Danielle Albers Szafir · Rita Borgo · Min Chen · Darren J. Edwards · Brian Fisher · Lace Padilla *Editors*

Visualization Psychology

Springer

Visualization Psychology

Danielle Albers Szafir • Rita Borgo • Min Chen •
Darren J. Edwards • Brian Fisher • Lace Padilla
Editors

Visualization Psychology

Springer

Editors
Danielle Albers Szafir
University of Colorado Boulder
Boulder, CO, USA

Rita Borgo
Department of Computer Science
King's College London
London, UK

Min Chen
OeRC, Department of Engineering Science
University of Oxford
Oxford, UK

Darren J. Edwards
Department of Public Health
Swansea University
Swansea, UK

Brian Fisher
School of Interactive Arts & Technology
Simon Fraser University
Surrey, BC, Canada

Lace Padilla
Northeastern University
Boston, USA

ISBN 978-3-031-34737-5 ISBN 978-3-031-34738-2 (eBook)
https://doi.org/10.1007/978-3-031-34738-2

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Foreword

Visualization Psychology is enormously diverse. Most of what is known about perception and cognition applies to some degree, and this knowledge is relevant to a huge diversity of visualization methods ranging from conventional scatter plots to people collaborating to interpret scientific data in a shared virtual reality workspace.

This volume provides an excellent introduction to this diversity, with cutting edge research and theory across the breadth of the field. To pick a few examples: We have basic perceptual research into what makes a color express a greater quality by Karen Schloss and her collaborators. The broader cognitive processes of Sense-Making is introduced by Margaret Pohl and collaborators and a more focused introduction to the importance of cognitive processes in graph comprehension provides depth. We have introductions to other to bodies of theory such as the cognitive processes of the visualization designers themselves by Paul Parsons and educational theory by Stoiber and collaborators. Enjoy the feast!

There was a time in the 1990 when visualization psychology did not exist and inventions were celebrated, with little attention paid to whether they worked; I recall that the hottest topics in visualization were a tree structure visualization invention called the Cone Tree, the CAVE virtual reality viewing environment, and line integral convolution (LIC) for visualizing flows in liquids and gasses. None of these turned out to be useful, and this could have been predicted with a little insight into the perceptual and cognitive issues. ConeTrees because of the mental gymnastics required for interaction; CAVES for many reasons including occlusion, vergence-focus conflict, poor interaction affordances, and lack of resolution at the critical fovea (it is not surprising that HMDs now dominate); and LIC because it provides a poor stimulus for orientation detectors and lacks perceptual cues for showing speed effectively.

We can only avoid such costly mistakes if the discipline of data visualization is grounded in both evaluation and psychological theory. And, although a careful evaluation can usually avoid egregious mistakes in design, evaluation without theory only applies after the fact. The proper application of perceptual and cognitive theory can inform visualization design from the outset.

We are entering a new Age of Visualization. The massive data sets being generated in every field of human endeavor can often only be understood with visualization. Who can comprehend a table with a thousand numbers? But it can be easy to comprehend a thousand data points represented graphically. Companies such as Tableau employ perceptual and cognitive scientists to ensure that their products present data in ways that are clear and not misleading. One good thing that has come from the COVID disaster is a huge growth in public facing visualizations; news websites now show maps, times series plots, and sometimes complex network diagrams. *Visualization Psychology* provides the theoretical underpinnings of effective visualization design and this book provides a snapshot of the current state of the art.

January 2022 Colin Ware

Research Professor, University of New Hampshire, New Hampshire, USA
Author of *Information Visualization: Perception for Design*, Morgan Kaufmann
and *Visual Thinking for Design*, Morgan Kaufmann

Preface

Data visualization emerged as an academic subject in 1987 following the NSF Panel Report on “Visualization in Scientific Computing” edited by Bruce H. McCormick, Thomas A. DeFanti, and Maxine D. Brown. For several decades, building strong connections between visualization and psychology has always been a research agenda in the field of Visualization and Visual Analytics (VIS in short). Many called for interdisciplinary research between VIS and psychology (e.g., “Information Visualization, Wings for the Mind,” Stuart Card, 1995, and “Illuminating the Path: The Research and Development Agenda for Visual Analytics,” James J. Thomas and Kristin A. Cook, 2005). Several psychologists have exerted most valuable influence on VIS (e.g., “Visual Thinking for Design,” Colin Ware, 2008).

However, the progress for building connections between VIS and psychology has not been as rapid as many other advancements in either field. Before 2010, each VIS conference typically featured 0-2 papers on empirical studies. The VisWeek 2010 in Salt Lake City became a turning point, and since then more and more empirical study papers have been presented at VIS. Between 2016 and 2019, there were some 60 empirical study papers in VIS/TVCG tracks. Many young talents who are knowledgeable in both VIS and psychology emerged in the VIS community, while many colleagues in psychology are authoring and co-authoring such papers and attending VIS conferences. It is therefore timely to ask both VIS and psychology communities: Is there a need for *Visualization Psychology* as a new interdisciplinary subject?

