by Nye, Graesser, and Hu), simulations and microworlds (Chapter 30, by Plass and Schwartz), games (Chapter 31, by Tobias et al.,), video (Chapter 32, by Derry, Sherin, and Sherin), multiple sources (Chapter 33, by Rouet and Britt), and e-courses (Chapter 34, by Clark).

In all of the chapters the focus is on empirical research evidence, including implications of research for theory and practice. Overall, each chapter is intended to showcase the research base in a sub-area of multimedia learning, note its limitations, and offer suggestions for future research.

What's New in the Second Edition?

Although the general goals remain the same (i.e., to take an evidence-based approach to the design of multimedia instruction), there are five major changes in this second edition of the *Handbook*: an increase in the research base, the addition of new topics, a broadening of contexts of studying multimedia learning, an identification of boundary conditions, and a refinement of theory.

Increase in the research base

The second edition reflects the strong growth of the empirical research base in the field of multimedia learning, with many new references beyond those found in the previous edition. The book contains all of the basic principles of multimedia learning (i.e., multimedia, split attention, modality, redundancy, signaling, coherence, spatial contiguity, temporal contiguity, segmenting, pre-training, modality, personalization, voice, and image) and most of the advanced principles of multimedia learning (i.e., guided discovery, worked examples, self-explanation, collaboration, and prior knowledge) found in the first edition, but the principles are now informed by a much richer evidence base.

In some basic multimedia principles, the research base has more than doubled since the publication of the first edition in 2005. For example, in Chapter 12 on the coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles, the total number of experimental comparisons across all five principles in the first edition was 40, compared with 99 in the second edition, reflecting an increase of more than 100%. Similarly, in Chapter 13 on the segmenting, pre-training, and modality principles, the total number of experimental comparisons across all three principles was 31 in the first edition, compared with 87 in the second edition, reflecting an increase of more than 100%. Finally, Chapter 14 on the personalization, voice, image, and embodiment principles reported a total of 23 experimental comparisons in the first edition, compared with 48 in the second edition, also reflecting an increase of more than 100%. The growing research base also is reflected in a proliferation of meta-analyses of multimedia principles that have appeared since the first edition (e.g., Adesope & Nesbit, 2012; Ginns, 2005, 2006; Ginns, Marin, & Marsh, 2013), compared with none reported in the first edition. Overall, this edition of *The Cambridge Handbook of Multimedia Learning* reflects strong growth in what we know about how to design effective multimedia instruction.

Addition of new topics

This second edition is organized into the same number of parts as the first edition, but each has undergone some change. Parts II and III of the second edition contain seven new chapters. First, there are now separate chapters on the signaling principle, animation principle, and learner control principle, which were only chapter sections in the first edition. Second, there are new chapters on the drawing principle, feedback principle, and multiple representation principle. Third, in addition to a chapter on the role of prior knowledge also found in the first edition, this edition adds a new chapter on the role of working memory as another important individual differences consideration in multimedia design. These seven additions reflect the growing number of evidence-based principles of multimedia instructional design that now have substantial research bases. Material from previous chapters on the navigational principle and the site map principle has been subsumed in other chapters, such as the newly added chapter on the signaling principle. The previous chapter on the cognitive aging principle has been removed to make way for areas that have shown greater research growth, such as the working memory principle.

Broadening of contexts of studying multimedia learning

In this second edition, there are more studies of multimedia learning with new media and in new contexts. Instead of there being a part on multimedia learning in subject areas, as was the case in the preceding edition, there is now a part (Part IV) on multimedia learning of cognitive processes, which contains chapters on multimedia learning of cognitive skills, metacognitive strategies, and reasoning about complex physical systems. This shift reflects a focus on promoting higher-level cognitive processing in multimedia learning environments. Material from previous chapters on learning in subject areas has been subsumed into other chapters.

The final part of the book (Part V), on advanced computer-based contexts, includes new chapters on multimedia learning with intelligent tutoring systems, games, simulations and microworlds, video, and multiple sources, reflecting a growing number of technology-based contexts that have been studied. Material from previous chapters on pedagogical agents, virtual reality, and hypermedia has been subsumed into these new chapters.

Identification of boundary conditions

An important development reflected in this second edition is the identification of *boundary conditions*; that is, there is now enough evidence in some cases to identify patterns in which a particular principle tends to apply under certain circumstances but not under others. For example, the modality principle (described in Chapters 9 and 12) tends to apply more strongly when the multimedia lesson is system-paced rather than learner-paced or when the verbal material is in short segments rather than long segments, and the pre-training principle (described in Chapter 13) tends to apply more strongly to low prior knowledge learners than to high prior knowledge learners. The discovery of a pattern of boundary conditions such as these provides a useful opportunity to test the predictions of theories of multimedia learning.

Refinement of theory

Part I on theoretical foundations contains updated versions of the same four theoretical chapters from the first edition – Chapter 2 on Sweller's cognitive load theory (which has received an evolutionary upgrade), Chapter 3 on Mayer's cognitive theory of multimedia learning (which adds a focus on the distinction among extraneous, essential, and generative forms of processing), Chapter 4 on Schnotz's integrated model of text and picture comprehension (which benefits from new data on how people build mental models from words and graphics), and Chapter 5 on van Merriënboer's four-component instructional design model (which includes reviews of supporting evidence for basic design principles).

The advances reflected in this second edition reflect a field of research that is flourishing in terms of increased empirical research base, theoretical depth, and practical application. However, you may detect that the job is not yet done. Some unfinished business in the study of multimedia learning includes incorporating metacognition and motivation into theories of multimedia learning, broadening the domain of study beyond multimedia presentations and lessons to include how multimedia principles apply to advanced media such as educational games or mobile learning environments, and expanding research venues beyond short-term lab studies to include more authentic learning in actual classrooms over longer time periods and with delayed tests.

Technology-Centered versus Learner-Centered Approaches to Multimedia Learning

Multimedia represents a potentially powerful learning technology – that is, a system for enhancing human learning. A practical goal of research on multimedia is to devise design principles for multimedia presentations. In addressing this goal, it is useful to distinguish between two approaches to multimedia design – a technology-centered approach and a learner-centered approach. The differences between the technology-centered and learner-centered approaches to multimedia design are summarized in Table 1.2.

Table 1.2. *Two approaches to the design of multimedia instruction*

Design approach	Starting point	Goal	Issues
Technology-centered	Capabilities of multimedia	Provide access to information	How can we use cutting-edge technology in designing multimedia instruction?
Learner-centered	How the human mind works	Aid human cognition	How can we adapt multimedia technology to aid human cognition?

Technology-centered approaches

The most straightforward approach to multimedia design is technology-centered. Technology-centered approaches begin with the functional capabilities of multimedia and ask, "How can we use these capabilities in designing multimedia presentations?" The focus is generally on cutting-edge advances in multimedia technology, so technology-centered designers might focus on how to incorporate multimedia into emerging communications technologies such as wireless access to the World Wide Web or the construction of interactive multimedia representations in virtual reality. The research issues often involve media research – that is, determining which technology is most effective in presenting information. For example, a media research issue is whether students learn as well from an online lecture – in which they can see a lecturer in a window on the computer screen – as from a live lecture – in which they are actually sitting in a classroom.

What's wrong with technology-centered approaches? A review of educational technologies of the 20th century shows that the technology-centered approach generally fails to lead to lasting improvements in education (Cuban, 1986). For example, when the motion picture was invented in the early 20th century, hopes were high that this visual technology would improve education. In 1922 the famous inventor Thomas Edison predicted that “the motion picture is destined to revolutionize our educational system and … in a few years it will supplant largely, if not entirely, the use of textbooks” (cited in Cuban, 1986, p. 9). Like current claims for the power of visual media, Edison proclaimed that “it is possible to teach every branch of human knowledge with the motion picture” (cited in Cuban, 1986, p. 11). In spite of the grand predictions, a review of educational technology reveals that “most teachers used films infrequently in their classrooms” (Cuban, 1986, p. 17). From our vantage point beyond the close of the 20th century, it is clear that the predicted educational revolution in which movies would replace books has failed to materialize.

Consider another disappointing example that may remind you of current claims for the educational potential of the World Wide Web. In 1932 Benjamin Darrow, founder of the Ohio School of the Air, proclaimed that radio could “bring the world to the classroom, to make universally available the services of the finest teachers, the inspiration of the greatest leaders …” (cited in Cuban, 1986, p. 19). His colleague, William Levenson, director of the Ohio School of the Air, predicted in 1945 that a “radio receiver will be as common in the classroom as the blackboard” and “radio instruction will be integrated into school life” (cited in Cuban, 1986, p. 19). As we rush to wire our schools and homes for access to the educational content of the Internet, it is humbling to recognize what happened to a similarly motivated movement for radio: “Radio has not been accepted as a full-fledged member of the educational community” (Cuban, 1986, p. 24).

In addition, consider the sad history of educational television – a technology that combined the visual power of the motion picture with the worldwide coverage of radio. By the 1950s, educational television was touted as a way to create a “continental classroom” that would provide access to “richer education at less cost” (Cuban, 1986, p. 33). Yet a review shows that teachers used television infrequently, if at all (Cuban, 1986).

Finally, consider the most widely acclaimed technological accomplishment of the 20th century – computers. The technology that supports computers is different from that of film, radio, and television, but the grand promises to revolutionize education are the same. Like current claims for the mind-enhancing power of computer technology, during the 1960s computer tutoring machines were predicted to eventually replace teachers. The first large-scale implementation occurred under the banner of computer-assisted instruction (CAI), in which computers presented short frames, solicited a response from the learner, and provided feedback to the learner. In spite of a large financial investment in CAI, sound evaluations showed that the two largest computer-based systems in the 1970s – PLATO and TICCIT – failed to produce better learning than traditional teacher-led instruction (Cognition and Technology Group at Vanderbilt, 1996).

What can we learn from the humbling history of the 20th century’s great educational technologies? Although different technologies underlie film, radio, television, and computer-assisted instruction, they all produced the same cycle. First, they began with grand promises about how the technology would revolutionize education. Second, there was an initial rush to implement the cutting-edge technology in schools. Third, from the perspective of a few decades later it became clear that the hopes and expectations had been largely unmet.

What went wrong with these technologies that seemed poised to tap the potential of visual and worldwide learning? I attribute the disappointing results to the technology-centered approach taken by the promoters. Instead of adapting technology to the needs of human learners, humans were forced to adapt to the demands of cutting-edge technologies. The driving force behind the implementations was the power of the technology rather than an interest in promoting human cognition. The focus was on giving people access to the latest technology rather than helping people to learn with the aid of technology.

Today, the most widely accepted cutting-edge technologies involve handheld devices such as smartphones, tablets, e-readers, and controllers. For example, school districts are told that the wave of the future requires purchasing one tablet for each student. Are we about to replicate the cycle of high expectations, large-scale implementation, and disappointing results in the realm of multimedia technology? In my opinion, the answer to that question depends on whether or not we continue to take a technology-centered approach. When we ask, “How can we give multimedia technology to students?” and when our goal is to “provide access to technology,” we are taking a technology-centered approach with a 100-year history of failure.

Learner-centered approaches

Learner-centered approaches offer an important alternative to technology-centered approaches. Learner-centered approaches begin with an understanding of how the human mind works and ask, “How can we adapt multimedia to enhance human learning?” The focus is on using multimedia technology as an aid to human cognition. Research questions focus on the relation between design features and the human information processing system, such as comparing multimedia designs that place light or heavy loads on the learner’s visual information processing channel. The premise underlying the learner-centered approach is that multimedia designs that are consistent with the way the human mind works are more effective in fostering learning than those that are not. This premise is the central theme of Part I, which lays out theories of multimedia learning.

The first successful development of multimedia learning technology was an instructional picture book for children entitled *Orbis Pictus* (The world in pictures), in which each page showed a drawing of a common scene such as a barbershop or birds in flight along with a printed name for each element in the drawing corresponding to a numbered key (Comenius, 1887). This multimedia book (i.e., using words and pictures) was first published in 1658 by John Comenius, and for more than a century it was the most popular textbook in Europe. Why was the world’s first multimedia textbook so successful? The answer lies in Comenius’s learner-centered approach based on the idea that words and things must go together because “there is nothing in understanding which was not before in the sense” (Comenius, 1887, p. xiv). In spite of the tremendous advances in multimedia technology we see today, the success of high-tech venues still depends on designing technologies that are in sync with how people learn.

Norman (1993, p. xi) eloquently makes the case for a learner-centered approach to technology design, which he refers to as human-centered technology: “Today we serve technology. We need to reverse the machine-centered point of view and turn it into a person-centered point of view: Technology should serve us.” Consistent with the learner-centered approach, Norman (1993, p. 3) shows how “technology can make us smart” – that is, technology can expand our cognitive capabilities. Norman (1993, p. 5) refers to tools that aid the mind as *cognitive artifacts*: “anything invented by humans for the purpose of improving thought or action counts as an artifact.” Examples include mental tools such as language and arithmetic, as well as physical tools such as paper and pencils; as the 20th century’s most important new cognitive artifact, computer technology represents a landmark invention that has the potential to assist human cognition in ways that were previously not possible.

Norman’s (1993, p. 9) assessment is that “much of science and technology takes a machine-centered view of the design of machines,” so that “the technology that is intended to aid human cognition … more often interferes and confuses.” In contrast, Norman’s (1993, p. 12) vision of a learner-centered approach to technology design is that “technology … should complement human abilities, aid those activities for which we are poorly suited, and enhance and help develop those for which we are ideally suited.” The design of multimedia technology to promote human cognition represents one exemplary component in the larger task of creating what Norman (1993) calls “things that make us smart.”

Three Metaphors of Multimedia Learning: Response Strengthening, Information Acquisition, and Knowledge Construction

In making decisions about how to design or select a multimedia learning environment, you may be influenced by your underlying conception of learning. Table 1.3 compares three views of multimedia learning – *multimedia learning as response strengthening*, *multimedia learning as information acquisition*, and *multimedia learning as knowledge construction*. If you view multimedia learning as response strengthening, then multimedia is a feedback delivery system. If you view multimedia learning as information acquisition, then multimedia is an information delivery system. If you view multimedia learning as knowledge construction, then multimedia is a cognitive aid.

Table 1.3. *Three metaphors of multimedia learning*

Metaphor	**Definition**	**Content**	**Learner**	**Teacher**	**Goal of multimedia**
Response strengthening	Strengthening and weakening connections	Connections	Passive receiver	Dispenser of rewards and punishments	Exercise system
Information acquisition	Adding information to memory	Information	Passive receiver	Dispenser of information	Delivery system
Knowledge construction	Building a coherent mental structure	Knowledge	Active sense maker	Cognitive guide	Cognitive guidance

Multimedia learning as response strengthening

According to the response strengthening view, learning involves increasing or decreasing the connection between a stimulus and a response. The underlying principle is that the connection is strengthened if a response is followed by reward and is weakened if the response is followed by punishment. This view entails assumptions about the nature of what is learned, the nature of the learner, the nature of the teacher, and the goals of multimedia presentations. First, learning is based on building connections, so "what is learned" is that a certain response is connected to a certain situation. Second, the learner's job is to make a response and receive feedback on the response; thus, the learner is a passive recipient of rewards and punishments. Third, the teacher's job – or, in some cases, the instructional designer's job – is to dispense rewards and punishments. Overall, the teacher controls the instructional episode by providing a prompt or question – such as "What is the definition of multimedia learning?" – and then providing feedback on the answer given by the learner – such as "Yes, that's correct" or "No, you left out _______." Finally, the goal of multimedia instruction is to provide practice in exercising skills, that is, to act as a trainer. The underlying metaphor is that multimedia is an exercise system, that is, a system for practicing skills with feedback.

The response strengthening view reflects the first major theory of learning proposed by educational psychologists in the early 1900s – the law of effect (Thorndike, 1913). According to Thorndike's law of effect, if a response is followed by a satisfying state of affairs it will be more likely to occur under the same circumstances, and if a response is followed by a unsatisfying state of affairs it will be less likely to occur under the same circumstances. This straightforward principle has been a pillar of psychology and education for more than 100 years (Mayer, 2001), dominating the field through the 1950s. The law of effect was the guiding principle for many early instructional programs delivered by teaching machines in the 1960s. This view of learning can still be seen in multimedia environments that emphasize drill and practice, such as an online game that teaches arithmetic computation by giving the learner points for each correctly answered arithmetic problem.

What is wrong with the response strengthening view (or more accurately, the response strengthening and weakening view)? My main objection is not that it is incorrect but rather that it is incomplete. Although certain cognitive skills (and motor skills, for that matter) can best be learned through drill and practice, the teaching of other kinds of knowledge – such as concepts and strategies – may best be taught with other methods of instruction based on other views of learning. For example, when the goal of instruction is to foster meaningful learning reflected in the ability to solve transfer problems, drill and practice aimed at response strengthening may be too limited. Thus, the response strengthening view may be appropriate for guiding the design of multimedia learning environments mainly when the goal of instruction is to help people learn certain specific skills. However, when the goal of instruction is to help people learn certain concepts and strategies that can be applied to new situations, the response strengthening view is not adequate.

Multimedia learning as information acquisition

According to the information acquisition view, learning involves adding information to one's memory. As with the other views, the information acquisition view entails assumptions about the nature of what is learned, the nature of the learner, the nature of the teacher, and the goals of multimedia presentations. First, learning is based on information – an objective item that can be moved from place to place (such as from the computer screen to the human mind). Second, the learner's job is to receive information; thus, the learner is a passive being who takes in information from the outside and stores it in memory. Third, the teacher's job – or the multimedia designer's job – is to present information. Fourth, the goal of multimedia presentations is to deliver information as efficiently as possible. The underlying metaphor is that of multimedia as a delivery system; according to this metaphor, multimedia is a vehicle for efficiently delivering information to the learner.

The information acquisition view is sometimes called the *empty vessel view* because the learner's mind is seen as an empty container that needs to be filled by the teacher pouring in some information. Similarly, this is sometimes called the *transmission view* because the teacher transmits information to be received by the learner. Finally, it is sometimes called the *commodity view* because information is seen as a commodity than can be moved from one place to another.

What's wrong with the information acquisition view? If your goal is to help people learn isolated fragments of information, then I suppose nothing is wrong with the information acquisition view. However, when your goal is to promote understanding of the presented material, the information acquisition view is not very helpful. Even worse, it conflicts with the research base on how people learn complex material (Mayer, 2009, 2011). When people are trying to understand presented material – such as a lesson on how a bicycle tire pump works – they do not carefully store each word like tape recorders. Rather, humans focus on the meaning of presented material and interpret it in light of their prior knowledge.

Multimedia learning as knowledge construction

According to the knowledge construction view, in contrast to the information acquisition view, multimedia learning is a sense-making activity in which the learner seeks to build a coherent mental representation from the presented material. Unlike information – which is an objective commodity that can be moved from one mind to another – knowledge is personally constructed by the learner and cannot be delivered in exact form from one mind to another. This is why two learners can be presented with the same multimedia message and come away with different learning outcomes. Second, according to the knowledge construction view, the learner's job is to make sense of the presented material; thus, the learner is an active sense maker who experiences a multimedia presentation and tries to integrate the presented material into a coherent mental representation. Third, the teacher's job is to assist the learner in this sense-making process; thus, the teacher is a cognitive guide who provides needed guidance to support the learner's cognitive processing. Fourth, the goal of multimedia presentations is not only to present information, but also to provide guidance for how to process the presented information – that is, for determining what to pay attention to, how to mentally organize it, and how to relate it to prior knowledge. Finally, the guiding metaphor is that of multimedia as a helpful communicator; according to this metaphor, multimedia is a sense-making guide – that is, an aid to knowledge construction.

Overall, I favor a knowledge construction view because it is more consistent with the research base on how people learn and because it is more consistent with my goal of promoting understanding of presented material. Rather than seeing the goal of multimedia presentations as exposing learners to vast quantities of information or exercising correct responses, my goal for multimedia is to help people develop an understanding of important aspects of the presented material. For example, Bransford, Brown, and Cocking (1999, p. xi) note that "in the last 30 years … views of how effective learning proceeds have shifted from the benefits of diligent drill and practice to focus on students' understanding and application of knowledge." In short, the knowledge construction view offers a more useful conception of learning when the goal is to help people understand and use what they have learned.

Three Kinds of Multimedia Learning Outcomes: No Learning, Rote Learning, and Meaningful Learning

There are two major kinds of goals of learning – remembering and understanding. Remembering is the ability to reproduce or recognize the presented material and is assessed by retention tests. The most common retention tests are recall – in which learners are asked to reproduce what was presented (such as writing down all they can remember of a lesson they read) – and recognition – in which learners are asked to select what was presented (as in a multiple-choice question) or judge whether a given item was presented (as in a true–false question). Thus, the major issue in retention tests involves quantity of learning – that is, how much was remembered.

Understanding is the ability to construct a coherent mental representation from the presented material; it is reflected in the ability to use the presented material in novel situations and is assessed by transfer tests. In a transfer test, learners must solve problems that were not explicitly given in the presented material – that is, they must apply what they learned to a new situation. An example is an essay question that asks learners to generate solutions to a problem, which requires going beyond the presented material. The major issue in transfer tests involves the quality of learning – that is, how well someone can use what he or she has learned. The distinction between remembering and understanding is summarized in Table 1.4. A major goal of the research presented in this handbook is to promote understanding as well as retention.

Table 1.4. ***Two goals of multimedia instruction***

Goal	Definition	Test	Sample test item
Remembering	Ability to reproduce or recognize presented material	Retention	Write down all you can remember from the presentation you just studied
Understanding	Ability to use presented material in novel situations	Transfer	List some ways to improve the reliability of the device you just read about

Table 1.5 summarizes three kinds of learning outcomes: no learning, rote learning, and meaningful learning. The distinguishing feature of no learning is poor performance on retention and transfer. In this case, the learner lacks knowledge. The distinguishing pattern for rote learning outcomes is good retention and poor transfer. In this case, the learner has what can be called *fragmented knowledge* or *inert knowledge* – knowledge that can be remembered but cannot be used in new situations. In short, the learner has acquired a collection of *factoids* – isolated bits of information. Finally, meaningful learning is distinguished by good transfer performance as well as good retention performance. In this case, the learner's knowledge is organized into an integrated representation. Overall, the chapters in this handbook examine design features of multimedia that foster meaningful learning – that is, ways of integrating words and pictures that foster meaningful learning.

Table 1.5. ***Three kinds of multimedia learning outcomes***

Learning outcome	Cognitive description	Test performance	

		Retention	Transfer
No learning	No knowledge	Poor	Poor
Rote learning	Fragmented knowledge	Good	Poor
Meaningful learning	Integrated knowledge	Good	Good

Two Kinds of Active Learning: Behavioral Activity versus Cognitive Activity

What is the best way to promote meaningful learning outcomes? The answer rests in *active learning* – meaningful learning outcomes occur as a result of the learner's activity during learning. However, does active learning refer to what is going on with the learner's physical behavior – such as the degree of hands-on activity – or to what is going on in the learner's mind – such as the degree of integrative cognitive processing? In short, if the goal is to foster meaningful learning outcomes, should multimedia presentations be designed to prime mainly behavioral activity or cognitive activity?