There are many branches of applied psychology, such as *clinical psychology*, *counselling psychology*, *educational psychology*, *forensic psychology*, *health psychology*, *industrial-organizational psychology*, *legal psychology*, *media psychology*, *music psychology*, *occupational psychology*, *sports psychology*, and so on. Almost all of these are widely recognized academic subjects and have their own conferences and journals. Since interactive visualization and visual analytics are activities most commonly encountered in human-centric processes in data science and real-world data intelligence workflows, many will argue for the necessity and feasibility for developing *Visualization Psychology*—as a branch of applied psychology—in a coherent and organized manner.

There has been existing activities for empirical research during VIS conferences, noticeably, the BELIV workshop series and the VISxVISION events. The BELIV workshops, as the name suggests, have focused on the "evaluation" of visualization methods and techniques, and there has been a strong emphasis on "beyond" the traditional controlled experiments. Meanwhile, the VISxVISION events have been successful in bridging to the vision science community, but as the name suggests, the scope of VISxVISION cannot easily cover the engagement with scientists and researchers with expertise in higher-order cognition (including topics such as analytical reasoning, problem-solving, and collaborative cognition) in studying complex phenomena in VIS processes. Both series of events are no doubt important to the development of VIS as a scientific discipline, while stimulating more interdisciplinary and empirical research.

Advanced data intelligence workflows likely involve both human-centric processes (e.g., visualization and interaction) and machine-centric processes (e.g., statistics and algorithms). Such workflows feature a diverse range of cognitive activities. Numerous phenomena in these processes cannot easily be explained using the existing theories and experiments in VIS and psychology, including some of the most fundamental questions such as "since visualization is not as precise as the data being depicted, what is visualization really for, and how visualization works?" Being able to answer such fundamental questions and explain numerous real-world phenomena in VIS processes is critical to VIS and data science as well as psychology. As VIS techniques are for augmenting human cognition, we must develop VIS techniques by building on the theoretical, empirical, and methodological knowledge that has already been acquired in psychology. At the same time, the field of VIS is a rich playground for discovering new knowledge relevant to both VIS and psychology.

VisPsych

The first IEEE VIS Workshop on Visualization Psychology took place during IEEE VIS2020, providing a venue for the experts in VIS and psychology to define the scope of this new subject of *Visualization Psychology* collectively, and stimulate new research directions and activities in both fields. The logo of the workshop features the abbreviations of "Vis" and "Psych" connected by one of the most popular continuous color maps in visualization. The goals of the workshop were:

- To broaden the scope of empirical research in VIS to involve more cognitive aspects in addition to considering visualization a vision or perception problem
- To provide researchers in VIS with a significant platform to develop their theories and experiments in addition to acquiring knowledge from psychology
- To enable researchers in psychology to explore VIS as a rich playground and carry out research beyond the existing molds

- To enable the development of the young talents in VIS and psychology through the development of a new interdisciplinary subject and the provision of a platform for research communication, publications, and collaboration

This book results from the initiative taken by the VisPsych workshop. The attendees of the workshop were motivated by the aforementioned goals and enthused by the technical developments and outlooks presented in the workshop. Many offered to transform their preliminary ideas, viewpoints, and findings to chapters to be included in this book. After some two years of enormous effort and great endurance, these authors produced the wonderful scholarly work featured in this book, which consists of 15 chapters organized into 3 parts:

- Part I—Visualization Psychology from a Psychology Perspective—contains five chapters that examine aspects of psychology, including existing theories, findings, and methodologies, and discuss how such acquired knowledge (e.g., findings on color semantics, process theories for graph comprehension, theories for mental models, and dual-processing models in decision-making) help understand and interpret phenomena in visualization or such established best practice (e.g., the diversity of research methods) can influence the development of *Visualization Psychology*.
- Part II—Visualization Psychology from a Visualization Perspective—contains five chapters, each of which focuses on an important topic in VIS and makes connections with aspects of psychology. The selected visualization topics include visualization literacy, visualization of health information, the cognition of visualization designers, and understanding eye tracking data captured in visualization processes. The discourses presented show that not only these topics can benefit from the previous work in psychology, but can also inspire researchers in *Visualization Psychology* to make new discoveries that are scientifically significant and practically useful. One chapter in this Part presents a coherent argument that the field of VIS is a fertile laboratory for exploring human cognition, while VIS research and VIS system development can be grounded in theories of perception and cognition.
- Part III—Visualization Psychology from an Experimental Perspective—contains five chapters, presenting a collection of experimental findings on several topics, including visualization tasks, perceptual biases, design preferences, uncertainty visualization, and sensemaking strategies. Through structured literature reviews, categorized descriptions, and analytical discourse, these chapters demonstrate that there is an abundance of intriguing and complex phenomena in visualization processes, which cannot easily be explained by the known theories and experiments in either VIS or psychology, but can benefit from further interdisciplinary research in a new subject *Visualization Psychology*.

If the subject of *Visualization Psychology* were a landscape to be painted collectively by the scientists and researchers in VIS and psychology, the process of painting this landscape has just begun. This book would not be in any way a piece of finished work. It would be better described as a number of earnest and thoughtful strokes brushed onto the canvas by a diverse group of authors attempting to sketch out some major components of the landscape. No doubt, we need many more scientists, researchers, and practitioners to join this long-term effort. The landscape would gradually but surely unveil itself with every new stroke resulting from future research in *Visualization Psychology*.

We are hugely grateful to all authors of the 15 chapters in this book, and appreciate their scientific and scholarly discourse as well as collaborative and enduring effort in completing their ambitious writing plans. In addition, we value tremendously the contributions made by other authors who submitted their papers to the VisPsych workshop.

We would like to record our enormous gratitude to all members of the Program Committee of the VisPsych workshop, including Alfie Abdul-Rahman, Nadia Boukhelifa, Spencer Castro, Michael Correll, Evanthia Dimara, Kristin M. Divis, Sarah Dryhurst, Madison Elliott, Steve Haroz, Lane Harrison, Kuno Kurzhals, Bongshin Lee, Laura Matzen, Christine Nothelfer, Alvitta Ottley, Khairi Reda, Irene Reppa, Karen Schloss, Yunhai Wang, and Cindy Xiong. They reviewed the submissions to the VisPsych workshop, and many of them also reviewed the chapters included in this book. Their time, effort, knowledge, and wisdom are deeply appreciated, and their comments, critiques, and suggestions have been indispensable to this book as well as the VisPsych workshop.

We would like to thank all members of the Advisory Board of the VisPsych workshop, including Sarah Creem-Regehr, Sara Fabrikant, David Laidlaw, Bradley Love, Sine McDougall, Melanie Tory, Barbara Tversky, and Colin Ware. In particular, we appreciate very much their advice and suggestions on the scope and future development of visualization psychology as a new academic subject. Our special thanks to Barbara Tversky for her keynote speech "How Graphics Communicate?" that provided an inspiring opening of the VisPsych workshop, and to Colin Ware for his scholarly Foreword that introduces this book from the perspective of a pioneer of *Visualization Psychology*.

Last but not least, we are in debt to Helen Desmond (Springer) for her advice and patience throughout this book project. We appreciate the support offered by the Springer teams for managing the Springer LaTeXstyle and for typesetting, cover design, web access, and many other matters in the publication process. The editors of this book also appreciate the editorial fees offered by Springer and have collectively decided to donate all the fees to IEEE VIS2022 as a sponsorship.

Oxford, UK	Min Chen
Surrey, BC, Canada	Brian Fisher
Boulder, CO, USA	Danielle Albers Szafir
London, UK	Rita Borgo
Swansea, UK	Darren J. Edwards
Boston, USA	Lace Padilla
December 2022	

Contents

Contributors

Alfie Abdul-Rahman King's College London, London, UK

Wolfgang Aigner St. Poelten University of Applied Sciences, St. Poelten, Austria

Danielle Albers Szafir University of North Carolina-Chapel Hill, Chapel Hill, NC, USA

Dietrich Albert University of Graz, Graz, Austria

Tomás Alves HUMAN-LAB, Graphics and Interaction, INESC-ID and Instituto Superior Técnico—University of Lisbon, Lisbon, Portugal

Melanie Bancilhon Washington University in St. Louis, St. Louis, MO, USA

Michael A. Bedek University of Graz, Graz, Austria

Rita Borgo King's College London, London, UK

Michael Burch University of Applied Sciences Graubünden, Chur, Switzerland

Min Chen University of Oxford, Oxford, UK

Deborah A. Cronin University of California, Davis, CA, USA

Kristin M. Divis Sandia National Laboratories, Albuquerque, NM, USA

Johanna Doppler Haider Vienna University of Technology, Vienna, Austria

Darren J. Edwards Swansea University, Swansea, UK

Brian Fisher Simon Fraser University, Surrey, BC, Canada

Amy Rae Fox University of California San Diego, La Jolla, CA, USA

Sandra Gama HUMAN-LAB, Graphics and Interaction, INESC-ID and Instituto Superior Técnico—University of Lisbon, Lisbon, Portugal

Daniel Gonçalves HUMAN-LAB, Graphics and Interaction, INESC-ID and Instituto Superior Técnico—University of Lisbon, Lisbon, Portugal

Florian Grassinger St. Poelten University of Applied Sciences, St. Poelten, Austria

Michael J. Haass Sandia National Laboratories, Albuquerque, NM, USA

Joana Henriques-Calado CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal

James D. Hollan University of California San Diego, La Jolla, CA, USA

Karl Horvath Medical University of Graz, Graz, Austria

Breannan Howell Sandia National Laboratories, Albuquerque, NM, USA

Klaus Jeitler Medical University of Graz, Graz, Austria

Maurice Koch University of Stuttgart, Stuttgart, Germany

Neesha Kodagoda Genetec UK, London, UK

Bettina Kubicek University of Graz, Graz, Austria

Kuno Kurzhals University of Stuttgart, Stuttgart, Germany

David H. Laidlaw Brown University, Providence, RI, USA

Laura E. Matzen Sandia National Laboratories, Albuquerque, NM, USA

Eva Mayr Danube University for Continuing Education Krems, Krems an der Donau, Austria