Consider the following situation. Alan is preparing for an upcoming test in meteorology. He sits in front of a computer and clicks on an interactive tutorial on lightning. The tutorial provides hands-on exercises in which he must fill in blanks by writing words. For example, the following sentence appears on the screen: "Each year approximately _____ Americans are killed by lightning." He types in an answer, and the computer then provides the correct answer. In this case, Alan is behaviorally active in that he is typing answers on the keyboard, but he may not be cognitively active in that he is not encouraged to make sense of the presented material.

In contrast, consider the case of Brian, who is preparing for the same upcoming meteorology test. Like Alan, he sits in front of a computer and clicks on a tutorial about lightning; however, Brian's tutorial is a short narrated animation explaining the steps in lightning formation. As he watches and listens, Brian tries to focus on the essential steps in lightning formation and to organize them into a cause-and-effect chain. Wherever the multimedia presentation is unclear about why one step leads to another, Brian uses his prior knowledge to help create an explanation for himself – which Chi, Bassok, Lewis, Reimann, and Glaser (1989) call a *self-explanation* (see also Chapter 17). For example, when the narration says that positively charged particles come to the surface of the earth, Brian mentally creates the explanation that opposite charges attract. In this scenario, Brian is behaviorally inactive because he simply sits in front of the computer; however, he is cognitively active because he is actively trying to make sense of the presentation.

Which type of active learning promotes meaningful learning? Research on learning shows that meaningful learning depends on the learner's cognitive activity during learning rather than on the learner's behavioral activity during learning. You might suppose that the best way to promote meaningful learning is through hands-on activity, such as a highly interactive multimedia program. However, behavioral activity per se does not guarantee cognitively active learning; it is possible to engage in hands-on activities that do not promote active cognitive processing – such as in the case of Alan or many highly interactive computer games. You might suppose that presenting material to a learner is not a good way to promote active learning because the learner appears to sit passively. In some situations, your intuitions would be right – presenting a long, incoherent, and boring lecture or textbook chapter is unlikely to foster meaningful learning. However, in other situations, such as the case of Brian, learners can achieve meaningful learning in a behaviorally inactive environment such as a multimedia instructional message. My point is that well-designed multimedia instructional messages can promote active cognitive processing in learners, even when learners seem to be behaviorally inactive.

Summary

In summary, this handbook explores how to promote multimedia learning – that is, learning from words and pictures. In 34 chapters, the book takes an evidence-based approach by examining what research has to say about how to design multimedia learning environments that help people learn. Overall, it examines the evidence for more than 20 principles of multimedia instructional design and explores their application in a variety of contexts ranging from computer-based presentations to educational games to tutoring systems. The approach taken here is learner-centered rather than technology-centered and seeks to foster meaningful learning rather than rote learning. Compared with the first edition, this second edition reflects a substantial growth in the research base, an increase in the number of evidence-based principles, a broadening of the domains of application, a better understanding of boundary conditions, and the refinement of learning theories.

Glossary

Boundary conditions:	Circumstances under which a design principle is most likely to apply and least likely to apply.
Information acquisition view:	Viewing learning as adding information to memory.
Knowledge construction view:	Viewing learning as building mental representations.
Learner-centered approach:	An approach to multimedia learning design based on adapting technology to the way people learn.
Meaningful learning:	Learning with understanding, as indicated by good performance on retention and transfer tests.
Multimedia:	Presenting words (such as printed text or spoken text) and pictures (such as illustrations, photos, animation, or video).
Multimedia instruction:	Presenting words and pictures that are intended to promote learning.
Multimedia learning:	Building mental representations from words and pictures.
Multimedia principle:	People learn more deeply from words and pictures than from words alone.
Response strengthening view:	Viewing learning as the strengthening and weakening of connections.
Rote learning:	Learning without understanding, as indicated by good performance on retention tests but not on transfer tests.
Technology-centered approach:	An approach to multimedia learning design based on making cutting-edge technology available to learners.

Acknowledgment

This chapter is partially based on “The Promise of Multimedia Learning” in Mayer (2009, ch. 1). Preparation of the chapter was supported by a grant from the Office of Naval Research. I appreciate the useful comments of Celeste Pilegard and Logan Fiorella.

---

[1] There may be some conditions in which words or pictures alone are better than words and pictures

combined, such as the redundancy effect described by Sweller and Kalyuga in Chapter 10 and the expertise reversal effect described by Kalyuga in Chapter 24.

References

Adesope, O. O., & Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments: A meta-analysis. *Journal of Educational Psychology*, 104, 250–263.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn*. Washington, DC: National Academy Press.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145–182.

Cognition and Technology Group at Vanderbilt (1996). Looking at technology in context: A framework for understanding technology in education. In D. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 807–840). New York: Macmillan.

Comenius, J. A. (1887). *Orbis pictus*. Syracuse, NY: Bardeen. [Reproduced version.]

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York: Teachers College Press.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, 15, 313–332.

Ginns, P. (2006). Integrating information: A meta-analysis of spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16, 511–525.

Ginns, P., Marin, A. J., & Marsh, H. M. (2013). Designing instructional text for conversational style: A meta-analysis. *Educational Psychology Review,* 25, 445–472.

Mayer, R. E. (2001). Changing conceptions of learning: A century of progress in the scientific study of education. In L. Corno (Ed.), *Education across a century: The centennial volume. One hundredth yearbook of the National Society for the Study of Education* (pp. 34–75). Chicago: University of Chicago Press.

Mayer, R. E. (2009). *Multimedia learning* (2d ed.). New York: Cambridge University Press.

Mayer, R. E. (2011). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83, 484–490.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84, 444–452.

Norman, D. A. (1993). *Things that make us smart*. Reading, MA: Addison-Wesley.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Paivio, A. (2006). *Mind and its evolution: A duel coding theoretical approach*. Mahwah, NJ: Lawrence Erlbaum.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction*, 13, 141–156.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Thorndike, E. L. (1913). *Educational psychology*. New York: Columbia University Press.

van Merriënboer, J. J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design*. New York: Lawrence Erlbaum.

Part I Theoretical Foundations

2 Implications of Cognitive Load Theory for Multimedia Learning

Fred Paas

Erasmus University Rotterdam

John Sweller

University of New South Wales

Abstract

Human cognitive architecture indicates the manner in which cognitive structures and processes are organized. In turn, that architecture can be used to hypothesize the relative effectiveness of alternative instructional designs. Over several decades, cognitive load theory has simultaneously identified those aspects of human cognition relevant to instructional issues and tested the resultant hypotheses using randomized, controlled experiments. The cognitive architecture used by cognitive load theory has continually been developed and refined over this period. Currently, that architecture is based on evolutionary principles. This chapter outlines the cognitive architecture used by cognitive load theory and provides a general indicator of its relevance to instructional design issues associated with multimedia instruction.

Introduction

Good instructional design is driven by our knowledge of human cognitive structures and the manner in which those structures are organized into a cognitive architecture. Without knowledge of relevant aspects of human cognitive architecture such as the characteristics of and intricate relations between working memory and long-term memory, the effectiveness of instructional design is likely to be random. Cognitive load theory (Sweller, 2011, 2012; Sweller, Ayres & Kalyuga, 2011; Sweller, van Merriënboer & Paas, 1998) has been one of the theories used to integrate our knowledge of human cognitive structures and instructional design principles. This chapter is concerned with the elements of that theory and its general implications for multimedia learning, specifically words presented in spoken or written form along with pictures or diagrams.

We suggest that the processes and structures of human cognition are closely analogous to the processes and structures associated with evolution by natural selection and that accordingly, evolutionary theory, which is much older and better developed than cognitive theory, can be used as a guide to assess which instructional procedures may or may not be effective. We begin by considering categories of knowledge from an evolutionary perspective, followed by an outline of those aspects of human cognitive architecture relevant to instructional concerns, also considered from an evolutionary perspective. Lastly, the instructional implications of human cognitive architecture, including categories of cognitive load, are discussed.

Categories of Knowledge

There are many, possibly an infinite number of ways in which knowledge can be categorized. From an instructional design perspective, only categories that require different instructional procedures are significant. Very few such categories have been identified. Geary's (2007, 2008, 2012) distinction between biologically primary and secondary knowledge, based on evolutionary theory, provides an example of a knowledge categorization scheme that is critically important from an instructional perspective.

Biologically primary knowledge is knowledge that we have specifically evolved to acquire over many generations. Examples are learning to recognize faces, learning to listen to and speak a first language, or learning to use general problem-solving strategies. An important characteristic of primary knowledge is that it is modular, in that we have probably evolved to acquire different primary skills during different evolutionary epochs and so various primary skills are likely to be independent. Because we have specifically evolved to acquire them, very complex primary skills can be acquired rapidly, easily, without mental effort and frequently unconsciously. Primary skills do not need explicit instruction. We do not need to be taught to listen to or speak our native language. Similarly, a general problem-solving strategy such as means–ends analysis (Newell & Simon, 1972) cannot be taught because we acquire knowledge of the strategy automatically without explicit instruction.

Biologically secondary knowledge is knowledge we have not evolved to acquire but that we need for cultural reasons. Schools and other educational institutions were invented to assist us to acquire biologically secondary knowledge, and so examples of this knowledge can be obtained by considering almost anything that is taught in educational institutions. Reading and writing, unlike listening and speaking, are examples of biologically secondary knowledge. While basic listening and speaking do not need to be taught in schools because we have evolved to acquire the skills merely by membership of a listening and speaking society, reading and writing require more than mere membership of a reading and writing society. Biologically secondary knowledge is acquired deliberately and with conscious effort. It needs to be explicitly taught.

While biologically primary knowledge and secondary knowledge are distinct, most knowledge we acquire consists of a combination of both. Frequently we use biologically primary knowledge to leverage the acquisition of secondary knowledge (Paas & Sweller, 2012). Table 2.1 indicates the important distinctions between biologically primary and secondary knowledge.

Table 2.1. ***Distinctions between biologically primary and secondary knowledge***

Biologically primary knowledge	Biologically secondary knowledge
Knowledge we have evolved to acquire	Cultural knowledge we have not evolved to acquire
Modular, with different types of knowledge unrelated to each other and acquired independently at different times and in different ways	Types of knowledge that bear some relation to each other and are acquired in a similar manner
Acquired easily, automatically and unconsciously	Acquired deliberately with conscious effort
Explicit instruction not required	Best acquired with explicit instruction

Human Cognitive Architecture

Cognitive load theory is concerned primarily with the acquisition of biologically secondary knowledge. The manner in which we process information associated with the acquisition of secondary knowledge is determined by human cognitive architecture. While we have evolved to acquire each example of biologically primary knowledge in a modular fashion independently of every other example of primary knowledge, we have evolved a general, cognitive architecture to process biologically secondary information (Sweller, 2003). The general principles that underlie the human cognitive architecture associated with secondary knowledge are identical to the principles that underlie the information processing characteristics of biological evolution. Both are examples of natural information processing systems (Sweller & Sweller, 2006). There are many ways of describing the information processing characteristics of human cognitive architecture and evolutionary biology, but cognitive load theory normally uses five basic principles.

The Information Store Principle

To function, natural information processing systems require a very large store of information. Long-term memory provides this role in human cognition, while a genetic code has the same role in biological evolution (Sweller & Sweller, 2006). Just as a genetic code heavily determines a biological life, so long-term memory heavily determines our cognitive lives. All the information in a genetic code has been determined by adaptation to an environment (evolution by natural selection), and similarly everything in long-term memory has been learned for the sake of cognitive adaptation to an environment. On this analogy between a genetic code and long-term memory, almost all human cognitive activity is determined by information held in long-term memory. This information must be learned over time just as the information held in a genetic code is acquired over time. Learning is defined as an alteration in long-term memory. If nothing has altered in long-term memory nothing has been learned. Accordingly, appropriate alterations in long-term memory's store of biologically secondary information should be the primary aim of instruction.

The suggestion that information in long-term memory is analogous to a genetic code, that most human cognitive activity is driven by information held in long-term memory and that the aim of instruction should be to alter long-term memory implies that the long-term memory's store of biologically secondary information is very large. The evidence for a very large long-term memory is overwhelming. The origin of this discovery is unusual: the game of chess.

De Groot (1965) studied the factors that permitted chess grand masters to almost invariably defeat less able players. The only factor he could find that distinguished between more able and less able chess players was memory for board configurations taken from real games. If shown a board configuration taken from a real game for a few seconds and then asked to reproduce that configuration, chess grand masters could replace most of the pieces correctly. Less able players could correctly replace few of the pieces. Chase and Simon (1973) replicated this result but found it could not be replicated using random board configurations. The result was obtainable only when board configurations taken from real games were used.

In the late 1970s and 1980s, a similar result was obtained many times in a variety of fields by several investigators (e.g., Egan & Schwartz, 1979; Jeffries, Turner, Polson & Atwood, 1981; Sweller & Cooper, 1985). Experts have a vastly superior memory to novices for problem states in their field of expertise. For example, Simon and Gilmartin (1973) have estimated that chess grand masters have memorized up to 100,000 board configurations. It is this store of biologically secondary information in long-term memory that constitutes expertise. As a consequence, problem-solving skill is critically determined by information in long-term memory concerning problem states and the best move associated with each state. Such knowledge held in long-term memory allows an expert to immediately recognize most of the situations faced and the actions required by that situation. That large body of biologically secondary knowledge permits the fluency shown by experts in their own area. A major function of instructional design is to assist learners to acquire fluency. Fluent procedures imply that the necessary biologically secondary knowledge that underpins skilled performance in any substantive area has been acquired.

The Borrowing and Reorganizing Principle

How do large amounts of information accumulate in an information store? In the case of genetic material, it is obtained from ancestors through either asexual or sexual reproduction. In this manner, all genetic information with the exception of mutations is borrowed from other stores. Furthermore, during sexual reproduction, it is reorganized with information from two parents combined.

In the case of human cognition, information also is borrowed from other stores. The borrowing and reorganizing principle assumes that we have evolved to acquire information from other people. We imitate what other people do (Bandura, 1986), listen to what they say and read what they write. By engaging in imitating, listening and reading, we can obtain new, biologically secondary knowledge from other people that we combine with existing knowledge, resulting in the alteration of the contents of long-term memory.

It should be noted that our ability to obtain biologically secondary information from other people via imitation and listening is itself a biologically primary task that does not need to be taught (Paas & Sweller, 2012). We have evolved to obtain information in this manner. We have not evolved to obtain information from others via reading, and so this task needs to be taught as a biologically secondary task, along with associated tasks such as learning to use a library or the Internet. The fact that listening is biologically primary while reading is biologically secondary may contribute to the modality effect (Chapter 8). While looking at a diagram, for example, we may be better at listening to associated speech rather than reading.

The huge stores of biologically secondary knowledge that we hold in long-term memory could not be acquired rapidly and efficiently without the borrowing and reorganizing principle. Cognitive load theory is based, at least in part, on the assumption that the purpose of instruction is to assist learners to acquire information from other people. Therefore, the way we present that information and the activities that we ask learners to engage in are important and constitute much of the subject matter of cognitive load theory.

The Randomness as Genesis Principle

While effective learning in substantive fields depends on obtaining biologically secondary information from others via the borrowing and reorganizing principle, that information must be created in the first instance before it can be transferred. The manner in which humans create biologically secondary information is again analogous to the manner in which evolution by natural selection creates information.

A species faced with a changing environment may evolve to handle the new circumstances. The manner in which it evolves is not predetermined. All variations between species and between individual members of species ultimately can be sourced to random mutations. In effect, whenever a mutation occurs, it is checked for effectiveness, with effective mutations resulting in more offspring and ineffective mutations resulting in fewer or even no offspring. In this manner, evolution by natural selection uses a random generate and test process. From an information processing perspective, this procedure is indistinguishable from human problem solving, which also depends on random generation followed by tests of effectiveness. The underlying logic of both systems is identical.

Consider a student learning a new task incorporating biologically secondary information such as how to navigate the Web. The student is faced with a screen page containing many buttons, each likely to represent a link to other pages and functions that also contain many more links and functions. He or she has to learn which buttons on the screen to press in order to successfully navigate. It is a new task and so the student has no knowledge informing him or her of the procedures to be followed. Assuming there is no one present to provide direct guidance, the student must engage in problem solving to determine an appropriate procedure. Failing knowledge (either one's own or someone else's knowledge), problem-solving search can function only by the learner randomly proposing a step and then testing that step for effectiveness. That random component when the learner is dealing with novel material that necessitates problem solving is quite unavoidable. If information is not available, the student must discover the new procedures required using a random generate and test procedure that is analogous to the random generate and test procedures required by a species faced with a new environment.

Random generate and test procedures provide another example of a biologically primary task that does not need to be taught but can be used in biologically secondary tasks (Paas & Sweller, 2012). We have evolved to learn how to use random generate and test and so do not need to be taught how to engage in it, although we may need to have the usefulness of the procedure pointed out to us for particular secondary tasks. Random generation has further structural implications for human cognitive architecture .

The Narrow Limits of Change Principle

Consider an information processing system that is severely limited in that it can combine only about 4 novel elements at any given time. There are many ways those elements could be combined, but let us assume they are being combined using the logic of permutations. With four elements, there are 4! = 24 permutations. It may be difficult to determine which of 24 permutations is best but it is likely to be possible. In contrast, assume a somewhat larger information processing system that can handle 10 rather than 4 elements. With 10 elements, there are 10! = 3,628,800 permutations. An information processing system structured to test the relative effectiveness of millions of new possibilities is likely to be unworkable. As a consequence, and paradoxically, a somewhat smaller system is likely to be more efficient than a larger one. Working memory provides the human cognitive system with the required characteristics.

When dealing with novel, biologically secondary information, human working memory has two severe limitations. Miller (1956) indicated that working memory is able to hold only about 7 elements of information. It can probably process in the sense of combine, contrast or manipulate no more than about 2–4 novel elements. On these numbers, the capacity of working memory when dealing with new, biologically secondary information is severely constrained. The duration of working memory is also constrained. Peterson and Peterson (1959) found that, without rehearsal, almost all the contents of working memory are lost within about 20 seconds. We may have evolved with these limitations in our ability to acquire biologically secondary knowledge because a larger, or worse, unlimited working memory may be counterproductive due to the unmanageable number of combinations of elements that could be generated. Furthermore, one of the functions of working memory is to determine which novel information should be used to alter the information held in long-term memory. A large, rapid change in the biologically secondary information held in long-term memory is likely to render that store dysfunctional. Small, incremental changes are less likely to have adverse consequences.

There is an analogous structure to working memory in evolutionary biology. The epigenetic system (Jablonka & Lamb, 2005; West-Eberhard, 2003) plays the same role in evolution by natural selection as working memory plays in human cognition (Sweller & Sweller, 2006). Just as working memory determines which external information will be processed, so the epigenetic system can speed up or slow down the rate of genetic mutations. It can also determine where mutations will occur in the same way that working memory determines which problems will be considered and which problem-solving steps will be taken. As is the case for the human cognitive system, changes to the information store (a genome) must be small to ensure that its functionality is not destroyed.

The Environmental Organizing and Linking Principle

The environmental organizing and linking principle provides the ultimate justification for a natural information processing system. It connects the information held in the information store with appropriate action in the external environment. Working memory and the epigenetic system again are central. When they take information from the information store in order to act within a given environment, their characteristics are very different to when they deal with novel information from the external environment.

Human cognitive architecture has evolved with an ingenious set of relations between long-term and working memory. The nature of those relations provides the centrepiece of human cognitive functioning and is critical to any theory of instructional design. The intellectual heights that humans have reached and to which they aspire are made possible by the manner in which biologically secondary information in long-term memory alters the characteristics of working memory. The environmental organizing and linking principle provides the necessary context for relations between long-term memory, working memory and the activity that is appropriate for a given environment.

The limitations of working memory were discussed previously. It must be emphasized that those limitations apply only to novel information fed to working memory through the sensory system (known as sensory memory). Information that has already been organized in long-term memory can also be fed into working memory. Neither the duration nor capacity limitations attached to novel information that is received from sensory memory apply to information from long-term memory. That information has no measurable limitations of either duration or capacity. It can be indefinite in size and duration. In effect, information in long-term memory vastly expands working memory. That expansion trivializes any biological differences between humans in the capacity of working memory. Basic differences between people in working memory capacity are likely to be irrelevant given the huge alterations in this processor that occur when it is dealing with organized information taken from long-term memory.

Ericsson and Kintsch (1995) with their concept of long-term working memory provided an important model of the relations between working and long-term memory. They suggested that because the characteristics of working memory when processing information from long-term memory are so dramatically different to its characteristics when processing information from sensory memory, it is appropriate to assume a separate processor – long-term working memory.

We can consider the relations between working and long-term memory in the following manner. At one end of a continuum, when one is dealing with unfamiliar information, working memory limitations are critical. They become successively less critical as familiarity increases, that is, as more and more information from long-term memory is used. At the other extreme, when one is dealing with information incorporated in well-entrenched knowledge, working memory limitations become irrelevant. Thus, the extent to which working memory limitations matter depends on the extent to which the information being dealt with has been organized in long-term memory. The characteristics of working memory and the manner in which working memory functions are critically dependent on what has been stored in long-term memory.

Relations between working and long-term memory mirror similar relations between the epigenetic and genetic systems. The epigenetic system can selectively use large amounts of genetic information to transform activity in the same way that working memory can selectively use large amounts of information from long-term memory in order to determine activity. A skin cell and a liver cell from a particular individual are structurally and functionally vastly different, despite having identical genetic information in their nuclei. The differences between them are due to the epigenetic rather than the genetic system.

We can summarize natural information processing systems in the following manner. In the case of both evolution and human cognition, large amounts of information can be dealt with only after they have been appropriately organized. Prior to being organized, the amount of information that can be dealt with is necessarily very small. In the case of genetic information, huge amounts of organized information can be dealt with by the epigenetic system and transmitted from generation to generation, but alterations to a genome are not and cannot be organized. Random alterations followed by effectiveness testing are unavoidable, and so any viable alterations will be relatively minuscule. Similarly, a huge amount of organized, biologically secondary information held in long-term memory can be and is used repeatedly, but failing direct guidance through instruction, changes to long-term memory cannot be organized. Random generation followed by effectiveness testing must be used, and this procedure cannot and should not result in rapid, massive, effective changes to long-term memory. Alterations must be small, and a small working memory when one is dealing with new information is a consequence. Table 2.2 indicates the function of each of the five principles.

Table 2.2. ***The function of each of the natural information store principles***

Principle	**Function**
Information store	Storing of information

principle	
Borrowing and reorganizing principle	Acquisition of information
Randomness as genesis principle	Generation of novel information
Narrow limits of change principle	Imposition of limits to the generation of novel information to ensure the continuing functionality of the information store
Environmental organizing and linking principle	Coordination of stored information with the external environment to generate appropriate action

The Structure of Knowledge in Long-Term Memory

Emphasizing the importance of accumulating biologically secondary knowledge in long-term memory as the primary goal of instruction is sometimes misinterpreted as an emphasis on rote learning. In fact, both rote learning and learning with understanding result in changes in long-term memory. Rote learning occurs when some connections between elements occur but other, essential connections are omitted. If a student learns to recite the letters of the alphabet but not how they can be used to produce written language or learns to recite a multiplication table but not that multiplication is a shorthand procedure for repeated addition, there are changes in long-term memory due to the rote-learned material. If the student begins to learn to read or learns to use multiplication instead of repeated addition to determine the cost of three pencils, as well as changes in long-term memory due to rote learning there are further changes due to the increased level of understanding. Understanding can be largely described by the additional changes in long-term memory. Without changes in long-term memory, nothing has been understood.