Kushin Mukherjee University of Wisconsin–Madison, Madison, WI, USA

Alvitta Ottley Washington University in St. Louis, St. Louis, MO, USA

Lace Padilla Northeastern University, Boston, USA

Paul Parsons Purdue University, West Lafayette, IN, USA

Margit Pohl Vienna University of Technology, Vienna, Austria

Benjamin Potzmann Vienna University of Technology, Vienna, Austria

Bárbara Ramalho HUMAN-LAB, Graphics and Interaction, INESC-ID and Instituto Superior Técnico—University of Lisbon, Lisbon, Portugal

Alisa Rogers Sandia National Laboratories, Albuquerque, NM, USA

Karen B. Schloss University of Wisconsin–Madison, Madison, WI, USA

Melissa A. Schoenlein University of Wisconsin–Madison, Madison, WI, USA

Tobias Schreck Graz University of Technology, Graz, Austria

Patrick Seidler Genetec UK, London, UK

Thomas Semlitsch Medical University of Graz, Graz, Austria

Lin Shao Graz University of Technology, Graz, Austria
Fraunhofer Austria Center for Data Driven Design, Graz, Austria

Andrea Siebenhoferr-Kroitzsch Medical University of Graz, Graz, Austria
Goethe University Frankfurt am Main, Frankfurt, Germany

Holger Stitz datavisyn GmbH, Linz, Austria

Christina Stoiber St. Poelten University of Applied Sciences, St. Poelten, Austria

Marc Streit Johannes Kepler University Linz, Linz, Austria

Markus Wagner St. Poelten University of Applied Sciences, St. Poelten, Austria

Colin Ware University of New Hampshire, Durham, New Hampshire, USA

Daniel Weiskopf University of Stuttgart, Stuttgart, Germany

Florian Windhager Danube University for Continuing Education Krems, Krems an der Donau, Austria

B. L. William Wong Auckland University of Technology, Auckland, New Zealand

Part I
Visualization Psychology from a Psychology Perspective

Visualizations have the power to inspire, compel, and even change our firmly held beliefs. At their best, visualizations effortlessly reveal the true nature of data to a wide range of audiences. On the other hand, such raw power can lead to serious communication failings. For example, the Cone of Uncertainty produced by the National Hurricane Center leads viewers to incorrectly think that all storms grow in size over time. Without understanding the psychology of how the mind processes visualizations, predicting when or why some visualizations confuse readers, whereas others are effortlessly understood, can be challenging. By understanding the psychological processes that drive our experience with visualizations, designers can avoid predictable pitfalls and create new visualizations that harness the immense processing power of the brain.

Fortunately, psychology has a long history of using visual stimuli to understand mental processes. Some of the earliest experiments in psychology (circa 1850) used participants' responses to visual stimuli to infer information about the visual system (relationship between the eye and brain). Today, psychological research offers a wealth of knowledge about how people perceive, reason, and make decisions with visual information. Only recently have researchers worked to generalize psychological findings to visualizations. As psychologists have historically conducted many studies with visual stimuli, these results may generalize to visualizations. However, visualizations may have unique characteristics that reveal new and unexplored aspects of human cognition, making research at the nexus of psychology and visualizations exciting to pursue.

The chapters in this part take a psychological perspective by using data visualization research to build on the empirical traditions of psychological sciences, gaining insights into mental processes. Chapter 1 reviews empirical research on the use of color in visualizations that examines the generalizability of seminal findings in color perception and reveals new insights into the relationships among perception, language, and data attributes. Chapter 1 also provides practical recommendations for using color in visualizations. Chapter 2 reviews prominent theories of graph comprehension, each of which offers high-level descriptions of the relationship between cognitive processes. This chapter details the evolution of psychological theories of graph comprehension and their application to data visualizations. Chapter 3 reviews theories of mental models of visualizations. Mental models describe the process by which people create and store internal representations of graphs. Mental model research can provide valuable insights for optimizing learning and memory of visualizations. Chapter 4 reviews theories of visualization decision-making and demonstrates the application of theoretical frameworks. This chapter highlights how new advances in decision-making can help improve visualizations intended for public or policy-level decisions. The final chapter (Chap. 5) compares the publication models for psychology and visualization research, highlighting a need for greater integration between the fields and alternative publication approaches.

The chapters in this part offer historical and modern perspectives on the psychology of visualizations, ranging from lower-level processing (e.g., perception) to higher-level cognition (e.g., decision-making). While revealing new insights about the mind, these works point to practical design recommendations informed by human capabilities.

Part I Editor: Lace Padilla
Northeastern University, Boston, USA

Chapter 1
Color Semantics for Visual Communication

Karen B. Schloss, Melissa A. Schoenlein, and Kushin Mukherjee

Abstract Visual communication through information visualizations (e.g., graphs, charts, maps, diagrams, and signage) is central to how people share knowledge. In information visualizations, visual features such as color are used to encode concepts represented in the visualization ("encoded mappings"). However, people have expectations about how colors map to concepts ("inferred mappings"), which influence the ability to interpret encoded mappings. Inferred mappings have an effect even when legends explicitly specify the encoded mappings and when encoded concepts lack specific, strongly associated colors. In this chapter, we will discuss factors that contribute to inferred mappings for visualizations of categorical information and visualizations of continuous data. We will then discuss how these different kinds of factors can be united into a single framework of assignment inference. Understanding how people infer meaning from colors will help design information visualizations that facilitate effective and efficient visual communication.

1.1 Introduction

When observers look at information visualizations such as weather maps, political polling charts, and airport terminal signage, the input they receive is just an array of light projected onto the retinas of their eyes. Yet, from this input, observers ultimately glean knowledge about the world. They find out if it is likely to rain during their afternoon walk, which political candidate is expected to win an election, or which direction to dash to reach their gate before their flight departs.

Extensive perceptual and cognitive processing is needed to go from light stimulating the retina to knowledge about the world. When interpreting information visualizations, this processing includes, but is not limited to, (1) detecting and

---

K. B. Schloss (✉) · M. A. Schoenlein · K. Mukherjee
University of Wisconsin–Madison, Madison, WI, USA
e-mail: kschloss@wisc.edu; schoenlein@wisc.edu; kmukherjee2@wisc.edu

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_1

discriminating visual features (e.g., color, shape, size, texture) [2, 7, 16], (2) mapping visual features onto the concepts they represent, and [16, 28, 39, 44] (3) using (1) and (2) to derive implications about information represented in the visualization [9, 37, 45, 51]. In this chapter, we will focus on (2) by asking: how do people infer meaning from visual features?

At first, it may seem like the answer is straightforward: people can simply examine legends, labels, or accompanying text to determine the meanings of visual features. However, the answer is not so simple. People have expectations about the meanings of visual features, and visualization designs that violate those expectations are harder to interpret. Let us consider two examples.

The first example is a study by Lin et al. [18] on the meanings of colors for visualizations of categorical information. In their study, Lin et al. presented participants with colored bar charts in which each color represented a different category (e.g., kinds of fruits) (Fig. 1.1a). In one condition, the colors of the bars were selected by an algorithm that maximized the fit between concepts and colors. In another condition, the colors were default colors used by Tableau visualization software (ignoring the concepts represented in the visualization). The charts included a legend to indicate the category corresponding to each bar color. Participants were asked to answer questions about the data in the chart, and their response time (RT) was recorded. RT is a measure of interpretability, such that faster RTs for correct responses suggest greater interpretability. RTs were faster when the colors were optimized to match people's expectations, compared to the default Tableau colors, even though there was a clear legend indicating the meaning of the colors in both conditions.

The second example is a study by Schloss et al. [37] on the meanings of color for visualizations of continuous data. In their study, Schloss et al. presented participants with colormap data visualizations, in which gradations of colors represented gradations of quantity (Fig. 1.1b). Participants were told that the colormaps represented alien animal sightings on the planet Sparl, and their task was to indicate whether there were more sightings early or late in the day. The colormap visualizations included a legend that specified the mappings between lightness (dark to light) and quantity (greater to fewer sightings). Overall, participants were faster at correctly interpreting the colormap when the legend specified darker colors mapped to more. This is because observers have a dark-is-more bias leading to the expectation that darker colors map to larger quantities (see Sect. 1.3.1.2).

In both of these examples, legends indicate the meanings of colors. But, when the encoding indicated in the legend violates people's expectations, visualizations are harder to interpret. Thus, understanding visual communication requires understanding people's expectation about the meaning, or *semantics*, of visual features.