The environmental organizing and linking principle with its emphasis on relations between working memory and long-term memory is central to an explanation of understanding (Marcus, Cooper & Sweller, 1996). Understanding occurs when all relevant elements of biologically secondary information can be processed simultaneously in working memory. Because of the limitation of working memory when dealing with novel, biologically secondary information, if faced with new material that must be learned, there may be too many elements to simultaneously process in working memory. If the elements are essential, understanding cannot occur until it becomes possible to process them. While the learner is studying the material, elements are organized and combined into knowledge held in long-term memory. When knowledge acquisition has progressed to the point where all of the elements essential to understanding a topic can be processed in working memory, understanding has occurred. On the basis of these interactions, understanding can be defined as the ability to simultaneously process required elements in working memory. On this definition, the relations and interplay between working and long-term memory are central to understanding.

We can get an intuitive feel for the power of information held in long-term memory by considering the cognitive processes required to read this page. Objectively, written text is an almost indescribably complex series of squiggles. A person can read because knowledge of individual letters permits an infinite number of shapes to be recognized (hence the ability to read handwriting); knowledge of combinations of letters that form words and combinations of words to form phrases permits extremely complex combinations of squiggles to be recognized. Further, additional knowledge connects these squiggles to objects, events and procedures, permitting meaning to be derived. This knowledge is acquired over very long periods of time and is all stored in long-term memory. In character and function, there is every reason to believe that knowledge for reading is identical in function to the knowledge acquired by chess grand masters for chessboard configurations. All skilled performance in complex domains requires the acquisition of large amounts of knowledge held in long-term memory. From a multimedia perspective, knowledge is held in long-term memory whether it is pictorial or verbal, written or spoken.

Instructional Consequences: Cognitive Load Theory

Cognitive load theory (Paas, Renkl, & Sweller, 2003; Sweller, 2011, 2012; Sweller et al., 1998; Sweller et al., 2011; van Merriënboer & Sweller, 2005) and the instructional principles it has generated are based on the assumptions discussed above concerning human cognitive architecture, especially the assumptions concerning working memory and long-term memory. Three categories of cognitive load are included the theory: intrinsic, extraneous and germane cognitive load. All categories of cognitive load are concerned with the acquisition, storage and use of biologically secondary information.

Intrinsic cognitive load is the cognitive load due to the natural complexity of the biologically secondary information that must be processed. It is determined by levels of element interactivity (Sweller, 2010). For example, if someone is learning to translate some of the nouns of a foreign language, each translation can be learned independently of every other translation. One can learn to translate the word ‘cat’ without learning to translate the word ‘dog’. In this example, element interactivity is low and so working memory load is low. In contrast, the elements that constitute other material may interact in the sense that one cannot meaningfully learn one element without simultaneously learning many other elements. For example, if learning the appropriate word order in English for the words ‘when learning a language’, one cannot attend to individual words to determine that ‘a language learning when’ is inappropriate. One must consider all of the words and the relations between them because they interact. Element interactivity is high, resulting in a high intrinsic cognitive load. This is a biologically secondary task in the case of a foreign language but a primary task in the case of a native language. While there are other reasons why learning can be difficult, such as the material including a very large number of elements irrespective of whether they interact, understanding and learning material with high element interactivity are difficult for a specific and important reason: because material with high element interactivity imposes a high working memory load.

For a given task with a specific knowledge level, intrinsic cognitive load is fixed. It can be changed but only by changing the nature of the task or the knowledge levels of participants. Multiple elements can be combined into a single element during learning, resulting in a reduction of intrinsic cognitive load. Readers of this book can treat ‘cat’ as a single element. When one is learning to read, ‘cat’ will consist of many interacting elements.

Extraneous cognitive load also is caused by high levels of element interactivity, but in this case the element interactivity is due to inappropriate instructional designs that unnecessarily increase the number of interacting elements that learners must process (Sweller, 2010). Inappropriate instructional designs require learners to use working memory resources to process elements that do not lead to knowledge acquisition.

There is a wide range of instructional design effects that are based on cognitive load theory. Each effect takes a commonly used instructional procedure, analyses it from the perspective of relevant aspects of human cognition and then redesigns the instruction to reduce working memory load and increase knowledge acquisition. Several of the effects based on cognitive load theory are discussed in this volume (e.g., Chapters 8, 9, and 10) and so will not be discussed further here.

Lastly, germane cognitive load is ‘effective’ cognitive load. It refers to working memory resources that are devoted to dealing with intrinsic cognitive load rather than extraneous cognitive load. The more working memory resources that are devoted to dealing with interacting elements associated with intrinsic cognitive load and the fewer working memory resources that are devoted to dealing with interacting elements associated with extraneous cognitive load, the more effective will be instruction. In effect, germane cognitive load refers to the working memory resources devoted to intrinsic cognitive load minus the resources devoted to extraneous cognitive load. The higher the value, the higher the germane cognitive load. In this sense, germane cognitive load is a synthesis or combination of intrinsic and extraneous cognitive load.

Extraneous cognitive load and intrinsic cognitive load are additive. The aim of instruction should be to reduce extraneous cognitive load caused by inappropriate instructional procedures. Reducing extraneous cognitive load frees working memory capacity and so may permit an increase in the working resources devoted to intrinsic cognitive load, resulting in an increase in germane cognitive load. Nevertheless, if intrinsic cognitive load is low, working memory resources devoted to intrinsic cognitive load may be adequate for learning even with high levels of extraneous cognitive load. Germane load may not need to be high if intrinsic cognitive load is low. In other words, how one designs instruction may not be particularly important when dealing with simple material that can be easily understood. Even with poor instructional designs, working memory capacity may not be exceeded. Instructional design may be critical only when one is dealing with complex material that imposes a heavy working memory load due to its intrinsic nature. When a heavy extraneous cognitive load is added to a heavy intrinsic cognitive load, working memory capacity may be exceeded, whereas when a heavy extraneous cognitive load is added to a light intrinsic cognitive load, capacity may not be exceeded. As a consequence, the cognitive load effects due to extraneous cognitive load can be demonstrated only with material that is high in element interactivity (Sweller & Chandler, 1994; Tindall-Ford, Chandler & Sweller, 1997). If element interactivity due to intrinsic cognitive load is low, material can frequently be understood and learned even if element interactivity due to extraneous cognitive load is high. Table 2.3 summarizes each category of cognitive load.

Table 2.3. *Categories of cognitive load*

Category	Source	Example
Intrinsic	Caused by interacting elements that are intrinsic to the task and must be processed simultaneously. Cannot be altered other than by changing the nature of the task or by increasing knowledge.	A novice solving a mathematical problem, such as $a/b = c$, solve for a. Because the elements interact, no problem-solving move can be made without all of the other elements being affected.
Extraneous	Caused by interacting elements introduced by an instructional design. This cognitive load should be reduced by altering the instructional design.	Requiring learners to learn by solving a problem rather than studying a worked example. Searching for a problem solution unnecessarily introduces a large number of interacting elements that are eliminated by the study of a worked example.
Germane	Refers to working memory resources dealing with intrinsic rather than extraneous cognitive load, thus facilitating learning.	Instructional designs that decrease extraneous load associated with problem-solving search increase working memory resources devoted to intrinsic rather than extraneous elements and so increase germane load.

Conclusions

Instructional design that proceeds without reference to human cognition is likely to be random in its effectiveness. Until relatively recently, that lamentable state of affairs was unavoidable because our knowledge of human cognitive architecture was too sparse to effectively apply to instruction. The immense expansion of that knowledge, including suggestions concerning the evolutionary origins of human cognitive architecture, has altered the instructional design landscape. The limitations of working memory when one is dealing with novel, biologically secondary information and the elimination of those limitations when one is dealing with well-known information have profound implications for instructional design in general and multimedia instruction in particular. Those implications have changed and are likely to continue to change instructional procedures.

Glossary

Biologically primary knowledge:	Knowledge that we have evolved to acquire.
Biologically secondary knowledge:	Knowledge that we need to acquire for cultural reasons but that we have not specifically evolved to acquire.
Borrowing and reorganizing principle:	Explains how natural information processing systems can acquire very large information stores.
Cognitive architecture:	The manner in which the cognitive structures used to learn, think and solve problems are organized.
Cognitive load:	The load imposed on working memory by information being presented.
Cognitive load theory:	An instructional theory based on our knowledge of human cognitive architecture which specifically addresses the limitations of working memory.
Direct instructional guidance:	Instruction in which procedures are directly demonstrated to learners. Can be contrasted with inquiry-based learning.
Element interactivity:	The extent to which elements of information that must be processed interact. If material that must be learned has high element interactivity, elements cannot be processed individually in working memory, and that material will then be seen as complex and difficult to understand and learn.
Environmental organizing and linking principle:	Explains how natural information processing systems link information held in the information store to appropriate action in the external environment.
Extraneous cognitive load:	The cognitive load that is imposed by nonessential, interacting elements (see element interactivity) that can be eliminated by altering the instructional design.
Germane cognitive load:	Working memory resources devoted to dealing with intrinsic rather than extraneous interacting elements.
Information store principle:	Deals with the need for natural information processing systems to store very large amounts of information; long-term memory in humans.
Inquiry-based learning:	Instruction in which learners, rather than having a procedure demonstrated, are required to discover it themselves. Can be contrasted with direct instructional guidance.
Intrinsic cognitive load:	The cognitive load that is imposed by essential, interacting elements (*see* element interactivity) that, because they interact, must be processed simultaneously rather than successively in working memory, resulting in a heavy load.
Learning:	Any change in long-term memory involving an accumulation of information.

Long-term memory:	The cognitive structure that stores our knowledge base. We are conscious only of those contents of long-term memory that are transferred to working memory.
Narrow limits of change principle:	Explains why natural information processing systems can make only small changes to their information stores. In humans, working memory when dealing with novel information.
Natural information processing systems:	Information processing systems that can be found in nature, such as biological evolution and human cognition.
Randomness as genesis principle:	Explains how natural information processing systems generate novel information.
Sensory memory:	The cognitive structure that permits us to perceive new information.
Working memory:	The cognitive structure in which we consciously process information. Notable for its severe capacity and duration limits when dealing with new information.

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

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

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

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

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

Geary, D. (2007). Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology. In J. S. Carlson & J. R. Levin (Eds.), *Psychological perspectives on contemporary educational issues* (pp. 1–99). Greenwich, CT: Information Age Publishing.

Geary, D. (2008). An evolutionarily informed education science. *Educational Psychologist*, 43, 179–195.

Geary, D. (2012). Evolutionary educational psychology. In K. Harris, S. Graham & T. Urdan (Eds.), *APA educational psychology handbook* (vol. 1, pp. 597–621). Washington, DC: American Psychological Association.

Jablonka, E., & Lamb, M. J. (2005). *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. Cambridge, MA: MIT Press.

Jeffries, R., Turner, A., Polson, P., & Atwood, M. (1981). Processes involved in designing software. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 255–283). Hillsdale, NJ: Lawrence Erlbaum.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, 88, 49–63.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.

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

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38, 1–4.

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review*, 24, 27–45.

Peterson, L., & Peterson, M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193–198.

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

Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning and motivation* (vol. 43, pp. 215–266). San Diego, CA: Academic Press.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review*, 22, 123–138.

Sweller, J. (2011). Cognitive load theory. In J. Mestre & B. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (vol. 55, pp. 37–76). Oxford: Academic Press.

Sweller, J. (2012). Human cognitive architecture: Why some instructional procedures work and others do not. In K. Harris, S. Graham & T. Urdan (Eds.), *APA educational psychology handbook* (vol. 1, pp. 295–325). Washington, DC: American Psychological Association.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12, 185–233.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2, 59–89.

Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, 4, 434–458.

Sweller, J., van Merriënboer, J. J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–296.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, 3, 257–287.

van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17, 147–177.

West-Eberhard, M. (2003). *Developmental plasticity and evolution*. New York: Oxford University Press.

3 Cognitive Theory of Multimedia Learning

Richard E. Mayer

University of California, Santa Barbara

Abstract

A fundamental hypothesis underlying research on multimedia learning is that multimedia instructional messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not so designed. The cognitive theory of multimedia learning is based on three cognitive science principles of learning: the human information processing system includes dual channels for visual/pictorial and auditory/verbal processing (i.e., dual-channel assumption), each channel has a limited capacity for processing (i.e., limited-capacity assumption), and active learning entails carrying out a coordinated set of cognitive processes during learning (i.e., active processing assumption). The cognitive theory of multimedia learning specifies five cognitive processes in multimedia learning: selecting relevant words from the presented text or narration, selecting relevant images from the presented graphics, organizing the selected words into a coherent verbal representation, organizing selected images into a coherent pictorial representation, and integrating the pictorial and verbal representations and prior knowledge. Three demands on the learner's cognitive capacity during learning are extraneous processing (which is not related to the instructional objective), essential processing (which is needed to mentally represent the essential material as presented), and generative processing (which is aimed at making sense of the material). Three instructional goals are to reduce extraneous processing (for extraneous overload situations), manage essential processing (for essential overload situations), and foster generative processing (for generative underuse situations). Multimedia instructional messages should be designed to guide appropriate cognitive processing during learning without overloading the learner's cognitive system.

The Case for Multimedia Learning

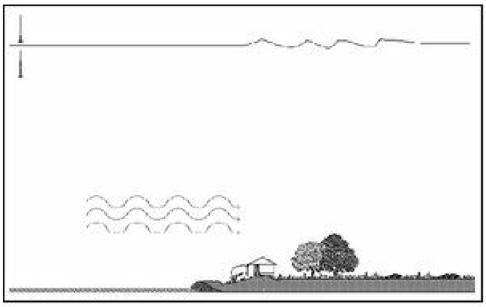
What Is the Rationale for a Theory of Multimedia Learning?

People learn more deeply from words and pictures than from words alone. This assertion – which can be called the *multimedia principle* – underlies much of the interest in multimedia learning. For thousands of years, words have been the major format for instruction – including spoken words and, within the past few hundred years, printed words. Today, thanks to advances in computer and communication technologies, pictorial forms of instruction are becoming widely available, including dazzling computer-based graphics. However, simply adding pictures to words does not guarantee an improvement in learning – that is, all multimedia presentations are not equally effective. In this chapter I explore a theory aimed at understanding how to use words and pictures to improve human learning.

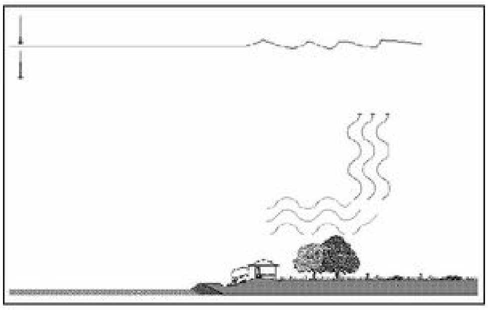
A fundamental hypothesis underlying research on multimedia learning is that multimedia instructional messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not so designed. For the past 25 years my colleagues and I at the University of California, Santa Barbara (UCSB) have been engaged in a sustained effort to construct an evidenced-based theory of multimedia learning that can guide the design of effective multimedia instructional messages (Mayer 2001, 2008, 2009; Mayer & Moreno, 2003).

What Is a Multimedia Instructional Message?

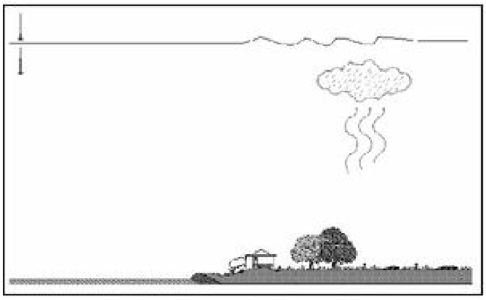
A multimedia instructional message is a communication containing words and pictures intended to foster learning. The communication can be delivered using any medium, including paper (i.e., book-based communications) and computers (i.e., computer-based communications), or even face to face (i.e., face-to-face communications). Words can include printed words (such as you are now reading) or spoken words (such as in a narration); pictures can include static graphics – such as illustrations, charts, and photos – or dynamic graphics – such as animation and video clips. This definition is broad enough to include textbook chapters containing text and illustrations, online lessons containing animation and narration, interactive simulation games including on-screen text and graphics, and face-to-face slideshow presentations involving graphics and spoken words. For example, Figure 3.1 presents frames from a narrated animation on lightning formation, which we have studied in numerous experiments (Mayer, 2009).



“Cool moist air moves over a warmer surface and becomes heated.”



“Warmed moist air near the earth’s surface rises rapidly.”



“As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud.”

Figure 3.1. Selected frames from a narrated animation on lightning formation.

Learning can be measured by tests of retention (i.e., remembering the presented information) and transfer (i.e., being able to use the information to solve new problems), as described in Chapter 1. Our focus is on transfer test performance because we are mainly interested in how words and pictures can be used to promote understanding. In short, transfer tests can help tell us how well people understand what they have learned. We are particularly interested in the cognitive processes by which people construct meaningful learning outcomes from words and pictures.

What Is the Role of a Theory of Learning in Multimedia Design?

Much of the work presented in this handbook is based on the premise that the design of multimedia instructional messages should be compatible with how people learn. In short, the design of multimedia instructional messages should be sensitive to what we know about how people process information. The cognitive theory of multimedia learning represents an attempt to accomplish this goal by describing how people learn from words and pictures, in a way that is consistent with empirical research evidence (e.g., Mayer, 2001, 2008, 2009; Mayer & Moreno, 2003) and consensus principles in cognitive science (e.g., Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Mayer, 2011).

In building the cognitive theory of multimedia learning, my colleagues and I were guided by four criteria: *theoretical plausibility* – the theory is consistent with cognitive science principles of learning; *testability* – the theory yields predictions that can be tested in scientific research; *empirical plausibility* – the theory is consistent with empirical research evidence on multimedia learning; and *applicability* – the theory is relevant to educational needs for improving the design of multimedia instructional messages. In this chapter, I describe the cognitive theory of multimedia learning, which is intended to meet these criteria. In particular, I summarize three underlying assumptions of the theory derived from cognitive science; describe three memory stores, five cognitive processes, and five forms of representation in the theory; examine three demands on the learner's cognitive capacity during learning and three resulting goals for coping with them; and then provide a historical overview and a conclusion.

Three Assumptions of a Cognitive Theory of Multimedia Learning

Decisions about how to design a multimedia message always reflect an underlying conception of how people learn – even when the underlying theory of learning is not stated (Mayer, 1992). In short, the design of multimedia messages is influenced by the designer's conception of how the human mind works. For example, when a multimedia presentation consists of a screen overflowing with multicolored words and images – flashing and moving about – this reflects the designer's conception of human learning. The designer's underlying conception is that human learners possess a single-channel, unlimited-capacity, and passive processing system. First, by not taking advantage of auditory modes of presentation, this design is based on a single-channel assumption – all information enters the cognitive system in the same way regardless of its modality. It follows that it does not matter which modality is used to present information – such as presenting words as sounds or text – just as long as the information is presented. Second, by presenting so much information, this design is based on an unlimited-capacity assumption – humans can handle an unlimited amount of material. It follows that the designer's job is to present information to the learner. Third, by presenting many isolated pieces of information, this design is based on a passive processing assumption – humans act as if they were tape recorders, adding as much information to their memories as possible. It follows that learners do not need any guidance in organizing and making sense of the presented information.

What's wrong with this vision of learners as possessing a single-channel, unlimited-capacity, passive processing system? Current research in cognitive psychology paints a quite different picture of how the human mind works (Mayer, 2009, 2011). Thus, a difficulty with this commonsense conception of learning is that it conflicts with what is known about how people learn. In this section, I explore three assumptions underlying a cognitive theory of multimedia learning – *dual channels*, *limited capacity*, and *active processing*. These assumptions are summarized in Table 3.1.

Table 3.1. *Three assumptions of a cognitive theory of multimedia learning*

Assumption	Description	Related citations
Dual channels	Humans possess separate channels for processing visual and auditory information	Paivio (1986), Baddeley (1992)
Limited capacity	Humans are limited in the amount of information that can be processed in each channel at one time	Baddeley (1992), Chandler and Sweller (1991)
Active processing	Humans engage in active learning by attending to relevant incoming information, organizing selected information into coherent mental representations, and integrating mental representations with other knowledge	Mayer (1999), Wittrock (1989)

Dual-Channel Assumption

The dual-channel assumption is that humans possess separate information processing channels for visually/spatially represented material and auditorily/verbally represented material. The relevance of the dual-channel assumption to the cognitive theory of multimedia learning lies in the proposal that the human information processing system contains an auditory/verbal channel and a visual/pictorial channel. When information is presented to the eyes (such as illustrations, animations, video, or on-screen text), humans begin by processing that information in the visual channel; when information is presented to the ears (such as narration or nonverbal sounds), humans begin by processing that information in the auditory channel. The concept of separate information processing channels has a long history in cognitive psychology and currently is most closely associated with Paivio's dual-coding theory (Clark & Paivio, 1991; Paivio, 1986, 2006) and Baddeley's model of working memory (Baddeley, 1999; Baddeley, Eysenck, & Anderson, 2009).

What is processed in each channel? There are two ways of conceptualizing the differences between the two channels – one based on *representation modes* and the other based on *sensory modalities*. The representation-mode approach focuses on whether the presented stimulus is verbal (such as spoken or printed words) or nonverbal (such as pictures, video, animation, or background sounds). According to the representation-mode approach, one channel processes verbal material and the other channel processes pictorial material and nonverbal sounds. This conceptualization is most consistent with Paivio's (1986, 2006) distinction between verbal and nonverbal systems.

In contrast, the sensory-modality approach focuses on whether learners initially process the presented materials through their eyes (such as for pictures, video, animation, or printed words) or ears (such as for spoken words or background sounds). According to the sensory-modality approach, one channel processes visually represented material and the other channel processes auditorily represented material. This conceptualization is most consistent with Baddeley's (1999; Baddeley, Eysenck, & Anderson, 2009) distinction between the visuospatial sketchpad and the phonological loop.

Whereas the representation-mode approach focuses on the format of the stimulus (i.e., verbal or nonverbal), the sensory-modality approach focuses on the sensory modality of the stimulus (i.e., auditory or visual). The major difference concerning multimedia learning rests in the processing of printed words (i.e., on-screen text) and background sounds. On-screen text is initially processed in the verbal channel in the representation-mode approach but in the visual channel in the sensory-modality approach; background sounds, including nonverbal music, are initially processed in the nonverbal channel in the representation-mode approach but in the auditory channel in the sensory-modality approach.

For purposes of the cognitive theory of multimedia learning, I have opted for a compromise in which I use the sensory-modality approach to distinguish between visually presented material (such as pictures, animations, video, and on-screen text) and auditorily presented material (such as narration and background sounds), as well as a representation-mode approach to distinguish between the construction of pictorially based and verbally based models in working memory. However, additional research is necessary to clarify the nature of the differences between the two channels and the implications for learning and instruction.

What is the relation between the channels? Although information enters the human information system via one channel, learners may be able to convert the representation for processing in the other channel. When learners are able to devote adequate cognitive resources to the task, it is possible for information originally presented to one channel to also be represented in the other channel. For example, on-screen text may initially be processed in the visual channel because it is presented to the eyes, but an experienced reader may be able to mentally convert images into sounds, which are processed through the auditory channel. Similarly, an illustration of an object or event such as a cloud rising above the freezing level may initially be processed in the visual channel, but the learner may also be able to mentally construct the corresponding verbal description in the auditory channel. Conversely, a narration describing some event such as "the cloud rises above the freezing level" may initially be processed in the auditory channel because it is presented to the ears, but the learner may also form a corresponding mental image that is processed in the visual channel. Cross-channel representations of the same stimulus play an important role in Paivio's (1986, 2006) dual-coding theory.

Limited-Capacity Assumption

The second assumption is that humans are limited in the amount of information that can be processed in each channel at one time. When an illustration or animation is presented, the learner is able to hold only a few images in the visual channel of working memory at any one time, reflecting portions of the presented material rather than an exact copy of the presented material. For example, if an illustration or animation of a tire pump is presented, the learner may be able to focus on building mental images of the handle going down, the inlet valve opening, and air moving into the cylinder. When a narration is presented, the learner is able to hold only a few words in the verbal channel of working memory at any one time, reflecting portions of the presented text rather than a verbatim recording. For example, if the spoken text is "When the handle is pushed down, the piston moves down, the inlet valve opens, the outlet valve closes, and air enters the bottom of cylinder," the learner may be able to hold the following verbal representations in auditory working memory: "handle goes up," "inlet valve opens," and "air enters cylinder." The conception of limited capacity in consciousness has a long history in psychology, and some modern examples are Baddeley's (1999; Baddeley, Eysenck, & Anderson, 2009; see also Chapter 25) theory of working memory and Sweller's (1999; Sweller, Ayres, & Kalyuga, 2011; see also Chapter 2) cognitive load theory.

What are the limits on cognitive capacity? If we assume that each channel has limited processing capacity, it is important to know just how much information can be processed in each channel. The classic way to measure someone's cognitive capacity is to give a memory span test (Miller, 1956; see also Mayer, 2011), although more recent advancements include the OSpan and RSpan tests, as described in Chapter 25. Although there are individual differences, on average, memory span is fairly small – approximately five to seven chunks.

With practice, of course, people can learn techniques for chunking the elements in the list, such as grouping the seven digits 8–7–5–3–9–6–4 into three chunks, 875–39–64 (e.g., "eight seven five" pause "three nine" pause "six four"). In this way, the cognitive capacity remains the same – five to seven chunks – but more elements can be remembered within each chunk (Mayer, 2011).

How are limited cognitive resources allocated? The constraints on our processing capacity force us to make decisions about which pieces of incoming information to pay attention to, the degree to which we should build connections among the selected pieces of information, and the degree to which we should build connections between selected pieces of information and our existing knowledge.

Metacognitive strategies are techniques for allocating, monitoring, coordinating, and adjusting these limited cognitive resources. These strategies are at the heart of what Baddeley (1999; Baddeley, Eysenck, & Anderson, 2009) calls the *central executive* – the system that controls the allocation of cognitive resources – and play a central role in modern theories of metacognition (Hacker, Dunlosky, & Graesser, 2009).

Active Processing Assumption

The third assumption is that humans actively engage in cognitive processing in order to construct a coherent mental representation of their experiences. These active cognitive processes include paying attention to relevant incoming information, organizing incoming information into a coherent cognitive structure, and integrating incoming information with other knowledge. In short, humans are active processors who seek to make sense of multimedia presentations. This view of humans as active processors conflicts with a common view of humans as passive processors who seek to add as much information as possible to memory, that is, as if they were tape recorders filing copies of their experiences in memory to be retrieved later.

What are the major ways that knowledge can be structured? Active learning occurs when a learner applies cognitive processes to incoming material – processes that are intended to help the learner make sense of the material. The desired outcome of active cognitive processing is the construction of a coherent mental representation, so active learning can be viewed as a process of model building. A *mental model* (*or knowledge structure*) represents the key parts of the presented material and their relations. For example, in a multimedia presentation of how lightning storms develop, the learner may attempt to build a cause-and-effect system in which a change in one part of the system causes a change in another part. In a lesson comparing and contrasting two theories, construction of a mental model involves building a sort of matrix structure that compares the two theories along several dimensions.

If the outcome of active learning is the construction of a coherent mental representation, it is useful to explore some of the typical ways that knowledge can be structured. Some basic knowledge structures include *process*, *comparison*, *generalization*, *enumeration*, and *classification* (Chambliss & Calfee, 1998; Cook & Mayer, 1988). Process structures can be represented as cause-and-effect chains and consist of explanations of how some system works. An example is an explanation of how the human ear works. Comparison structures can be represented as matrices and consist of comparisons among two or more elements along several dimensions. An example is a comparison between how two competing theories of learning view the role of the learner, the role of the teacher, and useful types of instructional methods. Generalization structures can be represented as a branching tree and consist of a main idea with subordinate supporting details. An example is an essay in support of lowering the voting age. Enumeration structures can be represented as lists and consist of a collection of items. An example is the names of principles of multimedia learning listed in this handbook. Classification structures can be represented as hierarchies and consist of sets and subsets. An example is a biological classification system for sea animals.

Understanding a multimedia message often involves constructing one or more of these kinds of knowledge structures. This assumption suggests two important implications for multimedia design: (1) the presented material should have a coherent structure, and (2) the message should provide guidance to the learner on how to build the structure. If the material lacks an underlying coherent structure – for example, if the material is mainly a collection of isolated facts – the learner's model-building efforts will be fruitless. If the message lacks guidance on how to structure the presented material, the learner's model-building efforts may be overwhelmed. Multimedia design can be conceptualized as an attempt to assist learners in their model-building efforts.

What are the cognitive processes involved in active learning? Table 3.2 summarizes three cognitive processes that are essential for active learning: selecting relevant material, organizing selected material, and integrating selected material with existing knowledge (Mayer, 2009; Wittrock, 1989). Selecting relevant material occurs when a learner pays attention to appropriate words and images in the presented material. This process involves bringing material from the outside into the working memory component of the cognitive system. Organizing selected material involves building structural relations among the elements – such as one of the five kinds of structures described in the preceding section. This process takes place within the working memory component of the cognitive system. Integrating selected material with existing knowledge involves building connections between incoming material and relevant portions of prior knowledge. This process involves activating knowledge in long-term memory and bringing it into working memory. For example, in a multimedia message on the cause of lightning, learners must pay attention to certain words and images, arrange them into a cause-and-effect chain, and relate the steps to prior knowledge such as the principle that hot air rises.

Table 3.2. ***Three cognitive processes required for active learning***

Process	**Description**
Selecting	Attending to relevant material in the presented lesson for transfer to working memory
Organizing	Mentally organizing selected information into a coherent cognitive structure in working memory
Integrating	Connecting cognitive structures with each other and with relevant prior knowledge activated from long-term memory

In sum, the implicit theory of learning underlying some multimedia messages is that learning is a single-channel, unlimited-capacity, passive processing activity. In contrast, I offer a cognitive theory of multimedia learning that is based on three basic assumptions about how the human mind works – namely, that the human mind is a dual-channel, limited-capacity, active processing system.

Three Memory Stores in the Cognitive Theory of Multimedia Learning

Figure 3.2 presents a cognitive model of multimedia learning intended to represent the human information processing system. The boxes represent memory stores, including sensory memory, working memory, and long-term memory, and the arrows represent the cognitive processes of selecting, organizing, and integrating. The top row represents the verbal channel and the bottom row represents the visual channel.

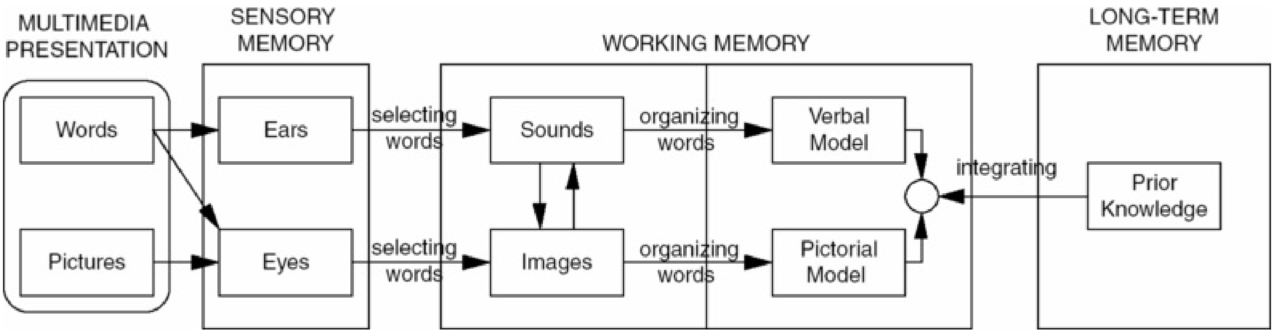


Figure 3.2. Cognitive theory of multimedia learning.

Table 3.3 summarizes the characteristics of the three memory stores in the cognitive theory of multimedia learning. Pictures and words come in from the outside world as a multimedia presentation (indicated on the left side of the figure) and enter sensory memory through the eyes and ears (indicated in the "Sensory Memory" box). Sensory memory allows for pictures and printed text to be held as exact visual images for a very brief time period in a visual sensory memory (at the top) and for spoken words and other sounds to be held as exact auditory images for a very brief time period in an auditory sensory memory (at the bottom). The arrow from "Pictures" to "Eyes" corresponds to a picture being registered in the visual sensory memory, the arrow from "Words" to "Ears" corresponds to spoken text being registered in the auditory sensory memory, and the arrow from "Words" to "Eyes" corresponds to printed text being registered in the visual sensory memory.

Table 3.3. ***Three memory stores in the cognitive theory of multimedia learning***

Memory store	**Description**	**Capacity**	**Duration**	**Format**
Sensory memory	Briefly holds sensory copies of incoming words and pictures	Unlimited	Very brief	Visual or auditory sensory images
Working memory	Allows for manipulating selected incoming information	Limited	Short	Verbal and pictorial representations
Long-term memory	Permanently stores organized knowledge	Unlimited	Permanent	Knowledge

The central work of multimedia learning takes place in working memory, so let's focus on the "Working Memory" box in Figure 3.2. Working memory is used for temporally holding and manipulating knowledge in active consciousness. For example, in reading this sentence you may be able to actively concentrate on only some of the words at one time, or in looking at Figure 3.2 you may be able to hold the images of only some of the boxes and arrows in your mind at one time. This kind of processing – namely, processing that involves conscious attention – takes place in working memory. The left side of the "Working Memory" box represents the raw material that comes into working memory – visual images of pictures and sound images of words – so it is based on the two sensory modalities that I call visual and auditory; in contrast, the right side of the "Working Memory" box represents the knowledge constructed in working memory – pictorial and verbal models and links between them – so it is based on the two representation modes that I call pictorial and verbal. I use the term *pictorial model* to refer to spatial representations rather than visual images. The arrow from "Sounds" to "Images" represents the mental conversion of a sound (such as the spoken word "cat") into a visual image (such as an image of a cat) – that is, when you hear the word "cat" you might also form a mental image of a cat; the arrow from "Images" to "Sounds" represents the mental conversion of a visual image (such as a mental picture of a cat) into a sound (such as the sound of the word "cat") – that is, you mentally hear the word "cat" when you see a picture of one.

Finally, the box on the right is labeled "Long-Term Memory" and corresponds to the learner's storehouse of knowledge. Unlike working memory, long-term memory can hold large amounts of knowledge over long periods of time, but to actively think about material in long-term memory it must be brought into working memory (as indicated by the arrow from "Long-Term Memory" to "Working Memory").

The major cognitive processing required for multimedia learning is represented by the arrows in Figure 3.2 labeled "selecting images," "selecting words," "organizing images," "organizing words," and "integrating" – which are described in the next section.

Five Processes in the Cognitive Theory of Multimedia Learning

For meaningful learning to occur in a multimedia environment, the learner must engage in five cognitive processes, indicated by the arrows in Figure 3.2: (1) selecting relevant words for processing in verbal working memory, (2) selecting relevant images for processing in visual working memory, (3) organizing selected words into a verbal model, (4) organizing selected images into a pictorial model, and (5) integrating the verbal and pictorial representations with each other and with relevant prior knowledge activated from long-term memory. The five cognitive processes in multimedia learning are summarized in Table 3.4. Although I present these processes as a list, they do not necessarily occur in linear order, so a learner might move from process to process in many different ways. Successful multimedia learning requires that the learner coordinate and monitor these five processes.

Table 3.4. *Five cognitive processes in the cognitive theory of multimedia learning*

Process	Description
Selecting words	Learner pays attention to relevant words in a multimedia message to create sounds in working memory
Selecting images	Learner pays attention to relevant pictures in a multimedia message to create images in working memory
Organizing words	Learner builds connections among selected words to create a coherent verbal model in working memory
Organizing images	Learner builds connections among selected images to create a coherent pictorial model in working memory
Integrating	Learner builds connections between verbal and pictorial models and with prior knowledge

Selecting Relevant Words

The first labeled step in Figure 3.2 involves a change in knowledge representation from the external presentation of spoken words (such as a computer-generated narration) to a sensory representation of sounds to an internal working memory representation of word sounds (such as some of the words in the narration). The input for this step is a spoken verbal message – that is, the spoken words in the presented portion of the multimedia message. The output for this step is a word sound base (called *sounds* in Figure 3.2) – that is, a mental representation in the learner's verbal working memory of selected words or phrases.

The cognitive process mediating this change is called *selecting relevant words* and involves paying attention to some of the words that are presented in the multimedia message as they pass through auditory sensory memory. If the words are presented as speech, this process begins in the auditory channel (as indicated by the arrows from "Words" to "Ears" to "Sounds"). However, if the words are presented as on-screen text or printed text, this process begins in the visual channel (as indicated by the arrow from "Words" to "Eyes") and later may move to the auditory channel if the learner mentally articulates the printed words (as indicated by the arrow from "Images" to "Sounds" in the left side of working memory). The need for selecting only part of the presented message occurs because of capacity limitations in each channel of the cognitive system. If the capacity were unlimited, there would be no need to focus attention on only part of the verbal message. Finally, the selection of words is not arbitrary; the learner must determine which words are most relevant – an activity that is consistent with the view of the learner as an active sense maker.

For example, in the lightning lesson partially shown in Figure 3.1, one segment of the multimedia presentation contains the words “Cool moist air moves over a warmer surface and becomes heated,” the next segment contains the words “Warmed moist air near the earth’s surface rises rapidly,” and the next segment has the words “As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud.” When a learner engages in the selection process, the result may be that some of the words are represented in verbal working memory – such as, “Cool air becomes heated, rises, forms a cloud.”

Selecting Relevant Images

The second step involves a change in knowledge representation from the external presentation of pictures (such as an animation segment or an illustration) to a sensory representation of unanalyzed visual images to an internal representation in working memory (such as a visual image of part of the animation or illustration). The input for this step is a pictorial portion of a multimedia message that is held briefly in visual sensory memory. The output for this step is a visual image base – a mental representation in the learner’s working memory of selected images.

The cognitive process underlying this change – *selecting relevant images* – involves paying attention to part of the animation or illustrations presented in the multimedia message. This process begins in the visual channel, but it is possible to convert part of it to the auditory channel (such as by mentally narrating an ongoing animation). The need to select only part of the presented pictorial material arises from the limited processing capacity of the cognitive system. It is not possible to process all parts of a complex illustration or animation simultaneously, so learners must focus on only part of the incoming pictorial material at a time. Finally, the selection process for images – like the selection process for words – is not arbitrary because the learner must judge which images are most relevant for making sense of the multimedia presentation.

In the lightning lesson, for example, one segment of the animation shows blue colored arrows – representing cool air – moving over a heated land surface that contains a house and trees; another segment shows the arrows turning red and traveling upward above a tree; and a third segment shows the arrows changing into a cloud with lots of dots inside. In selecting relevant images, the learner may compress all this into images of a blue arrow pointing rightward, a red arrow pointing upward, and a cloud; details such as the house and tree on the surface, the wavy form of the arrows, and the dots in the cloud are lost.

Organizing Selected Words

Once the learner has formed a word sound base from the incoming words of a segment of the multimedia message, the next step is to organize the words into a coherent representation – a knowledge structure that I call a *verbal model*. The input for this step is the word sound base – the word sounds selected from the incoming verbal message – and the output for this step is a verbal model – a coherent (or structured) representation in the learner’s working memory of the selected words or phrases.

The cognitive process involved in this change is *organizing selected words*, in which the learner builds connections among pieces of verbal knowledge. This process is most likely to occur in the auditory channel and is subject to the same capacity limitations that affect the selection process. Learners do not have unlimited capacity to build all possible connections so they must focus on building a simple structure. The organizing process is not arbitrary, but rather reflects an effort at sense making – such as the construction of a cause-and-effect chain.

For example, in the lightning lesson partially shown in Figure 3.1, the learner may build causal connections between the selected verbal components: “First: cool air is heated; second: it rises; third: it forms a cloud.” In mentally building a causal chain, the learner is organizing the selected words.

Organizing Selected Images

The process for organizing images parallels that for selecting words. Once the learner has formed an image base from the incoming pictures of a segment of the multimedia message, the next step is to organize the images into a coherent representation – a knowledge structure that I call a *pictorial model*. The input for this step is the visual image base – the images selected from the incoming pictorial message – and the output for this step is a pictorial model – a structured spatial representation in the learner's working memory based on the selected images.

This change from images to pictorial model requires the application of a cognitive process that I call *organizing selected images*. In this process, the learner builds connections among pieces of pictorial knowledge. This process occurs in the visual channel, which is subject to the same capacity limitations that affect the selection process. Learners lack the capacity to build all possible connections among images in their working memory, but rather must focus on building a simple set of connections. As in the process of organizing words, the process of organizing images is not arbitrary. Rather, it reflects an effort to build a simple structure that makes sense to the learner – such as a cause-and-effect chain.

For example, in the lightning lesson, the learner may build causal connections between the selected images: the rightward-moving blue arrow turns into a rising red arrow, which turns into a cloud. In short, the learner builds causal links in which the first event leads to the second and so on.

Integrating Word-Based and Picture-Based Representations

Perhaps the most crucial step in multimedia learning involves making connections between word-based and picture-based representations. This step involves a change from having two separate representations – a verbal model and a pictorial model – to having an integrated representation in which corresponding elements and relations from one model are mapped onto the other. The input for this step is the pictorial model and the verbal model that the learner has constructed so far, and the output is an integrated model, which is based on connecting the two representations. In addition, the integrated model includes connections with relevant prior knowledge.

I refer to this cognitive process as *integrating words and images* because it involves building connections between corresponding portions of the pictorial and verbal models as well as with relevant knowledge from long-term memory. This process occurs in visual and verbal working memory and involves the coordination between them. This is an extremely demanding process that requires the efficient use of cognitive capacity. The process reflects the epitome of sense making because the learner must focus on the underlying structure of the visual and verbal representations. The learner also can use prior knowledge activated from long-term memory to help coordinate the integration process, as indicated by the arrow from long-term memory to working memory.

For example, in the lightning lesson, the learner must see the connection between the verbal chain – "First, cool air is heated; second, it rises; third, it forms a cloud" – and the pictorial chain – the blue arrow followed by the red arrow followed by the cloud shape. In addition, the learner can apply prior knowledge to the transition from the first to the second event by remembering that hot air rises.

Each of the five processes in multimedia learning is likely to occur many times throughout a multimedia presentation. The processes are applied segment by segment rather than to the message as a whole. For example, in processing the lightning lesson, learners do not first select all relevant words and images from the entire passage, then organize them into verbal and pictorial models of the entire passage, and then connect the completed models with one another at the very end. Rather, learners carry out this procedure on small segments: they select relevant words and images from the first sentence of the narration and the first few seconds of the animation; they organize and integrate them; and then this set of processes is repeated for the next segment, and so on. Schnotz and Bannert's (2003; see also Chapter 4) integrated model of text and picture comprehension also addresses the issue of how learners integrate words and pictures.

Finally, another process (not shown in Figure 3.2 or Table 3.4) is *encoding*, which involves an arrow from working memory to long-term memory, signifying the transfer of the constructed representation from working memory to long-term memory for permanent storage within the learner's organized knowledge base.

Five Forms of Representation

As you can see in Figure 3.2, there are five forms of representation for words and pictures, reflecting their stage of processing. To the far left, we begin with *words and pictures in the multimedia presentation* – that is, the stimuli that are presented to the learner. In the case of the lightning message shown in Figure 3.1, the words are the spoken words presented through the computer's speakers and the pictures are the frames of the animation presented on the computer screen. Second, as the presented words and pictures impinge on the learner's ears and eyes, the next form of representation is *acoustic representations (or sounds) and visual representations (or images) in sensory memory*. The sensory representations fade rapidly, unless the learner pays attention to them. Third, when the learner selects some of the words and images for further processing in working memory, the next form of representation is *sounds and images in working memory*. These are the building blocks of knowledge construction – including key phrases such as "warmed air rises" and key images such as red arrows moving upward. The fourth form of representation results from the learner's construction of a *verbal model and pictorial model in working memory*. Here the learner has organized the material into coherent verbal and spatial representations, and also has mentally integrated them. The pictorial model should be considered a schematic spatial representation rather than a sensory-like visual image. Finally, the fifth form of representation is *knowledge in long-term memory*, which the learner uses for guiding the process of knowledge construction in working memory. Sweller (1999) refers to this knowledge as *schemas*. As new knowledge is constructed in working memory, it may be stored in long-term memory as prior knowledge to be used in supporting new learning. The five forms of representation are summarized in Table 3.5.

Table 3.5. ***Five forms of representation in the cognitive theory of multimedia learning***

Type of knowledge	Location	Example
Words and pictures	Multimedia presentation	Sound waves from computer speaker: "Cool moist air …"; pixel patterns on the computer screen showing a wavy blue arrow
Acoustic and iconic representations	Sensory memory	Received sounds in learner's ears: "Cool moist air …"; received image in learner's eyes corresponding to wavy blue arrow
Sounds and images	Working memory	Selected sounds: "Cool moist air moves"; selected images: wavy blue line moving rightward
Verbal and pictorial models	Working memory	Mental model of cloud formation
Prior knowledge	Long-term memory	Schema of differences in air pressure

Three Kinds of Demands on Cognitive Capacity

The challenge for instructional design is to guide the learner's appropriate cognitive processing during learning without overloading the learner's working memory capacity. Table 3.6 summarizes three kinds of demands on the learner's information processing system during learning: extraneous processing, essential processing, and generative processing.