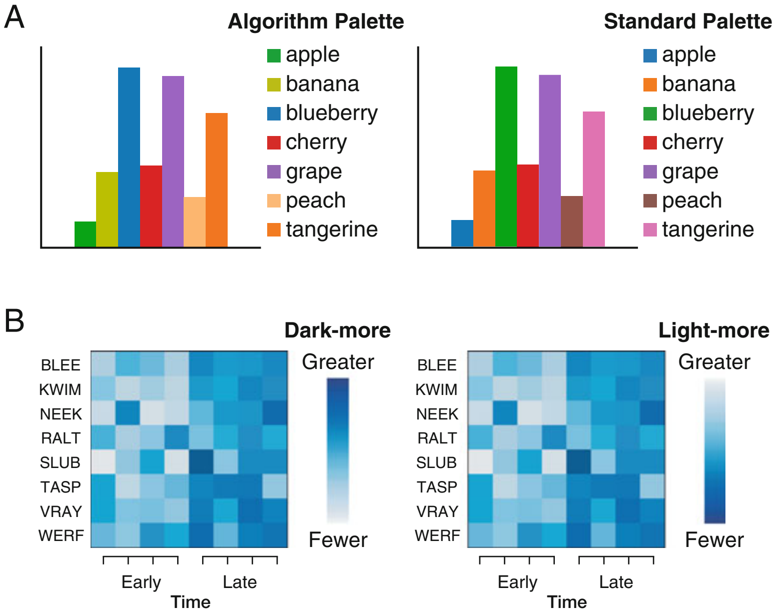

Fig. 1.1 (**a**) Bar charts representing fictitious data about fruit sales, with colors selected by an algorithm to maximize fit between concepts and colors (left) or colors determined by a standard Tableau palette order (right) (figure based on stimuli in Lin et al. [18]). (**b**) Colormap data visualization representing alien animal sightings at different times of day, with a legend specifying dark-more mapping (left) or light-more mapping (right) (figure based on stimuli in Schloss et al. [37])

1.1.1 Visual Semantics from Multiple Perspectives

When discussing visual semantics for visual communication, there are two perspectives to consider: the perspective of the designer and the perspective of the observer. If these perspectives are aligned, observers are more likely to interpret the message that the designer intended to convey through the visualization [12, 18, 25, 39, 50, 51].

Perspective of the Designer When we use the term "designer," we do so liberally to refer to anyone who creates a visualization. This could be a professional designer, but it could also be an undergraduate student creating a chart from data in their research methods course, or even a middle school student creating a diagram of the protocol for their science fair project [39]. In cases where people create visualizations for the purpose of exploring and finding patterns in their own data [10], the designer and the observer are the same person.

When a designer creates an information visualization, they use visual features to represent concepts. This mapping between concepts and visual features is called the ***encoded mapping***. For example, if the designer constructs a weather map in which darker colors signify larger amounts of rainfall, the map would have a "dark-more" encoded mapping. Designers may deliberately define the encoded mapping using their own knowledge, using recommendations from other experts, or using recommender system algorithms [18, 19, 22, 39, 43]. Alternatively, designers may rely on software defaults, which automatically assign colors to concepts in a predefined order, regardless of the concepts. In such cases, the encoded mapping is created though the designer's actions, but the designer may not explicitly consider the encoded mapping during visualization design.

Perspective of the Observer When we use the term "observer," we do so to refer to anyone who looks at visualizations with the goal of gleaning knowledge from what they see. Observers include the general public looking at public health data in the news, travelers looking at maps to find their way, students looking at diagrams to learn about mathematical or scientific processes in the classroom, and academics who look at charts to learn about the latest discoveries in their fields.

Observers' expectations about how visual features should map onto concepts are called ***inferred mappings*** [39]. As established earlier, it is harder for observers to interpret visualizations when the encoded mapping does not match their inferred mappings, even in the presence of a clear legend [18, 37, 47]. Moreover, when the encoded mapping matches their inferred mappings, observers can more easily interpret the meanings of visual features, even in the absence of a legend [8, 21, 22, 38, 39]. By understanding the nature of observers' inferred mappings, it is possible to design visualizations that match those expectations and thus facilitate visual communication.

1.1.2 Chapter Overview

In this chapter, we will use color as a lens to discuss factors that influence expectations about the meaning of visual features in information visualizations. We will discuss color semantics (i.e., the meaning of colors) in the context of two general kinds of information visualizations: visualizations of categorical information (Sect. 1.2) and continuous data (Sect. 1.3).

Historically, studies on inferred mappings discussed separate factors relevant for visualizations of categorical vs. continuous information. However, recent work suggests that they can be understood under a single framework [40], as we will discuss at the end of this chapter.

Defining the scope of artifacts that are considered to be "information visualizations" ("visualizations" for short) is a difficult endeavor (see Fox [9] and Chapter 9 of the present book). Stemming from issues raised in Fox [9], we use "information visualizations" broadly, in reference to external graphical representations (and corresponding verbal labels, if present) created to support visual communication. Here, the term "graphical" pertains to non-verbal markings in which visual features (e.g., color, shape, size, and texture) are used by a designer to communicate their intended message [2]. Although this definition of information visualizations includes visualizations of data (e.g., charts), it extends to any encoding system in which designers use non-verbal visual features to communicate their intended message [51, 53]. For example, an encoding system for recycling bins, in which a designer uses different colors to represent different kinds of trash/recyclables, is considered an information visualization. Using this broad definition enables researchers to identify generalizable psychological principles of how people infer meaning from visual features, which transcend specific design formats.

We aim for this chapter to serve two key purposes. First, it will help readers develop an understanding of psychological factors relevant to visual communication. Second, it will provide designers with knowledge that they can apply to help make visual communication effective and efficient. However, color semantics for visual communication is an active field of research. This chapter presents a snapshot of the field as it is today, but we anticipate that the ideas discussed here will evolve with new discoveries about how people infer meaning from visual features.