Table 3.6. *Three demands on cognitive capacity during multimedia learning*

Name	Description	Caused by	Learning processes	Example
Extraneous processing	Cognitive processing that is not related to the instructional goal	Poor instructional design	None	Focusing on irrelevant pictures
Essential processing	Cognitive processing to represent the essential presented material in working memory	Complexity of the material	Selecting	Memorizing the description of essential processing
Generative processing	Cognitive processing aimed at making sense of the material	Motivation to learn	Organizing and integrating	Explaining generative processing in one's own words

Extraneous processing refers to cognitive processing that does not support the instructional goal and is caused by poor instructional design. For example, when a figure is printed on one page and the words describing the figure are printed on another page, a learner may have to scan back and forth, resulting in extraneous processing that wastes precious cognitive capacity. Extraneous processing does not result in any useful knowledge being constructed in the learner's working memory. Extraneous processing is analogous to *extraneous cognitive load* in cognitive load theory, as described in Chapter 2.

Essential processing refers to cognitive processing aimed at mentally representing the presented material in working memory and is caused by the complexity of the material. For example, less essential processing is required to mentally represent the definition of working memory than is required to mentally represent the information processing system summarized in Figure 3.2. Essential processing involves selecting relevant information from the presentation and organizing it as presented. Thus, essential processing results in the construction of verbal and pictorial representations in working memory that correspond to the presented material, analogous to a *textbase* in Kintsch's (1998) construction-integration theory of text processing. Essential processing is analogous to *intrinsic cognitive load* in cognitive load theory, as described in Chapter 2.

Generative processing refers to cognitive processing aimed at making sense of the presented material and is caused by the learner's motivation to learn. For example, when the material is presented by a likable instructor, the learner may exert more effort to understand what the instructor is presenting. Generative processing involves reorganizing the incoming information and integrating it with relevant prior knowledge. Thus, generative processing results in the construction of an integrated mental model, analogous to a *situation model* in Kintsch's (1998) construction-integration theory of text processing. Generative processing is analogous to *germane cognitive load* in cognitive load theory, as described in Chapter 2. Both generative and essential processes are directed at the instructional goal.

Each of the key concepts – cognitive capacity, extraneous processing, essential processing, and generative processing – is relative to the learner and the learner's interaction with the instructional situation. For example, learners differ in terms of their working memory capacity (as explored in Chapter 25), which affects their ability to handle each of the three kinds of demands on cognitive capacity. Learners differ in their cognitive and metacognitive strategies for engaging in generative processing and essential processing. They differ in terms of their prior knowledge that can help them handle the extraneous processing caused by poorly designed instructional situations or guide their essential and generative processing of familiar material. For example, individual differences in prior knowledge are an important consideration in the instructional design of multimedia instruction (see Chapter 24). Thus, the identical multimedia lesson may be overloading for one learner and not be overloading for another because of differences in the capacities, knowledge, skills, and beliefs (e.g., beliefs about how learning works) that learners bring to the learning situation.

The learner has a limited amount of cognitive capacity to process information in each channel in working memory during learning, so capacity that is used for extraneous processing cannot be used for essential and generative processing. In short, consistent with cognitive load theory (Plass, Moreno, & Brunken, 2010; Sweller, Ayres, & Kalyuga, 2011; also see Chapter 2), the sum of extraneous processing plus essential processing plus generative processing cannot exceed the learner's cognitive capacity. Given that the learner's cognitive capacity is limited and the three demands on cognitive capacity are additive, if the learner increases one kind of processing then another one must be decreased. The instructional implications of this triarchic model of cognitive processing demands are explored in the next section.

Three Learning Scenarios

Figure 3.3 summarizes three learning scenarios based on the triarchic model of cognitive processing demands. First, in the top frame, consider what happens when the instructional message is so poorly designed that the learner is forced to expend large amounts of processing capacity on extraneous processing, thereby leaving insufficient capacity for essential and generative processing. This scenario, which can be called *extraneous overload*, can be addressed by devising instructional methods aimed at reducing extraneous processing. Examples of techniques aimed at reducing extraneous processing include the coherence principle, signaling principle, redundancy principle, spatial contiguity principle (or split-attention principle), and temporal contiguity principle, as described in Chapters 8, 10, 11, and 13. The goal of these instructional techniques, which are summarized in Table 3.7, is to reduce extraneous processing so that available cognitive capacity can be used for essential and generative processing.

Table 3.7. ***Three instructional goals in multimedia learning***

Goal	Representative technique	Description of technique	Chapter
Minimize extraneous processing	Coherence principle	Eliminate extraneous material	12
	Signaling principle	Highlight essential material	11, 12
	Redundancy principle	Do not add printed text to spoken text	10, 12
	Spatial contiguity principle	Place printed text near corresponding graphic	8, 12
	Temporal contiguity principle	Present narration and corresponding graphic simultaneously	12

	Segmenting principle	Break presentation into parts	13
Manage essential processing	Pre-training principle	Describe names and characteristics of key elements before the lesson	13
	Modality principle	Use spoken rather than printed text	9, 13
	Multimedia principle	Use words and pictures rather than words alone	7
Foster generative processing	Personalization principle	Put words in conversational style	14
	Voice principle	Use human voice for spoken words	14
	Embodiment principle	Give on-screen characters humanlike gestures	14
	Guided discovery principle	Provide hints and feedback as learner solves problems	15
	Self-explanation principle	Ask learners to explain a lesson to themselves	17
	Drawing principle	Ask learners to make drawings for the lesson	18

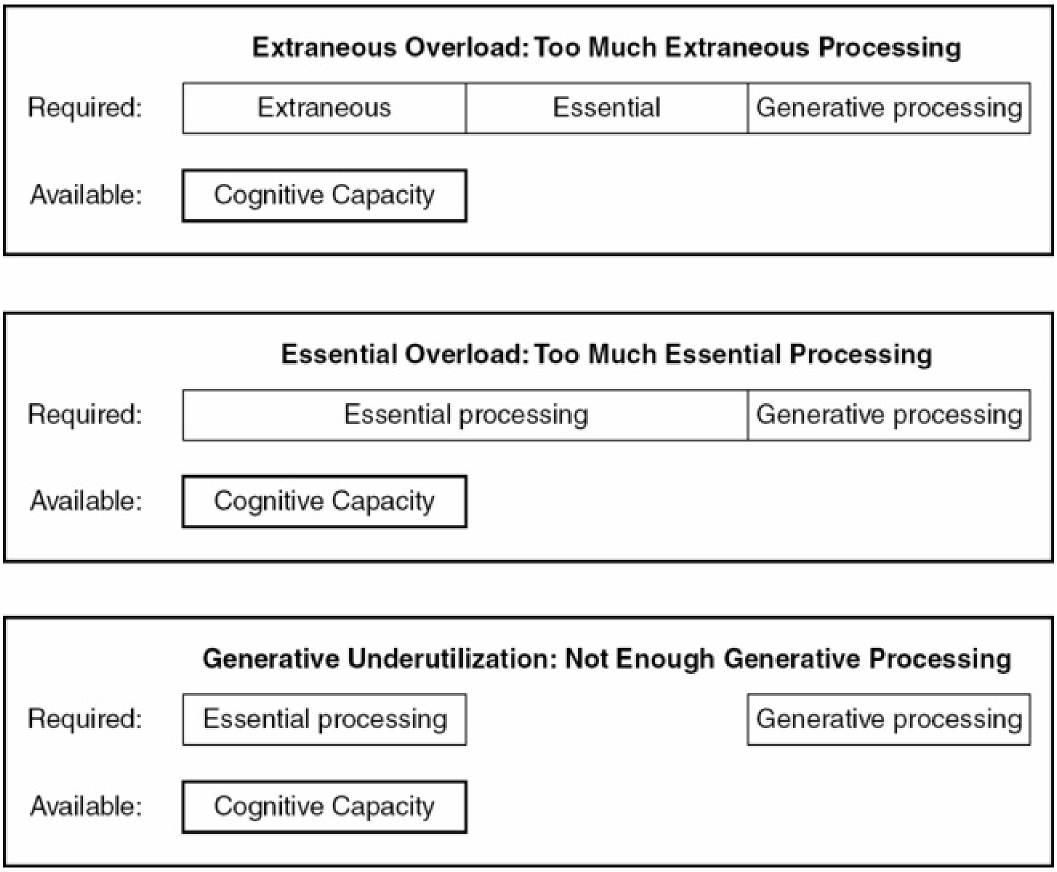


Figure 3.3. Three learning scenarios.

Next, consider what might happen when the learner is given an instructional message that is well designed so it does not create high levels of extraneous processing. The second frame in Figure 3.3 represents the *essential overload* scenario, in which the material is so complicated that the learner does not have enough cognitive capacity to mentally represent it as presented. Essential processing is needed to mentally represent the to-be-learned material, so it is not appropriate to seek to reduce essential processing. In this case, a reasonable instructional goal is to *manage essential processing*. As summarized in Table 3.7, some instructional techniques aimed at managing essential processing are the segmenting principle, pre-training principle, and modality principle, as described in Chapters 9 and 12.

Finally, suppose that the learner receives multimedia instruction that is designed to minimize extraneous processing and manage essential processing, so there is cognitive capacity available for generative processing. The third frame in Figure 3.3 represents the *generative underutilization* scenario, in which the learner has cognitive capacity available to engage in generative processing but does not exert the effort to do so. The solution to this instructional problem is to *foster generative processing*, as summarized in Table 3.7. In short, the goal is to motivate learners to exert and maintain effort to make sense of the material at a sufficient level of intensity. Some instructional design techniques aimed at fostering generative processing include the multimedia principle, personalization principle, voice principle, and embodiment principle (as explored in Chapters 7 and 13). Some learning strategies aimed at priming generative processing during learning include the guided discovery principle (Chapter 15), self-explanation principle (Chapter 17), and drawing principle (Chapter 18).

In summary, the cognitive theory of multimedia learning suggests three primary goals of instructional design: reduce extraneous processing, manage essential processing, and foster generative processing. The instructional techniques described in this handbook can be analyzed in terms of the kind of instructional goals they seek to address – helping students reduce their extraneous processing during learning (which was the original focus of much research in multimedia learning), helping students manage their essential processing during learning (in which the modality principle has enjoyed the most attention), or helping students engage in generative processing during learning (which is a newer and less researched domain).

Historical Overview

The Past: Evolution of the Cognitive Theory of Multimedia Learning

The cognitive theory of multimedia learning has evolved within the body of research papers and books produced by my colleagues and me at UCSB during the past 25 years. Although the name has changed over the years, the underlying elements of the theory – that is, dual channels, limited capacity, and active processing – have remained constant. Some names used early in the research program – such as "model of meaningful learning" (Mayer, 1989) and "cognitive conditions for effective illustrations" (Mayer & Gallini, 1990) – emphasized the active processing element; other names used later – such as "dual-coding model" (Mayer & Anderson, 1991, 1992) and "dual-processing model of multimedia learning" (Mayer & Moreno, 1998; Mayer, Moreno, Boire, & Vagge, 1999) – emphasized the dual-channel element; and yet other names – such as "generative theory" (Mayer, Steinhoff, Bower, & Mars, 1995) and "generative theory of multimedia learning" (Mayer, 1997; Plass, Chun, Mayer & Leutner, 1998) – emphasized all three elements. The current name, "cognitive theory of multimedia learning," was used in Mayer, Bove, Bryman, Mars, and Tapangco (1996), Moreno and Mayer (2000), and Mayer, Heiser, and Lonn (1991) and was selected for use in major reviews (Mayer, 2001, 2008, 2009; Mayer & Moreno, 2003) as well as the previous edition of *The Cambridge Handbook of Multimedia Learning* (Mayer, 2005).

An early predecessor of the flow chart representation shown in Figure 3.2 was a dual-coding model shown in Mayer and Sims (1994, fig. 1), which contained the same two channels and three of the same five cognitive processes but lacked two of the cognitive processes and sensory memory. Mayer, Steinhoff, Bower, and Mars (1995, fig. 1) and Mayer (1997, fig. 3) presented an intermediate version that was almost identical to the flow chart shown in Figure 3.2 except that it lacked long-term memory and sensory memory. Finally, the current version of the flow chart appeared in Mayer, Heiser, and Lonn (2001) and was reproduced in subsequent reviews (Mayer, 2001, fig. 3.2; 2002, fig. 7; 2003, fig. 2; 2005, fig. 3.2; 2009, fig. 3.1 Mayer & Moreno, 2003, fig. 1). Thus, the model has developed by the addition of components – both cognitive processes and mental representations – and the clarification of their role. The result is the cognitive theory of multimedia learning that is represented in the flow chart in Figure 3.2 of this chapter.

The primary addition represented in this chapter is the triarchic model of three demands on cognitive capacity (summarized in Table 3.6) and the three learning scenarios (summarized in Figure 3.3). These elements seek to link the cognitive theory of multimedia learning to an instructional framework; that is, the goal of these additional elements is to bridge the science of learning (represented in Figure 3.2) with the science of instruction (represented in the three kinds of instructional goals summarized in Table 3.7).

The Present: Progress Report

In the first edition of this handbook (Mayer, 2005), I called for work in (1) fleshing out the details of the mechanisms underlying the five cognitive processes and the five forms of representation, (2) integrating the various theories of multimedia learning, and (3) building a credible research base. In the ensuing decade, we have seen important progress on each of these goals. First, studying the mechanisms of cognitive processing during multimedia learning has been aided by the increasing use of new methodologies, including eye-tracking techniques (e.g., Johnson & Mayer, 2012; Scheiter & van Gog, 2009). Second, the theoretical focus has been strengthened by a focus on the three demands on cognitive capacity (as summarized in Table 3.6) as an organizing and unifying theme. Third, the research base has grown dramatically, as is indicated by the growing number of meta-analyses (Ginns, 2005, 2006; Ginns, Martin, & Marsh, 2013) and by the increasing focus on boundary conditions – that is, pinpointing the conditions under which design principles are more or less likely to apply, including the role of the learner's prior knowledge (see Chapter 24) and the learner's working memory capacity (see Chapter 25).

The Future: Incorporating Motivation and Metacognition

How will the cognitive theory of multimedia learning evolve? A useful next step would be to better incorporate the role of motivation and metacognition in multimedia learning. The rationale for this suggestion is that in addition to being able to engage in appropriate cognitive processing during multimedia learning, successful learners must want to engage in appropriate cognitive processing (i.e., motivation) and know how to manage their cognitive processing (i.e., metacognition).

Motivation to learn (which can be called *academic motivation*) refers to a learner's internal state that initiates and maintains goal-directed behavior (Mayer, 2011). According to this definition, academic motivation is (1) personal (i.e., it occurs within a learner), (2) activating (i.e., it initiates learning behavior), (3) energizing (i.e., it fosters persistence and intensity during learning), and (4) directed (i.e., it is aimed at accomplishing a learning goal). In sum, motivation to learn is reflected in the learner's willingness to exert effort to engage in appropriate cognitive processing during learning (such as the processes of selecting, organizing, and integrating that are needed for meaningful learning).

Metacognition in multimedia learning refers to the learner's awareness and control of cognitive processing during learning (Mayer, 2011). Metacognition plays a crucial role in multimedia learning by guiding the learner's cognitive processing during learning, such as when a learner knows which cognitive activity would be best for a particular learning task and adjusts cognitive activity on the basis of how well it is helping learning. In short, effective multimedia learning includes helping learners become self-regulated learners – that is, learners who take responsibility for managing their cognitive processing during learning.

Although the learner's motivation to learn is part of the definition of generative processing (as summarized in Table 3.6), the overall role of motivation and metacognition is an underdeveloped aspect of the cognitive theory of multimedia learning (Mayer, 2014). Moreno's (2007; Moreno & Mayer, 2007) cognitive affective theory of learning with media seeks to expand multimedia learning theory by more explicitly incorporating the role of motivation and metacognition, highlighted by adding arrows from long-term memory back to the cognitive processing arrows of selecting, organizing, and integrating. Consistent with this approach, Figure 3.4 (adapted from Mayer, 2011) presents a modified version of the cognitive theory of multimedia learning that takes a preliminary step in acknowledging the role of motivation and metacognition in multimedia learning by adding arrows from long-term memory back to the cognitive processing arrows of selecting, organizing, and integrating.

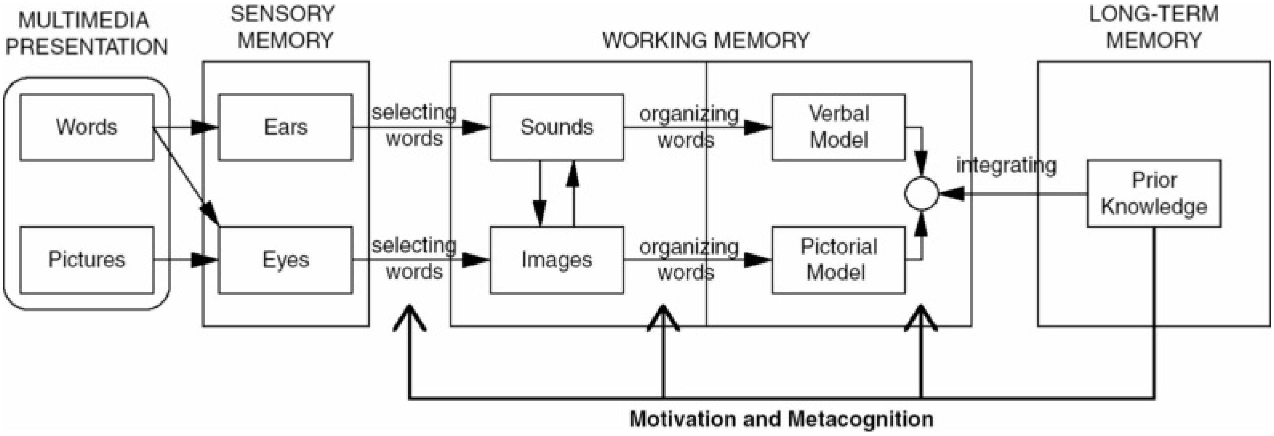


Figure 3.4. Incorporating motivation and metacognition into a cognitive theory of multimedia learning.

Future research is needed to spell out in greater detail the mechanisms of motivation and metacognition (i.e., how the added arrows work) and to test relevant instructional techniques for promoting academic motivation, such as using emotional design principles to create appealing but relevant graphics (Um, Plass, Hayward, & Homer, 2011). In addition, work is needed to develop dependent measures for learning outcomes, including the use of delayed tests of retention and transfer; and better measures of the learning process, including measures of cognitive load, motivation, and metacognitive control. Methodological advances including EEG, fMRI, eye-tracking methods, and physiological measures may contribute to these efforts.

Conclusion

In summary, multimedia learning takes place within the learner's information processing system – a system that contains separate channels for visual and verbal processing, a system with serious limitations on the capacity of each channel, and a system that requires appropriate cognitive processing in each channel for active learning to occur. In particular, multimedia learning is a demanding process that requires selecting relevant words and images, organizing them into coherent verbal and pictorial representations, and integrating the verbal and pictorial representations with each other and with relevant prior knowledge. In the process of multimedia learning, material is represented in five forms – as words and pictures in a multimedia presentation, acoustic and iconic representations in sensory memory, sounds and images in working memory, verbal and pictorial models in working memory, and knowledge in long-term memory. During learning, cognitive capacity must be allocated among extraneous, essential, and generative processing, so the goal of instructional design is to develop effective techniques for reducing extraneous processing, managing essential processing, and fostering generative processing.

The theme of this chapter is that multimedia messages should be designed to facilitate multimedia learning processes. Multimedia messages that are designed in light of how the human mind works are more likely to lead to meaningful learning than those that are not. This proposition is tested empirically in the chapters of this handbook.

Glossary

Cognitive theory of multimedia learning:	A theory of how people learn from words and pictures, based on the idea that people possess separate channels for processing verbal and visual material (dual-channel assumption), each channel can process only a small amount of material at a time (limited-capacity assumption), and meaningful learning involves engaging in appropriate cognitive processing during learning (active processing assumption).
Essential processing:	Cognitive processing during learning that is needed to represent the essential presented material in working memory and is caused by the complexity of the material.
Extraneous processing:	Cognitive processing during learning that does not serve the instructional objective and is caused by poor instructional design.
Generative processing:	Cognitive processing during learning that is aimed at making sense of the essential material in the lesson and is caused by the learner's motivation to exert effort.
Integrating:	A cognitive process in which the learner builds connections between visual and verbal representations in working memory and between them and relevant prior knowledge activated from long-term memory.
Long-term memory:	A memory store that holds large amounts of knowledge over long periods of time.
Multimedia instructional message:	A communication containing words and pictures intended to foster learning.
Multimedia principle:	People learn more deeply from words and pictures than from words alone.
Organizing:	A cognitive process in which the learner mentally arranges the incoming information in working memory into a coherent cognitive representation.
Selecting:	A cognitive process in which the learner pays attention to relevant incoming

	material and transfers it to working memory for further processing.
Sensory memory:	A memory store that holds pictures and printed text impinging on the eyes as exact visual images for a very brief period and that holds spoken words and other sounds impinging on the ears as exact auditory images for a very brief period.
Working memory:	A limited-capacity memory store for holding and manipulating sounds and images in active consciousness.

Acknowledgments

I wish to thank Jeroen van Merriënboer, John Sweller, and Wolfgang Scnotz for their many useful comments. Preparation of this chapter was supported by a grant from the Office of Naval Research. The chapter is a revision of chapter 3 of the first edition of *The Cambridge Handbook of Multimedia Learning*.

References

Baddeley, A. D. (1992). Working memory. *Science*, 255, 556–559.

Baddeley, A. D. (1999). *Human memory*. Boston: Allyn & Bacon.

Baddeley, A. D., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. Hove: Psychology Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn*. Washington, DC: National Academy Press.

Chambliss, M. J., & Calfee, R. C. (1998). *Textbooks for learning*. Oxford: Blackwell.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3, 149–210.

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology*, 80, 448–456.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, 15, 313–332.

Ginns, P. (2006). Integrating information: A meta-analysis of spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16, 511–525.

Ginns, P., Martin, A. J., & Marsh, H. M. (2013). Designing instructional text in a conversational style: A meta-analysis. *Educational Psychology Review*, 25, 445–472.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.) (2009). *Handbook of metacognition in education*. New York: Routledge.

Johnson, C., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied*, 18, 178–191.

Kintsch, W. (1998). *Comprehension*. New York: Cambridge University Press.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, 81, 240–246.

Mayer, R. E. (1992). Cognition and instruction: Their historic meeting within educational psychology. *Journal of Educational Psychology*, 84, 405–412.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32, 1–19.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2002). Multimedia learning. In B. H. Ross (Ed.), *The psychology of learning and motivation*, vol. 41 (pp. 85–139). San Diego, CA: Academic Press.

Mayer, R. E. (2003). The promise of multimedia learning: Using the same instructional design methods across different media. *Learning and Instruction*, 12, 125–141.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, 63(8), 760–769.

Mayer, R. E. (2009). *Multimedia learning* (2d ed). New York: Cambridge University Press.

Mayer, R. E. (2011). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.

Mayer, R. E. (2014). Incorporating motivation into multimedia learning. *Learning and Instruction*, 24, 171–173.

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of the dual-coding hypothesis. *Journal of Educational Psychology*, 83, 484–490.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84, 444–452.

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, 88, 64–73.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, 82, 715–726.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93, 187–198.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90, 312–320.

Mayer, R. E. & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38, 43–52.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology*, 91, 638–643.

Mayer, R. E., & Sims, V. K., (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, 86, 389–401.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology*

Research & Development, 43, 31–43.

Miller, G. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.

Moreno, R. (2007). Optimising learning from animations by minimizing cognitive load: Cognitive and affective consequences of signaling and segmentation methods. *Applied Cognitive Psychology*, 21, 765–781.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, 92, 117–125.

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19, 309–326.

Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford, England: Oxford University Press.

Paivio, A. (2006). *Mind and its evolution: A dual coding approach*. Mahwah, NJ: Lawrence Erlbaum.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, 90, 25–36.

Plass, J. L., Moreno, R., & Brunken, R. (Eds.) (2010). *Cognitive load theory*. New York: Cambridge University Press.

Scheiter, K., & van Gog, T. (2009). Using eye tracking in applied research to study and stimulate the processing of information from multi-representational sources. *Applied Cognitive Psychology*, 23, 1209–1214.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13, 141–156.

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell: ACER Press.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2011). Emotional design in multimedia learning. *Journal of Educational Psychology*, 104, 485–498.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, 24, 345–376.

4 Integrated Model of Text and Picture Comprehension

Wolfgang Schnotz

University of Koblenz-Landau

Abstract

This chapter presents an integrated model of text and picture comprehension that takes into account that learners can use multiple sensory modalities combined with different forms of representation. The model encompasses listening comprehension, reading comprehension, visual picture comprehension, and auditory picture comprehension (i.e., sound comprehension). The model's cognitive architecture consists of modality-specific sensory registers, working memory, and long-term memory. Within this architecture, a distinction is made between perception-bound processing of text surface or picture surface structures, on the one hand, and cognitive processing of semantic deep structures, on the other hand. The perception-bound processing of text surface structures includes phonological and graphemic input analyses with graphemic–phonemic conversion, leading to lexical patterns. The perception-based processing of picture surface structures includes visual or acoustic nonverbal feature analyses, leading to visuospatial patterns or sound patterns. The cognitive processing includes descriptive processing of lexical patterns (via parsing), which leads to propositional representations, and depictive processing of spatial or sound patterns (via structure mapping), which leads to mental models. Propositional representations and mental models interact via model construction and model inspection processes. After presenting the integrated model of text and picture comprehension, the chapter derives predictions, which can be empirically tested. It reports research findings that can be explained by the model, and it derives practical suggestions for instructional design. Finally, the chapter discusses limitations of the model and points out directions for further research.

Introduction

The term *multimedia* has different meanings at different levels. At the level of technology, it means the use of multiple delivery media such as computers, screens, and loudspeakers. At the level of presentation formats, it means the use of different forms of representation such as texts and pictures. At the level of sensory modalities, it means the use of multiple sensory organs such as the eye and the ear. Although highly important in terms of practical reliability, the level of technology is not very interesting from a psychological point of view. Comprehending a text printed on paper does not fundamentally differ from comprehending the same text on a computer screen. In fact, comprehension is highly dependent on what kind of information is presented and how it is presented. The psychology of multimedia learning focuses therefore on the level of presentation formats and on the level of sensory modalities.

What is multimedia learning? From a psychological point of view, the core of multimedia learning is the combined comprehension of text and pictures (Mayer, 1997). This does not necessarily require high technology. Multimedia learning is also possible with printed books or blackboards instead of computer screens and with the human voice instead of loudspeakers. Multimedia learning is therefore not a modern phenomenon. Instead, it has a long tradition going back to Comenius (1999), who emphasized the importance of adding pictures to texts in his pioneer work *Orbis Sensualium Pictus* (published first in 1658).

Multimedia learning can occur in different forms. Learners can listen to a lecture accompanied by pictures (i.e., lecture-based multimedia learning). They can read a book with pictures (i.e., book-based multimedia learning). Finally, they can read an illustrated text from the Internet on a computer screen or listen to a text accompanied by pictures from a loudspeaker (i.e., computer-based multimedia learning) (Mayer, 2009).

Individuals usually combine these different kinds of multimedia learning. Consider the following example. A teacher explains to her class of eighth graders the migration of birds in Europe. She presents a map of the European continent (shown in Figure 4.1a) that indicates where some birds live in summer and where they stay in winter. While pointing to the map, she gives oral explanations like the following:

> (a) “Many birds breed in middle and northern Europe in summer, but do not stay there during winter. Instead, they fly in September to warmer areas in the Mediterranean area. These birds are called ‘migrant’.”

After the lesson, Daniel, one of her students, has to learn as a homework task about a specific bird, the marsh harrier, and to give a report to his classmates the next day. Daniel walks into a library and opens a printed encyclopedia of biology, where he finds a drawing of the marsh harrier (Figure 4.1b) and the following text:

> (b) “The marsh harrier is a bird of prey with an average wingspan of 47″ and a face similar to that of an owl. The drawing shows the typical gliding position of the bird. The marsh harrier is usually found in wetlands, especially in marshes, swamps, and lagoons. It feeds mostly on small birds or mammals (like rodents or rabbits) and on reptiles. The marsh harrier is migrant.”

As the encyclopedia does not contain further information about the bird’s migration, Daniel decides to search the Internet, where he finds a Web site including a bar graph (Figure 4.1c) and the following text:

> (c) “The marsh harrier is found all year round in Spain, France, and around the Mediterranean. In other areas of Europe the bird is migrant, breeding in middle and northern Europe while wintering in tropical marshes and swamps in North Africa. The bar graph shows a typical frequency pattern of marsh harriers in a middle European habitat.”

Furthermore, the Web site offers a sound button. After clicking on it, Daniel hears the typical call of a marsh harrier near its breeding place.

Figure 4.1a. Map of bird migration in Europe.

Figure 4.1b. Drawing of a marsh harrier.

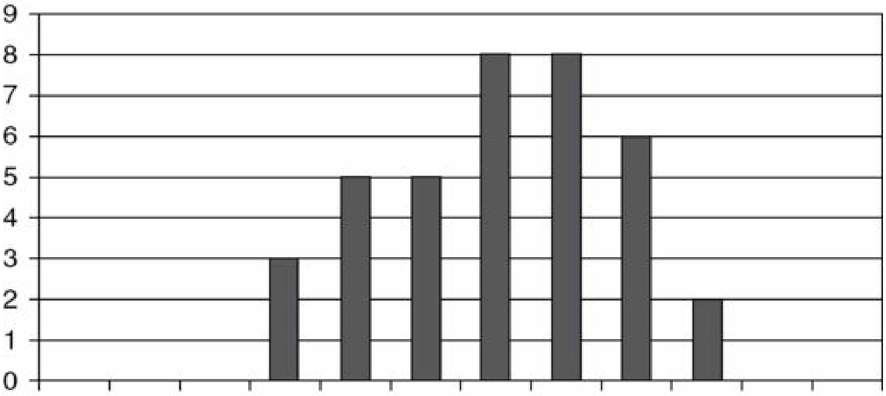


Figure 4.1c. Bar graph of the marsh harrier's observation frequency in a middle European habitat.

Altogether, Daniel has practiced three kinds of multimedia learning using various external sources of information. At school, he has performed lecture-based multimedia learning, using the map and the teacher's oral text as information sources. In the library, he has performed book-based multimedia learning, using the drawing of the bird and the printed text as information sources. With the Internet, he has performed computer-based multimedia learning, using the bar graph, the on-screen text, and the sound pattern as information sources. In each case, information was presented to him in different formats, such as visual texts, visual pictures (a map, a drawing, a bar graph), and sound, and he processed information through different sensory modalities: the visual modality (written text and pictures) and the auditory modality (oral text and sound).

As the example demonstrates, multimedia learning environments can be rather complex and they can involve a variety of external representations of the learning content. These representations can take different forms, such as spoken text, written text, maps, drawings, graphs, and sound. Multimedia learning occurs when an individual understands what is presented, that is, when the individual uses external representations in order to construct internal (mental) representations of the learning content in working memory and if he or she stores these representations in long-term memory.

In the first part of this chapter, a distinction between two different forms of representations is made and applied to both external and internal representations. The second part investigates how multimedia comprehension and learning are constrained by the human cognitive architecture. In the third part, the theoretical concepts that have been introduced will be combined into an integrated model of text and picture comprehension, which involves listening comprehension, reading comprehension, visual picture comprehension, and auditory picture comprehension (i.e., sound comprehension). The fourth part presents empirical evidence for the integrated model, and the fifth part explains what kind of instructional consequences can be derived from the integrated model. Finally, the sixth part points out the limitations of the model and suggests directions of future research in the area.

External and Internal Representations

Forms of Representation

How many basic forms of representation exist? Despite numerous variants of representations, there are only two basic forms: descriptions and depictions. Texts are the most common kind of descriptions. However, there are also other kinds of descriptive representations. Mathematical expressions such as $V = s^3$ (describing the relation between a cube's size and its volume) and the formula $F = m * a$ in physics (describing the relation between force, mass, and acceleration according to Newton's second law) are also descriptive representations. Descriptive representations consist of symbols. Symbols are signs that have no similarity with their referent (Peirce, 1931–1958). The word *bird*, for example, has no similarity to a real bird. It is a symbol, and its meaning is based on a convention. In texts, we use nouns (such as *bird* and *breeding*) as symbols for objects and events. We use verbs and prepositions (such as *feed* and *on*) as symbols for relations, and we use adjectives (such as *small* and *migrant*) as symbols for attributes.

Pictures such as photographs, drawings, paintings, and maps are depictive representations. It should be noted, however, that pictures are not the only kind of depictive representations. A miniature model of a building, a line graph, or the swing of a measuring tool pointer are also depictive representations. Depictive representations consist of icons. Icons are signs that are associated with their referent by similarity or by another structural commonality. A map such as that in Figure 4.1a and a drawing of a bird as in Figure 4.1b are graphical objects that have some similarity to the corresponding referent (i.e., the European continent or the marsh harrier). Graphs have a more abstract structural commonality with their referent. The meaning of the bar graph shown in Figure 4.1c, for example, is based on an analogy: the height of the bars corresponds to the number of marsh harriers observed in a habitat during the corresponding month, and the sequence of the bars corresponds to the sequence of months during the year.

Descriptive representations and depictive representations have different uses for different purposes. On the one hand, descriptive representations are more powerful in expressing abstract knowledge. For example, it is no problem to say, “The marsh harrier feeds on mammals or reptiles,” which connects abstract concepts (e.g., *mammals*, *reptiles*) by a disjunctive *or*. In a depictive representation, on the contrary, it is possible to show only a specific mammal (e.g., a mouse) or a specific reptile (e.g., a lizard). The disjunctive *or* cannot be represented by only one picture. It requires a series of pictures (e.g., one picture showing the bird eating a mouse and another picture showing the bird eating a lizard). On the other hand, depictive representations have the advantage of being informationally complete. A map, for example, includes all geometric information of the depicted geographical area, and a picture of a marsh harrier eating a mouse includes not only information about the shape of the bird and the shape of a mouse, but necessarily also information about their size, their orientation in space, how the bird holds its prey, and so on. Depictive representations are therefore more useful for drawing inferences, because the new information can be read off directly from the representation (Kosslyn, 1994).

Mental Representations

Does the distinction between descriptive and depictive representations apply also to internal (i.e., mental) representations? Research on text comprehension suggests that learners reading a text or listening to a text construct three kinds of mental representations (Graesser, Millis, & Zwaan, 1997; Kintsch, 1998; McNamara, 2007; van Dijk, & Kintsch, 1983; van Oostendorp & Goldman, 1999; Weaver, Mannes, & Fletcher, 1995). For example, when a learner reads a sentence like “Some migrant birds fly to the south of Europe for wintering,” he or she forms a mental representation of the text surface structure. This text surface representation cannot be referred to as understanding yet, but it allows repetition of what has been read. On the basis of this surface representation, the reader then constructs a propositional representation. This representation includes the ideas expressed in the text at a conceptual level, which is independent of the specific wording and syntax of the sentence. In the preceding example, this would include the idea that migrant birds in Europe fly south in September, represented by the proposition FLY(agent: MIGRANT BIRDS, location: EUROPE, aim: SOUTH, time: SEPTEMBER). Finally the reader constructs a mental model of the text content. In the preceding example, this could be a mental map of Europe, including a movement from the north to the south.

Research on picture comprehension suggests that when learners understand a picture, they also construct multiple mental representations (Kosslyn, 1994; Lowe, 1996). Accordingly, learners create a perceptual representation (i.e., a visual image) of the picture, and they then construct a mental model of the picture’s content. For example, when a learner understands the bar graph shown in Figure 4.1c, the learner perceives vertical bars on a horizontal line and creates a corresponding visual image. On the basis of this visual image, he or she constructs a mental model of a middle European habitat that includes different numbers of marsh harriers during the course of the year. The mental model can be used for reading specific information as, for example, that the birds stay in this habitat during the summer. The information read-off from the model is again encoded in a propositional format such as, for example, STAY(agent: BIRDS; location: HABITAT, time: SUMMER).

The distinction between descriptive and depictive representations previously mentioned applies also to these mental representations. A text surface representation and a propositional representation are descriptive representations, as they use symbols to describe the subject matter. A visual image and a mental model, on the contrary, are depictive representations, as they are assumed to have an inherent structure that corresponds to the structure of the subject matter (Johnson-Laird, 1983; Kosslyn, 1994). A visual image is sensory-specific because it is linked to the visual modality, whereas a mental model is not sensory-specific because it is able to integrate information from different sensory modalities. It is possible, for example, to construct a mental model of some spatial configuration based on visual, auditory, and touch information. This implies that a mental model is more abstract than a visual image. In picture comprehension, mental models and visual images can also differ in terms of their information content. On the one hand, irrelevant details of the picture, which are included in the visual image, may be ignored in the mental model. On the other hand, the mental model contains additional information from prior knowledge that is not included in the visual image. In understanding bird migration, for example, a mental model of the European continent might include snowfall in northern areas during winter, although no snow is indicated on the map.

On the basis of the distinction between descriptive and depictive representations, Schnotz and

Bannert (2003) have proposed a theoretical framework for analyzing text and picture comprehension. The framework, which is shown in Figure 4.2, includes a branch of descriptive representations (left side) and a branch of depictive representations (right side) with correspondingly different types of information processing. The descriptive branch involves the external text, the mental text surface representation, and the mental propositional representation of the subject matter. Information processing in the descriptive branch implies (subsemantic and semantic) analysis of symbol structures. The depictive branch involves the external picture, the visual image of the picture, and the mental model of the subject matter. Information processing in the depictive branch implies analog structure mapping (based on perception and thematic selection). The framework corresponds to the dual-coding concept of Paivio (1986), who assumes a verbal system and an image system in the human mind with different forms of mental codes. However, contrary to the traditional dual-coding theory, the framework assumes that multiple representations are formed in text comprehension as well as in picture comprehension.

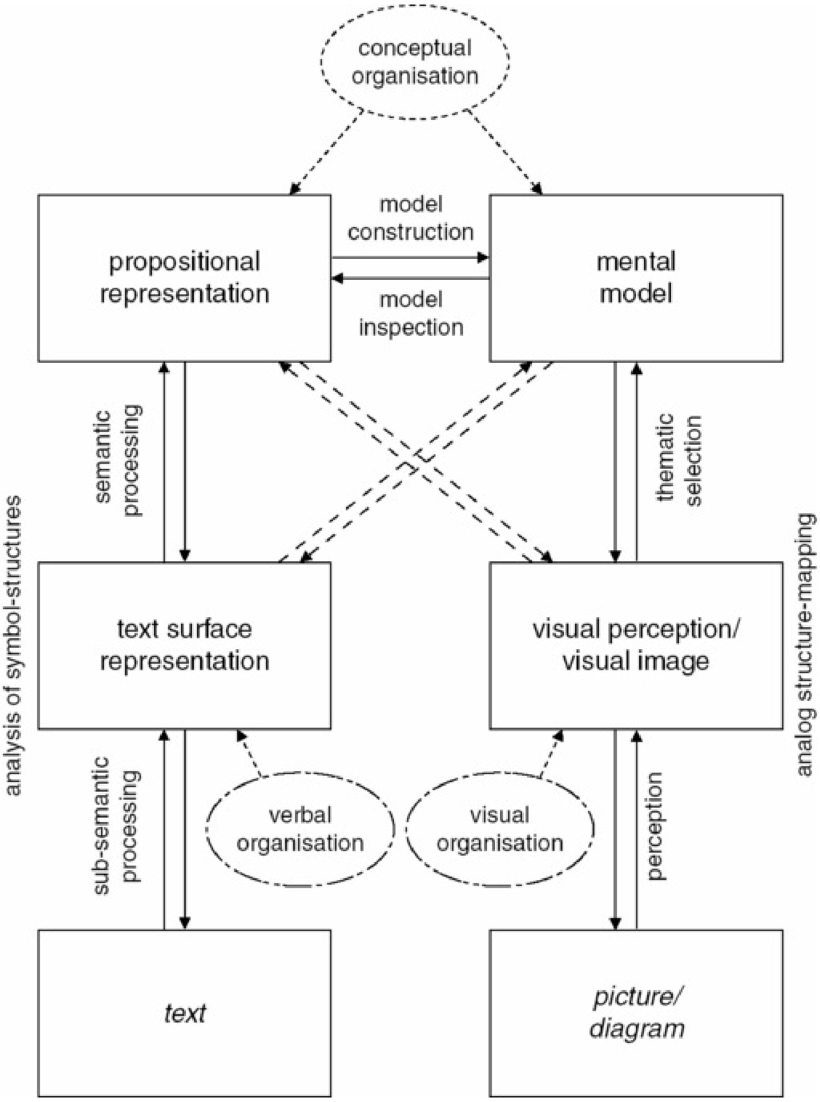


Figure 4.2. Theoretical framework for analyzing text and picture comprehension proposed by Schnotz and Bannert (2003). A distinction is made between processing of descriptions (symbol structures) and processing of depictions (analog structures).

Cognitive Architecture for Text and Picture Comprehension

When learners understand texts and pictures, they construct multiple mental representations in their cognitive system. Research in cognitive psychology suggests that the architecture of the human cognitive system includes multiple memory systems. A common view proposed by Atkinson and Shiffrin (1971) distinguishes three memory subsystems – sensory registers, working memory, and long-term memory – with different functions and different constraints on processing texts and pictures.

Sensory Registers

Information enters the cognitive system from the outside world through sensory organs, which convey the information through sensory channels to working memory. It should be noted that there is no inherent relationship between sensory modalities and representational formats. For example, written text is usually visual language read with the eyes, but it can also be read with the fingers (e.g., in the case of blind people reading Braille). Similarly, pictures are usually seen with the eyes, but they can sometimes also be perceived by touch (e.g., maps for blind people). Spoken text is usually perceived by the ear, but deaf people can also read lips and touch the vibrating larynx. Auditory pictures (i.e., sound patterns imitating an original sound as, for example, the call of a bird) are perceived by the ear too. Although multiple sensory modalities can be involved in text and picture comprehension, we will consider in the following only the visual and the auditory modalities.

Visual information that meets the eye is stored very briefly (less than 1 second) in a visual register. If attention is directed to information in the visual register, the information gets transmitted to visual working memory. Auditory information that meets the ear is stored briefly (less than 3 seconds) in an auditory register. If attention is directed to information in the auditory register, the information gets transmitted to auditory working memory.

Working Memory

Written or spoken text and visual or auditory pictures are further processed in a working memory with a highly constrained capacity for storing and processing of information (see Chapter 25). According to Baddeley (1986), working memory consists of a central executive and different subsystems for the storage of information.

Two of these subsystems have received much attention in research: auditory working memory and visual working memory. Auditory working memory is conceived as a phonological-articulatory loop. Visual working memory is conceived as a visuospatial sketchpad. The phonological-articulatory loop specializes in verbal material presented in the auditory modality, but it can also deal with nonverbal sound. It has limited capacity, corresponding on average to what can be articulated within about 2 seconds. Nevertheless, people with a highly reduced phonological-articulatory loop are still capable of normal language comprehension (Baddeley, 2000; Vallar & Shallice, 1990). Spoken text activates phonological lexical patterns, whereas auditory pictures activate sound patterns in auditory working memory. The visuospatial sketchpad specializes in spatial information presented in the visual modality. It has a limited capacity of about 5 units on the average. Written text activates graphemic lexical patterns, whereas visual pictures activate visuospatial patterns in visual working memory.

As working memory plays an important role at higher levels of text comprehension too (Daneman & Carpenter, 1983), one can furthermore assume a propositional subsystem that allows holding a limited number of propositions simultaneously in working memory (Kintsch & van Dijk, 1978). Propositions result from descriptive processing of phonological or graphemic lexical patterns through parsing the incoming word sequences combined with prior knowledge. Finally, research findings suggest a subsystem for mental model construction in working memory. Mental model construction seems to be influenced by the visuospatial sketchpad rather than the phonological-articulatory loop (Friedman & Miyake, 2000). More specifically, it is highly related to spatial cognitive processing (Sims & Hegarty, 1997). Consistent with these findings, research by Knauff and Johnson-Laird (2002) indicates that visual imagery and spatial reasoning are based on different cognitive subsystems. This suggests a distinction between a visual working memory (or sketchpad) for visual images and a spatial working memory for mental model construction. Accordingly, mental models result from depictive processing of visuospatial or sound patterns through structure mapping.

Long-Term Memory

Text comprehension and picture comprehension require prior knowledge stored in long-term memory, which includes lexical knowledge as well as perceptual and cognitive world knowledge. Lexical knowledge encompasses the mental phonological lexicon and the mental graphemic lexicon, which include knowledge about auditory or visual word forms. The phonological lexicon (also called *auditory lexicon*) includes phonological lexical patterns, which represent knowledge about the sound of spoken words required for spoken word recognition. Listening to a text implies activation of such phonological lexical patterns in working memory. Individuals who suffer from word deafness (due to brain injuries) have a deficient phonological lexicon: they can hear sounds but cannot separate and identify words when listening to spoken language. Individuals who suffer from word meaning deafness can repeat spoken words without understanding them, although they can understand written words. These individuals possess a phonological lexicon, but this is unconnected to semantic (long-term) memory. The graphemic lexicon (also called *visual* or *orthographic lexicon*) includes graphemic lexical patterns, which represent knowledge about the visual appearance of written words required for written word recognition. Reading a text implies activation of such graphemic lexical patterns in working memory. Individuals who suffer from pure alexia (due to illiteracy or brain injuries) have a deficient graphemic lexicon: they can understand spoken words but cannot understand written words, although their vision is intact (Ellis & Young, 1996).

Perceptual world knowledge refers to the appearance of objects – for example, what different kinds of birds typically look like. This knowledge is needed for the visual perception or imagination of objects, that is, for the creation of corresponding visuospatial patterns in working memory (Kosslyn, 1994; Rosch, 1978). Objects can be recognized faster and more easily when they are presented from a typical perspective (such as the bird shown in Figure 4.1b) than when they are presented from an unusual perspective (Palmer, Rosch, & Chase, 1981). Conceptual world knowledge refers to the relations within a domain – for example, the breeding of birds and the meteorological conditions in different seasons. This knowledge is needed both for the construction of a propositional representation and for the construction of a mental model (e.g., of bird migration) in working memory.