1.2 Color Semantics for Categorical Information

In visualizations of categorical information, discrete colors are used to represent distinct concepts. For example, Fig. 1.2a (top) shows a chart in which distinct colors represent different sectors that emit greenhouse gases, and Fig. 1.2a (bottom) shows a chart in which distinct colors represent different kinds of management for drinking water facilities. Visualizations of categorical information can be understood in contrast with visualizations of continuous data. Instead of representing discrete categories, visualizations of continuous data represent gradations of quantity, such as farm size across the world and the number of African elephants across Africa in Fig. 1.2b. In this section, we will focus on visualizations of categorical information, and we will return to visualizations of continuous data in Sect. 1.3.

One way to consider color semantics for categorical information is to focus only on the strength of the association between a color and the concept it encodes. Say, the concepts are watermelon and mango, and the chart is about fruit preferences. Mango is strongly associated with shades of orange, and watermelon is strongly associated with shades of red. So, if presented with the bar chart in Fig. 1.3a, observers would easily infer that orange encodes mango and red encodes watermelon.

But, what if concepts do not have specific, strongly associated colors, such as the more abstract concepts in Fig. 1.3b? And, what about cases when multiple concepts have similarly associated colors, such as the recycling related concepts in Fig. 1.3c? If one thinks about inferred mappings only in terms of associations between a single concept and single color, they may conclude that color cannot meaningfully encode concepts under such conditions. However, recent work suggests that color semantics is not so limited [22, 39]. To understand why, we must first draw a distinction between color–concept associations and inferred mappings.

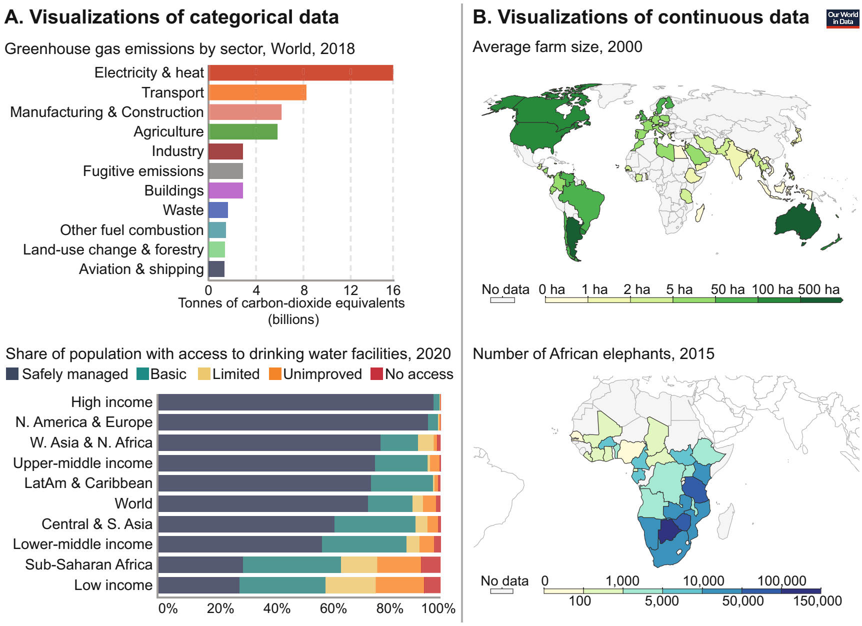

Fig. 1.2 Examples of visualizations in which color encodes (**a**) categorical data and (**b**) continuous data. Figures have been adapted from [30–33]

1.2.1 Color–Concept Associations vs. Inferred Mappings

It may be tempting to think that people's expectations about the meanings of colors in information visualizations simply depend on the association between an individual color (e.g., yellow) and an individual concept (e.g., banana) represented in the visualization. However, their expectations, or inferred mappings, are far more interesting and complex, as we explain below.

1.2.1.1 Color–Concept Associations

Color–concept associations are the degree to which individual colors are associated with individual concepts. Evidence suggests that people form color–concept associations through experiences in the world [41], at least for concepts with directly observable colors. As for more abstract concepts, some have proposed color–concept associations are formed by extension from related concrete objects that do have directly observable colors [36, 48].

For any concept, one can quantify the degree to which that concept is associated with every possible color that a human can perceive. In practice, when researchers measure color–concept associations, they sample colors over perceptual color space