Text and picture comprehension is therefore based not only on external sources of information (the text and the picture), but also on prior knowledge as an internal source of information. Prior knowledge can partially compensate for a lack of external information, for lower working memory capacity (Adams, Bell, & Perfetti, 1995; Miller & Stine-Morrow, 1998), and for deficits of the propositional representation (Dutke, 1996; McNamara, Kintsch, Songer, & Kintsch, 1996; Soederberg Miller, 2001). There seems to be a trade-off between the use of external and internal information sources: pictures are analyzed more intensively if the content is difficult and the learner's prior knowledge is low (Carney & Levin, 2002).

Integrated Comprehension of Text and Pictures

The idea of a cognitive architecture, including multiple memory systems with multiple sensory channels of limited capacity and a working memory of limited capacity operating on descriptive and depictive representations, is incorporated into an integrative model of text and picture comprehension (or ITPC model). The model integrates the concepts of multiple memory systems (Atkinson & Shiffrin, 1971), working memory (Baddeley, 1986, 2000), and dual coding (Paivio, 1986). It furthermore integrates the idea of multiple forms of mental representations in text comprehension or picture comprehension (Kosslyn, 1994; van Dijk & Kintsch, 1983) and neuropsychological models of word recognition and reading (Ellis & Young, 1996). Naturally, the model has commonalities with the precursive model of text and picture comprehension of Schnotz and Bannert (2003), as well as with the cognitive theory of multimedia learning (CTML) of Mayer (2009; see also Chapter 3). The model, which is schematically shown in Figure 4.3, aims at representing the single or combined comprehension of spoken text, written text, visual pictures, and auditory pictures (i.e., sound images). It is based on the following assumptions:

1. Text and picture comprehension take place in a cognitive architecture including modality-specific sensory registers as information input systems, a working memory of limited capacity, and a long-term memory.
2. Verbal information (i.e., information from spoken or written texts) and pictorial information (i.e., information from visual pictures or from sound pictures) is transmitted to working memory through visual channels and auditory channels. The channels have limited capacity to

process and transmit information.

3. Further semantic processing in working memory takes place in two different subsystems: a descriptive subsystem and a depictive subsystem. Text (spoken or written) is first processed in the descriptive subsystem and then in the depictive subsystem. Pictures (visual or auditory) are first processed in the depictive subsystem and then in the descriptive subsystem.
4. Text comprehension and picture comprehension are active processes of coherence formation. In comprehension, individuals engage in building coherent knowledge structures from the available external verbal and pictorial information and from their prior knowledge.

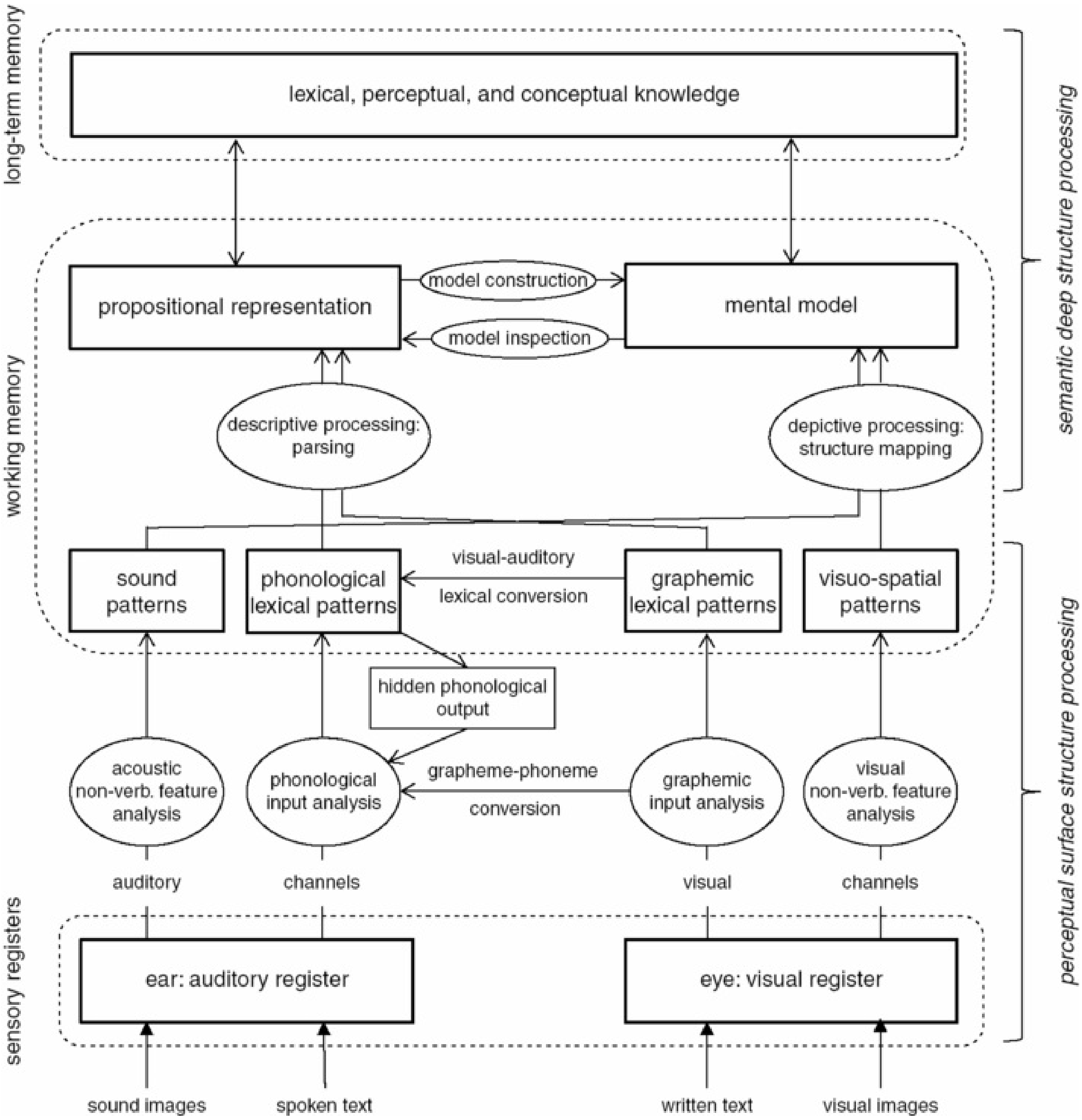


Figure 4.3. Integrated model of text and picture comprehension.

A distinction between perceptual surface structure processing and semantic deep structure processing can be made within the model. Perceptual surface structure processing refers to the information transfer from the surface structure of texts and pictures to working memory. It is characterized by (verbal) phonological or graphemic analyses and (nonverbal) visual or acoustic feature analyses leading to corresponding input patterns in auditory or visual working memory. Semantic deep structure processing refers to the cognitive processing within working memory that results in propositional representations and mental models as well as the information exchange between long-term and working memory. It is characterized by the functioning of the descriptive and the depictive subsystems and their interaction.

Listening comprehension. If a spoken text is understood, auditory verbal information enters the auditory register through the ear and is then the object of phonological input analysis, which identifies phonemes within the acoustic input, leading to phonological lexical patterns. Further descriptive processing (parsing of word sequences and further semantic analysis) leads to a propositional representation, which finally triggers the construction or elaboration of a mental model. In the example of a text on bird migration, phonological analysis of the spoken word *bird* leads (via the mental lexicon) to the activation of its phonological pattern in auditory working memory. Further processing through the descriptive subsystem results in the activation of the concept BIRD, which is then included in a propositional representation. This representation finally triggers the construction of a mental model of bird migration.

Reading comprehension. If a written text is understood, visually presented verbal information enters the visual register through the eye and is then subjected to graphemic input analysis, which identifies graphemes within the visual input. In skilled reading, this analysis leads to graphemic lexical patterns. These patterns are further processed in the descriptive subsystem. This results in the formation of a propositional representation, which in turn triggers the construction or elaboration of a mental model. In a text on bird migration, for example, graphemic analysis of the written word *bird* leads (via the mental lexicon) to the activation of its graphemic pattern in visual working memory. Further processing through the descriptive subsystem results in the activation of the concept BIRD, which is included into a propositional representation. This representation finally triggers the construction of a mental model of bird migration.

In nonskilled reading due to a deficient graphemic mental lexicon (e.g., by reading beginners), the individual has to apply grapheme–phoneme conversion rules by engaging in tedious phonological recoding of the visual input, which finally allows understanding of the internally spoken text. The grapheme–phoneme conversion rules, which are neither lexical nor semantic, convert letter strings into phoneme strings (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). With a comprehensive graphemic lexicon, on the contrary, texts can be understood via the activation of graphemic lexical patterns without the inclusion of any acoustic patterns (Ellis & Young, 1996, p. 219). Nevertheless, even skilled readers engage at least to some extent in graphemic–phonemic lexical conversion (Rieben & Perfetti, 1991) operating at the whole-word (lexical) level instead of the (sublexical) grapheme–phoneme level. It should be noted that this conversion represents per se a nonsemantic lexical route of word recognition: it allows word recognition even when the meaning of the word is not understood (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001).

Thus, when familiar written words are recognized, the activated graphemic lexical patterns usually also activate phonological lexical output patterns that allow the reader to pronounce these words. The pronunciation does not imply reading aloud. It can also take the form of *inner speech* as a hidden phonological output that can be heard by the reader through his or her *inner ear*. The inner pronunciation feeds into the phonological input analysis and activates phonological lexical input patterns, which are further processed through the descriptive subsystem, as already mentioned. As a result, read words can be processed via both graphemic and phonological lexical patterns (Ellis & Young, 1996, p. 219). Graphemic–phonemic lexical conversion seems to be especially important to trigger parsing procedures (e.g., analyzing word order in sentence structure analysis). Caramazza, Berndt, and Basili (1983) found that anomalies in the syntax of sentences are more easily detected when inner speech is possible than when it is suppressed. Although there is a direct route in reading from graphemic lexical patterns to further semantic analysis in the descriptive subsystem, this route does obviously not trigger syntactic analysis of the sentences. Syntactic processes “appear to operate upon a speech-based code, so that written sentences which are to undergo syntactic analysis must first be converted into spoken form and then recycled back to auditory comprehension processes” (Ellis & Young, 1996, p. 221).

Visual picture comprehension. If a visual picture is understood, visual pictorial information enters the visual register through the eye and is then subjected to visual feature analysis, which results in visuospatial patterns in working memory as a visual perceptual representation of the picture. Further depictive processing through the mapping of selected perceptual structures leads to the construction or elaboration of a corresponding mental model. This model can then be used by model inspection for reading new information that is encoded in a propositional format in working memory. For example, if a map about bird migration in Europe such as in Figure 4.1b is understood, the visual pattern of the map creates via visual feature analysis an internal visual image of the map in visual working memory. Selected information is then further processed via structure mapping, which results in the construction or elaboration of a mental model of bird migration in Europe. The individual can then read further information from the model (such as the fact that migrant birds fly from northern Europe to the Mediterranean area in the fall).

Auditory picture comprehension (sound comprehension). If a sound is understood, auditory pictorial information enters the auditory register through the ear and is then the object of acoustic feature analysis, which results in sound patterns in working memory as an auditory perceptual representation of the sound. Further depictive processing through the mapping of selected perceptual structures leads to the construction or elaboration of a corresponding mental model. This model can then be used by model inspection for reading new information that is encoded in a propositional format in working memory. For example, if the call of a marsh harrier (as a bird of prey) and the call of a small bird (as its possible prey) are heard, acoustic feature analysis leads to sound patterns forming an auditory perception (i.e., an auditory internal image) in auditory working memory. If an individual has sufficient knowledge about different birds, selected information can be further processed via structure mapping, which leads to the construction or elaboration of the mental model of a predator–prey scenario. The individual can then read further information from the mental model (e.g., that a small bird is in danger of falling prey to a marsh harrier).

It should be noted that according to this theoretical model, picture comprehension provides more direct access to mental model construction than does text comprehension, because pictures are immediately processed by the depictive subsystem, whereas texts are first processed by the descriptive subsystem, which usually leaves some ambiguity that has to be removed via the depictive subsystem (cf. Ainsworth, 1999)

Which cognitive processes lead to meaningful learning? Meaningful learning from text and pictures requires a coordinated set of cognitive processes, including the selection of information, organization of information, activation of prior knowledge, and active coherence formation by the integration of information from different sources. In the comprehension of written or spoken texts, the learner selects relevant verbal information from words, sentences, and paragraphs as an external source of information. He or she organizes the information, activates related prior knowledge as an internal source of information, and constructs both a coherent propositional representation and a coherent mental model. In the comprehension of visual pictures, the learner selects relevant pictorial information from a drawing, a map, or a graph as an external source of information, organizes the information, activates related prior knowledge as a further source of information, and constructs a coherent mental model complemented by a propositional representation. In the comprehension of auditory pictures (sound comprehension), the learner selects relevant acoustic information, organizes the information, activates related prior knowledge as an internal source of information, and constructs a coherent mental model complemented by a propositional representation.

As shown earlier, the ITPC model is embedded into a broader framework of human cognition, which incorporates

- concepts from semiotics (distinguishing between symbols and icons, or descriptions and depictions, respectively);
- concepts from text processing research (distinguishing between text surface representations, propositional representations, and mental models);
- concepts from picture processing research (distinguishing between visual imagery and mental models);
- concepts from cognitive neuropsychology (distinguishing between phonological and graphemic mental lexicons as well as different kinds of graphemic–phonemic conversion); and
- concepts from memory research combined with general ideas on the human cognitive architecture (multiple memory stores, including the substructure of working memory).

Furthermore, the ITPC model takes the active and constructive nature of comprehension and learning into account. Most important, the model offers a framework for the analysis of text and picture comprehension that makes it possible to explain a broad variety of empirical findings.

Empirical Evidence

In order to demonstrate its validity, the ITPC model should be able to predict under which conditions combinations of text and pictures will be beneficial for learning. However, the model should also be able to predict under which conditions such combinations will have detrimental effects. This part of the chapter analyzes how far the ITPC model is able to successfully predict or explain positive and negative effects of using text and pictures instead of using text alone or pictures alone.

Positive Effects of Combining Text and Pictures

Numerous studies have shown that students usually learn better from words and pictures than from words alone (Levie & Lentz, 1982; Levin, Anglin, & Carney, 1987). This is what Mayer (1997) has called the *multimedia effect* (see Chapter 7). The effect is bound to specific conditions.

Reading skills and prior knowledge. The ITPC model considers text comprehension and picture comprehension to be different routes for constructing mental models and propositional representations using prior knowledge as a third source of information. If one route does not work well or if one source provides little information, the other sources and routes become more important. When learners are poor readers, picture comprehension becomes more important. Thus, the ITPC model predicts that poor readers profit more from illustrations in written texts than good readers. This prediction corresponds to various empirical findings reported by Cooney and Swanson (1987), Levie and Lentz (1982), and Mastropieri and Scruggs (1989).

As text comprehension and picture comprehension are considered to be different routes for the construction of mental representations, the ITPC model also implies the possibility that one route replaces the other one to some extent: pictures can be used instead of a text, and a text can be used instead of pictures. The model therefore predicts that if a picture is added to a text and if the same amount of mental effort is invested in learning, text information becomes less important due to the additional picture information. The text will therefore be processed less deeply, resulting in lower memory for text information than if the text had been processed without pictures. Corresponding findings have been reported by Mayer and Gallini (1990) and by Schnotz and Bannert (1999).

When learners have low prior knowledge, they possess a poor internal source of information. Mental model construction only from written text can become too difficult under these conditions. Adding a picture as another source of information can then considerably enhance comprehension, because it offers an additional route for mental model construction. Learners with high prior knowledge, on the contrary, are able to construct a mental model also without pictorial support. The integrated model therefore predicts that learners with low prior knowledge profit more from pictures in texts than learners with high prior knowledge. This corresponds to the results of various studies which found that pictures in texts are more beneficial for students with low prior knowledge than for those with high prior knowledge (Mayer, 2009; see also Chapter 24).

Redundancy. Contrary to the dual-coding theory, which assumes that adding pictures to text always leads to better learning, because two codes in memory are better than one, the ITPC model predicts that the combination of text and pictures can also have detrimental effects, because high prior knowledge can suspend the multimedia effect. Learners with high prior knowledge frequently do not need both text and pictures as information sources, because one source provides all the information required for mental model construction. In this case, adding a picture to a written text means adding redundant, unneeded information. Although one of the two information sources is not needed, the eye wanders between the two sources, which implies split attention. Thus, the learner loses time and mental effort with processing redundant information without a benefit for learning. This negative effect has been called the *redundancy effect* (Chandler & Sweller, 1996; Sweller, van Merriënboer, & Paas, 1998; see also Chapter 10). This effect implies that a combination of text and pictures that has a positive effect on mental model construction when learners have low prior knowledge may have a negative effect on learning when prior knowledge is high. Experts possibly perform better with only one information source (i.e., text or picture) instead of two (i.e., text and pictures). Corresponding findings have been reported by Kalyuga, Chandler, and Sweller (2000), who have named this the *expertise reversal effect*.

Coherence and contiguity. Students learn better from words and pictures than from words alone, if the words and pictures are semantically related to each other (the *coherence condition*) and if they are presented close together in space or in time (the *contiguity condition*). These findings are explained by the ITPC model in a way that is similar to their explanation by the CTML of Mayer (2009). The ITPC model assumes that a text and a picture can contribute to joint mental model construction only if the text and the picture are semantically related. This corresponds to the coherence condition. The model further assumes that text and picture can contribute to joint mental model construction only if the corresponding text information and picture information are simultaneously available in working memory. As information decays quickly from working memory, this requires the combined presentation of words and pictures as far as possible. This corresponds to the contiguity condition (see Chapter 13).

If a picture is combined with written text, all information has to enter working memory through the visual register. The eye has to switch between pictures and words (i.e., between visual nonverbal feature analysis and graphemic input analysis) so that only one kind of information can be processed at the same time. This split attention implies unproductive search processes from the picture to the text and vice versa, and it affects the simultaneous availability of verbal and pictorial information in working memory (see Chapter 8). When pictures and related written words are presented close to each other (*spatial contiguity*), visual search processes are reduced. Spatial contiguity is a way to minimize the loss of information due to split attention and to allow an approximately simultaneous availability of pictorial and verbal information in working memory. In other words, spatial contiguity is a means to maximize temporal contiguity in working memory under the condition of a picture with written text. Fully simultaneous availability, however, can be ensured only when a picture is combined with auditory text, because pictorial and verbal information can then be processed at the same time (*temporal contiguity*) and be kept simultaneously in working memory. In this case, no split attention is required because learners can devote their full visual attention to the picture and their full auditory attention to the text (Mousavi, Low, & Sweller, 1995). This has led to the assumption of a modality effect.

Modality. According to the *modality effect*, students learn better from multimedia instructional messages when text is spoken rather than written (Ginns, 2005; Mayer & Moreno, 1998; Moreno & Mayer, 1999; see also Chapter 9). The modality effect is a derivative of the multimedia effect, because the rationale behind the modality effect is to take full advantage of text–picture combinations (i.e., of the multimedia effect) by maximizing the contiguity of verbal and pictorial information or by minimizing any obstacles to the simultaneous availability of verbal and pictorial information in working memory, respectively. The key to minimizing the obstacles and to maximizing contiguity is the combination of auditory presentation of text and visual presentation of pictures. As the modality effect is a derivative of the multimedia effect, a modality effect is to be expected only if there is also a multimedia effect. If there is no multimedia effect, no modality effect is to be expected either.

Currently, there is no straightforward answer to the question of where the modality effect comes from. The most popular explanation is the avoidance of split attention, as already mentioned (Leahy, Chandler, & Sweller, 2003; Mayer & Moreno, 1998; Mousavi, Low, & Sweller, 1995). Split attention is indeed a fundamental problem when written text is combined with animation: as soon as the learner reads some text, he or she is at risk of missing important pictorial information, which can be avoided by using spoken text. Besides split attention, Moreno and Mayer (1999) have argued for an additional explanation of the modality effect. They presented text and pictures to learners in a consecutive way and thus avoided split attention. Nevertheless, spoken text with pictures resulted in better learning than written text with pictures. The authors argued that part of the modality effect results from the amount of working memory capacity involved. Text comprehension and picture comprehension are enhanced if both visual memory and auditory working memory are involved, even if the two systems receive their input only in a consecutive manner. Although this explanation seems to be plausible, the ITPC model does not support this assumption because both the comprehension of spoken text and the comprehension of written text involve auditory working memory. Research findings suggest that even experienced readers engage in graphemic–phonemic lexical conversion and recode at least parts of the visual information into auditory information (Ellis & Young, 1996; Rieben & Perfetti, 1991). Similarly, Baddeley (1999) assumes that verbal information – either spoken or written – is generally processed in the phonological loop rather than the visuospatial sketchpad. Rummer, Schweppe, Fürstenberg, Seufert, and Brünken (2010) have suggested an *auditory-recency explanation* of the modality effect when text material consists of single sentences presented alternately with pictures. The authors argue that due to the longer duration of acoustic information in the auditory register than in the visual register, a sentence can be better maintained in working memory after it has been heard than after it has been read. Last but not least, a modality effect can be due to learners' literacy. Auditory language is ubiquitous, whereas mastering written language requires educational effort. It is possible that an illiterate person can understand auditory text with pictures but is unable to read the corresponding written text. This will result in a strong (and absolutely trivial) modality effect.

It seems that the modality effect does not result from a unitary set of causal relationships. Instead, findings suggest that heterogeneous factors lead to similar outcomes due to rather different processing mechanisms (Schnotz, 2011; Schüler, Scheiter, & Schmidt-Weigand, 2011). The ITPC model is in agreement with the split-attention explanation and with the auditory-recency explanation of a modality effect, whereas it does not agree with an explanation based on increased working memory capacity. Similar to the multimedia effect, which is counteracted by the redundancy effect as a *reversed multimedia effect*, the ITPC model can also predict a *reversed modality effect* (i.e., written text with pictures can be better for learning than spoken text with pictures), which counteracts the regular modality effect under specific conditions. Written text provides more control of cognitive processing. Readers can pause or slow down their reading, or reread difficult passages, and in this way adapt their perceptual processing to the needs of their cognitive processing, which is much more difficult or impossible with spoken text. Thus, if a text is difficult to understand and if the accompanying picture is neither animated nor too complex and if learning time is not severely limited, the ITPC model would predict a reversed modality effect, namely better learning with pictures accompanied by written text than by spoken text. This is in line with recent research indicating that the modality effect occurs only under specific conditions (Gyselinck, Jamet, & Dubois, 2008; Leahy, Chandler, & Sweller, 2003; Stiller, Freitag, Zinnbauer, & Freitag, 2009).

Interference Effects in Combining Texts and Pictures

Sequencing. Sometimes a picture is too large and too complex to be presented simultaneously with corresponding text. In this case, strict contiguity is hard to achieve. Instead, the picture has to be presented either before or after the text. Various studies have shown that it is better to present a picture before a corresponding text than after the text (Kulhavy, Stock, & Caterino, 1994). Eitel, Scheiter, and Schüler (in press) have recently demonstrated with the help of eye tracking that even a very short (less than 2 seconds) presentation of a picture can have a scaffolding function in mental model construction. The ITPC model explains this scaffolding function by the direct access of pictures to mental model construction through the mapping of analog structures in the depictive subsystem, whereas text comprehension has to make a detour through the descriptive subsystem. The sequencing effect is explained by the ITPC model through the inherent ambiguity of text. A text never describes the subject matter in enough detail to fit just one picture or one mental model. Instead, it allows some degrees of freedom for pictures and for mental model construction. If a mental model is constructed only from a text, the model will therefore most likely differ to some extent from a picture presented to illustrate the subject matter, even if it fully corresponds to its verbal description. Thus, if the picture is presented after the text, the picture will most likely interfere with the previously text-based constructed mental model. Such interference is avoided when the picture is presented before the text even if the learner looks only briefly at the picture to benefit from its mental model scaffolding function.

Verbal redundancy across modalities. Multimedia designers frequently try to adapt to the needs of individual learners, who are assumed to prefer either spoken text or written text. They therefore present pictures simultaneously with both written text and spoken text. Learners are in this way supposed to choose their preferred sensory modality: those who prefer to listen can focus on the spoken text, and those who prefer to read can focus on the written text. However, the ITPC model predicts that individuals do not learn better from pictures accompanied by spoken *and* by written text, but that they learn better from pictures combined with either *only spoken* or *only written* text. The model provides two reasons for this prediction. The first reason is that even if the same text is presented in an auditory manner, it is difficult for learners to ignore a simultaneously presented written text. Thus, the presentation of a picture combined with a written text will result in split visual attention, despite the simultaneous auditory presentation of the same text. The second reason is a problem of synchronization between listening and reading. Skilled readers are often able to read a text faster than the auditory text is spoken. When they create (based on graphemic–phonemic lexical conversion) inner speech that they can hear with their inner ear, interference between external listening and reading (i.e., internal listening) is likely to occur. Various studies of Mayer and his co-workers have demonstrated that the performance of individuals who learn from pictures accompanied by spoken and written text is poorer than that of individuals who learn from pictures and only spoken text (Mayer, 2009; see also Chapter 10).

Structure mapping. The same subject matter can often be visualized in different ways. Contrary to the dual-coding theory (Paivio, 1986), which assumes that verbal and pictorial coding is generally better for learning than single coding, the ITPC model considers the form of visualization an important predictor of a multimedia effect. Pictures are beneficial for learning only if task-appropriate forms of visualization are used, whereas they are harmful in the case of task-inappropriate forms of visualization. This prediction derives from the assumption that pictures are processed in the depictive subsystem by structure mapping. This implies that the form of visualization is mapped onto the structure of the mental model. Accordingly, the ITPC model predicts that the efficiency of a mental model for a specific task corresponds to the picture's efficiency for this task (Larkin & Simon, 1987). Corresponding empirical findings were reported by Schnotz and Bannert (2003), who studied learning from text combined with different pictures when the pictures were informationally equivalent but used different forms of visualization. The authors found that pictures enhanced comprehension only if the learning content was visualized in a task-appropriate way. If the learning content was visualized in a task-inappropriate way, the pictures interfered with the construction of a task-appropriate mental model. Thus, well-designed pictures are not only important for learners with low prior knowledge who need pictorial support for mental model construction. They are also important for learners with high prior knowledge, because mental model construction can be negatively affected by inappropriate forms of visualization.

Cognitive Economy

The ITPC model finally provides a framework for considerations of cognitive economy in learning from multiple external representations, especially from texts and pictures (see Chapter 20). Multiple external representations support comprehension because each representation both constrains and elaborates the interpretation of other representations. However, an understanding of each representation also creates cognitive costs. In the case of understanding multiple texts and pictures, the benefits and the costs of processing an information source depend on the ease or difficulty of using the corresponding sensory and representational channels. When more and more representations about one topic are processed, it is possible that the additional benefit for comprehension is not worth the additional cognitive costs. If the benefits from processing an additional information source are lower than the required costs, the learner will follow the principle of cognitive economy and will not engage in further cognitive processing. Instead, the learner will consider only some representations and ignore the other ones. This could explain why individuals in self-directed learning frequently ignore information sources. This finding has been reported repeatedly in research on learning from multiple representations (Ainsworth, 1999; Sweller, van Merriënboer, & Paas, 1998).

According to the ITPC model, the benefits of combining text with pictures (the multimedia effect) is not due to the superiority of dual versus single coding of information. Instead, because text is first processed in the descriptive subsystem followed by the depictive subsystem, whereas pictures are first processed in the depictive subsystem followed by the descriptive subsystem, text and pictures are assumed to have fundamentally different functions in comprehension and learning. Hochpöchler et al. (2013) found in an eye-tracking study of text–picture integration that processing was primarily text-driven during an initial phase of mental model construction, while brief looks at the accompanying picture indicated that pictures were used only for some scaffolding of the initial mental model. After initial mental model construction, on the contrary, the text was merely used for task-specific model updates, whereas the picture was now used intensively, depending on the task at hand, as an easily accessible visual tool. In other words, text processing was less task-dependent than picture processing. It seems that texts guide the reader's conceptual analysis systematically by describing the subject matter step by step, whereas pictures function as external cognitive tools that can be used on demand as a substitute for the subject matter.

Instructional Implications

What does the ITPC model contribute to instructional design? The model suggests various guidelines for instructional design that focus on the use of text and pictures in multimedia learning environments. Some guidelines correspond to those derived from the cognitive theory of multimedia learning developed by (Mayer, 2009; see also Chapter 3). Other guidelines go beyond the suggestions of CTML, and some further guidelines make contradictory suggestions. A fundamental commonality between the ITPC model and CTML is that both reject simple rules of thumb, such as the suggestion to use multiple forms of representations and multiple sensory modalities whenever possible. Instead, both views agree that instructional design for multimedia learning should be guided by sufficient understanding of human perception and human cognitive processing based on careful empirical research. The ITPC model suggests the following guidelines for instructional design:

- Conditional use of multimedia. Use text combined with content-related pictures when learners have low prior knowledge but sufficient cognitive abilities to process both the text and the pictures (see Chapters 7 and 24).
- Text–picture coherence. Use pictures only when they are semantically clearly related to the content of the text (see Chapter 13).
- Spatial and temporal contiguity. If written text is used, present it in close spatial proximity to the picture. If spoken text is used, present it in close temporal proximity to the picture (see Chapter 13).
- Avoidance of redundancy. Do not combine text and pictures if learners have sufficient prior knowledge and cognitive ability to construct a mental model from one source of information, as the other source would be redundant for them (see Chapter 10).
- Text modality for animated pictures. When animations are combined with text, use spoken text instead of written text due to the fluent nature of the animation in order to avoid split attention (see Chapters 9 and 22).
- Text modality for static pictures. When static pictures are used and learning time is not limited, split attention becomes less important. In this case, one should balance the advantage of auditory text (i.e., avoidance of split attention), which predicts a positive modality effect, against the possible advantage of written text (i.e., higher control of cognitive processing), which predicts a reversed modality effect. If the text is difficult to understand, learning time is not limited, and picture complexity is low, use written text rather than spoken text (see Chapters 8 and 9).
- Verbal redundancy across modalities. Do not add written text that duplicates spoken text combined with pictures (see Chapters 13 and 20).
- Sequencing. Do not present a text that is semantically related to a picture before the picture can be observed by the learner.
- Structure mapping. If the subject matter can be visualized by different pictures in different ways that are informationally equivalent, use a picture with the form of visualization that is most appropriate for solving future tasks.

A general message that emerges from these suggestions is that designers of instructional material should resist the temptation to add irrelevant bells and whistles to multimedia learning environments. Simply speaking, less can be more.

Limitations of the Integrated model and Directions for Future Research

Despite its relative complexity, the integrated model still simplifies things considerably and therefore needs further elaboration. For example, there might be multiple levels of propositional representations instead of only one level, ranging from micro-propositions (i.e., very detailed descriptions) to various levels of macro-propositions (i.e., more course-grained descriptions) based on macro-operations (van Dijk, 1980; van Dijk & Kintsch, 1983). Similarly, there might be multiple levels of mental models, ranging from coarse-grained overview models to detailed partial models of high granularity. Furthermore, the interaction between the descriptive subsystem and the depictive subsystem might occur not only between propositions and a mental model, as shown in Figure 4.3. When learners are highly familiar with a domain, mental models can also be constructed directly from phonological or graphemic input without a propositional detour (Perfetti & Britt, 1995). Similarly, it is possible to create a proposition directly from a perceptual representation of a visual picture without a mental model. These "shortcuts" are not included in Figure 4.3.

Another aspect not included in the ITPC model is that learning from text and pictures requires not only an understanding of the verbal and pictorial information, but also knowledge of where each kind of information can be found. In multimedia environments, texts and pictures are frequently distributed across a complex nonlinear hyperspace. In this case, the learner has to construct not only a mental model of the learning content, but also a mental model of the hyperspace.

More research is also needed to predict more precisely under which conditions the combination of text and pictures is beneficial and under which circumstances it is harmful for learning. In other words, the relative strengths of the different effects under different conditions need further specification. The effects of combining text and pictures can be considered to be a result of the different levels of efficiency of perceptual processing and cognitive processing under specific external and internal conditions of processing. External conditions include, for example, the structure and content of the written or spoken text, text–picture coherence, text–picture redundancy, contiguity of text–picture presentation, time constraints, and learning objectives. Internal conditions include, for example, prior knowledge, cognitive abilities, and individual preferences. Corresponding studies should also estimate the relative size of the various effects for different types of texts and for different forms of visualization in different domains.

The ITPC model deals only with perceptual and cognitive processing of texts and instructional pictures. However, most learning material also includes decorative pictures that are expected to make the material aesthetically pleasing or perhaps to remove some pressure from the learning situation (Pozzer & Roth, 2003; Takahashi, 1995). Because these pictures provide little information about the learning content, they cannot contribute much to mental model construction directly. Instead, they can be suspected to distract the learner's attention and act therefore as an impediment to learning (cf. Harp & Mayer, 1998; Sanchez & Wiley, 2006). However, Lenzner, Schnotz, and Müller (2013) found that decorative pictures captured very little attention. However, they induced a better mood, alertness, and calmness in the learner, which in turn could be assumed to enhance more concentrated cognitive processing. Decorative pictures moderated the beneficial effect of instructional pictures on learning. Instructional pictures combined with decorative pictures were more beneficial for learning than those without decorative pictures. When learners had low prior knowledge, the *combined* cognitive effect of instructional pictures and affective impact of decorative pictures led to especially successful learning. Further research is needed to clarify this issue.

Future elaborations of the model should address learners' strategies of selecting relevant verbal or pictorial information and of giving special emphasis to specific mental representations according to the aims of learning. As far as learners follow the principle of cognitive economy in knowledge acquisition, the efficiency of the different paths for constructing mental representations is a central concept for the analysis of strategic self-directed learning. Further research should investigate to what extent individuals follow this principle in learning from text and pictures. Individuals may prefer descriptive information processing to depictive processing. For example, so-called verbalizers are assumed to prefer verbal information, whereas so-called visualizers prefer pictorial information (Kirby, Moore, & Schofield, 1988; Plass, Chun, Mayer, & Leutner, 1998). Future research should also analyze whether there are preferences with regard to the visual or the auditory modality in multimedia learning.

The ITPC model of text and picture comprehension provides a framework for the analysis of learning from multiple representations including spoken or written text, visual pictures, and sound pictures. It is embedded in a broader framework of human cognition and incorporates concepts from various disciplines of cognitive science. The model aims at contributing to a deeper understanding of learning from text and pictures and to enable better-informed decisions in instructional design. Both aims require a balancing act between complexity and simplicity. The graphical representation of the ITPC model in Figure 4.3 may, on the one hand, be viewed as relatively complex in terms of the kinds of processing entailed and in terms of the products of processing as compared with other models of comprehending text and pictures. However, it can also be viewed as an oversimplification of the subject matter. For example, because the model deals primarily with comprehension, it takes into account the graphemic and the phonological input lexicon, but not the phonological output lexicon (which includes the motor patterns for producing speech sounds), although it refers to the possibility of inner speech in reading heard by readers with their inner ear. The maxim of making things as simple as possible but not simpler (attributed to Einstein) is a special challenge in research on multimedia learning. Future research will clarify whether the ITPC model is a useful tool for the analysis of text–picture integration.

Glossary

Cognitive deep structure processing:	Semantic processing of verbal and pictorial information in working memory, resulting in propositional representation and mental models.
Cognitive economy:	A principle of cognitive processing that tries to meet cognitive aims with a minimum of cognitive effort.
Coherence condition:	A condition of the multimedia effect, which corresponds to high semantic relatedness between text and picture.
Contiguity condition:	A condition of the multimedia effect, which corresponds to close proximity of text and picture in space or time.
Cross-modality verbal redundancy:	The use of written text that duplicates spoken text combined with pictures.
Depictive representation:	A form of representation that uses iconic signs (such as visual pictures) to show characteristics of the subject matter.
Descriptive representation:	A form of representation that uses symbols (such as natural language) to describe characteristics of the subject matter.
Grapheme–phoneme conversion:	The nonlexical conversion of letter strings into phoneme strings.
Graphemic input analysis:	The identification of graphemes within visual verbal input.
Graphemic–phonemic lexical conversion:	The lexicon-based conversion of whole-word letter strings into whole-word phoneme strings.
Integrated model of text and picture comprehension (ITPC model):	A model of how individuals understand text and pictures presented in different sensory modalities, based on the assumption that the human perceptual system includes multiple sensory channels, whereas the cognitive system includes two representational channels – a verbal (descriptive) channel and a pictorial (depictive) channel –and that these channels have a limited capacity for information processing and active coherence formation.
Listening comprehension:	The construction of propositional representations and mental models based on spoken text.
Mental model:	A mental representation of the subject matter by an internal structure that is

	analogous to the subject matter.
Modality effect:	Students learn better from text and pictures if the text is presented as spoken rather than as written text, mainly because of avoidance of visual split attention, if specific conditions are met.
Multimedia effect:	Students learn better from text and pictures than from text alone.
Parsing:	Syntactic-semantic analysis of spoken or written sentences (i.e., segmentation of word strings) with regard to their constituent structure.
Perceptual surface structure processing:	The transfer of information from the surface structure of texts and pictures to working memory, encompassing phonological or graphemic verbal information or visual or acoustic pictorial information.
Phonemic input analysis:	The identification of phonemes within acoustic verbal input.
Picture–text sequencing principle:	If a written text and a picture cannot be presented simultaneously, the picture should be presented before the text instead of after the text.
Propositional representation:	A mental representation of ideas expressed in a text or in a picture without reference to specific words and phrases.
Reading comprehension:	The construction of propositional representations and mental models based on written text.
Redundancy:	The combination of texts and pictures when learners have sufficient prior knowledge and cognitive ability to construct a mental model from one source only, which makes the additional information source redundant for the learner and creates a reversed multimedia effect.
Reversed modality effect:	Students learn better from text and pictures if the text is presented as written rather than as spoken text (mainly because written text provides more control of cognitive processing than spoken text) if the text is difficult to understand and if the accompanying picture is neither animated nor too complex and if learning time is not severely limited.
Sensory register:	A memory store that holds information from a specific sensory modality (e.g., the eye or the ear) for a very short time as a basis for further information processing.
Sound comprehension (auditory picture comprehension):	The construction of mental models and propositional representations based on sounds (as auditory pictures).
Split attention:	The use of one information channel for different sources of information.
Structure mapping:	The transfer of a structure consisting of elements and relations between the elements onto another structure with different elements but the same relations.
Text surface representation:	A mental representation of a text, including exact wording and syntax structure.
Visual picture comprehension:	The construction of mental models and propositional representations based on visual pictures (such as drawings, maps, or graphs).
Working memory:	A memory store that holds and manipulates information that is in the

individual's focus of attention, including a visual store, an auditory store, a propositional store, and a spatial mental model store.

References

Adams, B. C., Bell, L., & Perfetti, C. (1995). A trading relationship between reading skill and domain knowledge in children's text comprehension. *Discourse Processes*, 20, 307–323.

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, 33, 131–152.

Atkinson, C., & Shiffrin, R.M. (1971). The control of short-term memory. *Scientific American*, 225, 82–90.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Baddeley, A. D. (1999). *Essentials of human memory*. Hove: Psychology Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science*, 4, 417–423.

Caramazza, A., Berndt, R. S., & Basili, A. G. (1983). The selective impairment of phonological processing: A case study. *Brain and Language*, 18, 128–174.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, 14, 5–26.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, 10, 151–170.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256.

Comenius, J. A. (1999). *Orbis sensualium pictus* [Facsimile of the 1887 edition]. Whitefish, MT: Kessinger.

Cooney, J. B. & Swanson, H. L. (1987) Memory and learning disabilities: An overview. In H. L. Swanson (Ed.), *Memory and learning disabilities: Advances in learning and behavioral disabilities* (pp. 1–40). Greenwich, CT: JAI.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9, 561–583.

Dutke, S. (1996). Generic and generative knowledge: Memory schemata in the construction of mental models. In W. Battmann & S. Dutke (Eds.), *Processes of the molar regulation of behavior* (pp. 35–54). Lengerich: Pabst Science.

Eitel, A., Scheiter, K., & Schüler, A. (in press). The time course of information extraction from instructional diagrams. *Perceptual and Motor Skills*. doi: 10.2466/22.23.PMS.115.6.

Ellis, A. W., & Young, A. W. (1996). *Human cognitive neuropsychology: A textbook with readings*. Hove: Psychology Press.

Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General*, 129, 61–83.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning & Instruction*, 15, 313–331.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of*

Psychology, 48, 163–189.

Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory components in multimedia comprehension. *Applied Cognitive Psychology*, 22, 353–374.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414–434. doi:10.1037/0022-0663.90.3.414.

Hochpöchler, U., Schnotz, W., Rasch, T., Ullrich, M., Horz, H., McElvany, N., Schroeder, S., & Baumert, J. (2013). Dynamics of mental model construction from text and graphics. *European Journal of Psychology of Education,* 28(4), 1105–1126 doi: 10.1007/s10212-012-0156-z.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Cambridge University Press.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, 92, 126–136.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.

Kirby, J. R., Moore, P. J., & Schofield, N. J. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology*, 13, 169–184.

Knauff, M., & Johnson-Laird, P. (2002). Visual imagery can impede reasoning. *Memory & Cognition*, 30, 363–371.

Kosslyn, S. M. (1994). *Image and brain*. Cambridge, MA: MIT Press.

Kulhavy, R.W., Stock, W. A., & Caterino, L. C. (1994). Reference maps as a framework for remembering text. In W. Schnotz & R. W. Kulhavy (Eds.), *Comprehension of graphics* (pp. 153–162). Amsterdam: Elsevier Science.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11, 65–99.

Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology*, 17, 401–418.

Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science,* 41(5), 811–831. doi: 10.1007/s11251-012-9256-z.

Levie, H. W., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology Journal*, 30, 195–232.

Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton, (Eds.), *The psychology of illustration* (vol. 1, pp. 51–86). New York: Springer.

Lowe, R. K. (1996). Background knowledge and the construction of a situational representation from a diagram. *European Journal of Psychology of Education*, 11, 377–397.

Mastropieri, M. A., & Scruggs, T. E. (1989). Constructing more meaningful relationships: Mnemonic instruction for special populations. *Educational Psychology Review*, 1, 83–111.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32, 1–19.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning* (2d ed.). New York: Cambridge University Press.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, 82, 715–726.

Mayer, R. E., & Moreno R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90, 312–320.

McNamara, D. S. (Ed.) (2007). *Reading comprehension strategies: Theories, interventions, and technologies*. New York: Lawrence Erlbaum.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1–43.

Miller, L. M. S., & Stine-Morrow, E. A. L. (1998). Aging and the effects of knowledge on on-line reading strategies. *Journal of Gerontology: Psychology Sciences*, 53B, 223–233.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, 91, 358–368.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by minimizing auditory and visual presentation modes. *Journal of Educational Psychology*, 87, 319–334.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance* (vol. 9, pp. 135–151). Hillsdale, NJ: Lawrence Erlbaum.

Peirce, C. S. (1931–1958). *Collected writings*, 8 vols. (Ed. C. Hartshorne, P. Weiss & A. W Burks). Cambridge, MA: Harvard University Press.

Perfetti, C. A., & Britt, M. A. (1995). Where do propositions come from? In C. A. Weaver III, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 11–34). Hillsdale, NJ: Lawrence Erlbaum.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, 90, 25–36.

Pozzer, L. L., & Roth, W.-M. (2003). Prevalence, function and structure of photographs in high school biology textbooks. *Journal of Research in Science Teaching*, 40(10), 1089–1114. doi:10.1002/tea.10122.

Rieben, L., & Perfetti, C. (1991). *Learning to read: Basic research and its implications*. Hillsdale, NJ: Lawrence Erlbaum.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.

Rummer, R., Schweppe, J., Fürstenberg, A., Seufert, T., & Brünken, R. (2010). Working memory interference during processing texts and pictures: Implications for the explanation of the modality effect. *Applied Cognitive Psychology*, 24, 164–176.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, 34(2), 344–355.

Schnotz, W. (2011). Colorful bouquets in multimedia research: A closer look at the modality effect. *Zeitschrift für Pädagogische Psychologie*, 25, 269–276.

Schnotz, W., & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Bild- und Textverstehen [Effects of the visualization form on the construction of mental models in picture and text comprehension]. *Zeitschrift für Experimentelle Psychologie*, 46, 216–235.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction*, 13, 141–156.

Schüler, A., Scheiter, K., & Schmidt-Weigand, F. (2011). Boundary conditions and constraints of the modality effect. *Zeitschrift für Pädagogische Psychologie*, 25, 211–220.

Sims, V. K., & Hegarty, M. (1997). Mental animation in the visuospatial sketchpad: Evidence from dual-tasks studies. *Memory & Cognition*, 25, 321–332.

Soederberg Miller, L. M. (2001). The effects of real-world knowledge on text processing among older adults. *Aging, Neuropsychology and Cognition*, 8, 137–148.

Stiller, K. D., Freitag, A., Zinnbauer, P., & Freitag, C. (2009). How pacing of multimedia instructions can influence modality effects: A case of superiority of visual texts. *Australasian Journal of Educational Technology*, 25, 184–203.

Sweller, J., van Merriënboer, J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychological Review*, 10, 251–296.

Takahashi, S. (1995). Aesthetic properties of pictorial perception. *Psychological Review*, 102(4), 671–683. doi:10.1037/0033-295X.102.4.671.

Vallar, G., & Shallice, T. (Eds.) (1990). *Neuropsychological impairments of short-term memory*. Cambridge: Cambridge University Press.

van Dijk, T. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, NJ: Lawrence Erlbaum.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

van Oostendorp, H., & Goldman, S. R. (Eds.) (1999). *The construction of mental representations during reading*. Hillsdale, NJ: Lawrence Erlbaum.

Weaver, C. A., III, Mannes, S., & Fletcher, C. R. (1995). *Discourse comprehension. Essays in honor of Walter Kintsch*. Hillsdale, NJ: Lawrence Erlbaum.