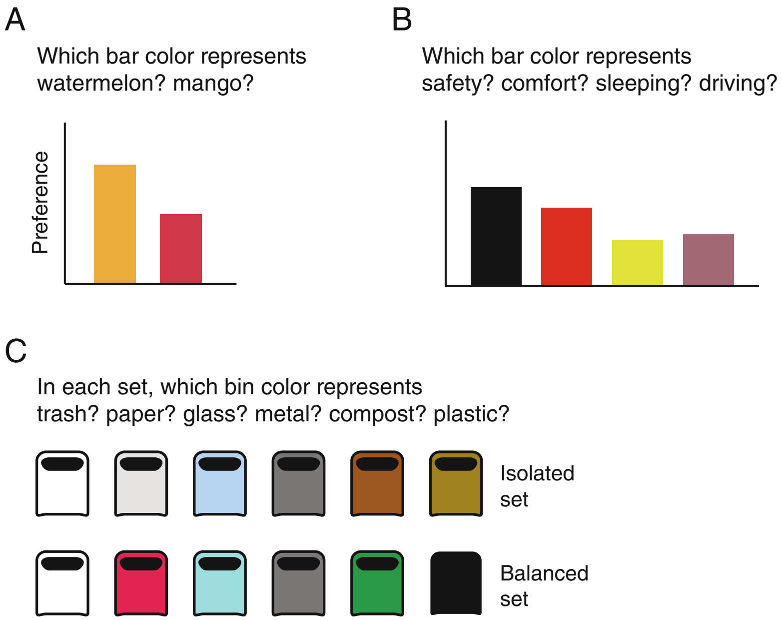

Fig. 1.3 Examples of visualizations in which colors are used to encode categories, which have been used as stimuli in experiments on inferred mappings. (**a**) Bar chart representing data about watermelon and mango, which are concepts with strong, specific associations [37]. (**b**) Bar chart representing data about safety, comfort, sleeping, and driving, which are more abstract concepts with less specific associations [22]. (**c**) Bins for discarding trash/recyclables, where multiple concepts have similar associations (see Fig. 1.6) [39]

to make the measurements more tractable [18, 19, 22, 26, 29, 38, 39]. This sample of colors is called the **color library** [22]. Figure 1.4 shows color–concept associations for the concepts banana, celery, sleeping, and driving [22]. The color library is the "UW-71" colors, which includes 58 colors uniformly sampled over CIELAB color space [29], plus an additional set of light colors required to include saturated yellows (see [22] for details).

Color–concept associations can be measured in multiple ways, including asking people to make judgments of association strength [22, 26, 38, 39, 41] and implementing algorithms that estimate associations from image or language databases [18, 19, 29, 43]. The mean associations in Fig. 1.4 were obtained by presenting participants with a concept at the top of the screen and a color patch below. They rated the association strength between each color and concept on a scale from "not at all" (0) to "very much" (1). Ratings near the middle of the scale (0.5) indicate a color was neither strongly associated nor strongly *not* associated with the given concept. For example, in the case of banana in Fig. 1.4, yellows are strongly associated, most blues are strongly *not* associated, and greens are in the middle around 0.5.

Concepts vary in the degree to which they have strong, specific associated colors within a given color library [18, 24], called **specificity** [22]. Here we focus on cases

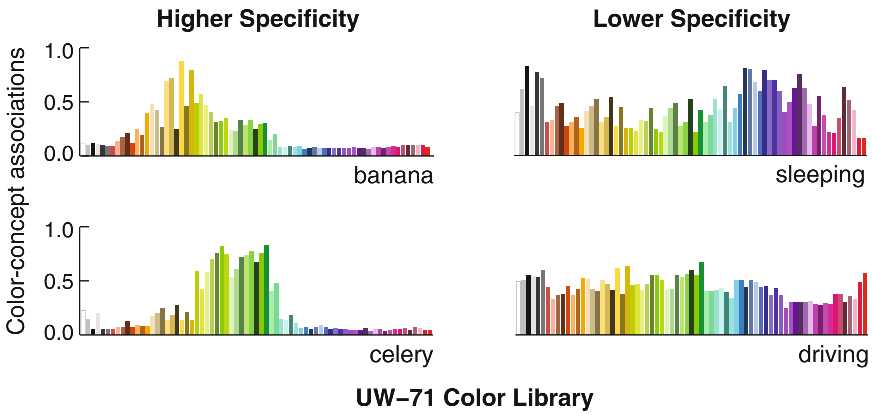

Fig. 1.4 Mean color–concept association ratings for the concepts banana and celery (higher specificity) and sleeping and driving (lower specificity) from [22]. Data were collected by asking participants to rate how much they associated each color with each concept on a scale from "not at all" to "very much." Thus, the middle of the scale (0.5) indicated neutral. The color library was the UW-71 colors, sampled over CIELAB space. Here, the colors are sorted (left to right) according to hue angle and chroma, with the achromatic colors placed leftmost

when specificity is based on the mean associations across a group of participants, but specificity could also be defined based on a single person's associations. Concepts have high specificity if they are strongly associated with some colors and weakly associated with the remaining colors in the color library. For example, Fig. 1.4 shows that celery has high specificity because it is strongly associated with greens and is weakly associated with the remaining colors. As such, concepts with high specificity have "peaky" distributions of associations over the color library. In contrast, concepts have lower specificity if they have more uniform distributions over the color library. In a fully uniform distribution, all colors would be equally associated with the concept (i.e., equal bar heights in Fig. 1.4). As shown in Fig. 1.4, the concepts sleeping and driving have lower specificity than banana and celery because their distributions are closer to uniform. Specificity can be quantified using entropy [22, 24], a mathematical measure of the "peakiness" vs. uniformity of a distribution.

Color–concept associations are essential to interpretations of color meanings in visualizations, but they are only part of the story. This brings us to inferred mappings.

1.2.1.2 Inferred Mappings

Inferred mappings are people's *expectations* about the meanings of each color in an encoding system that maps colors to concepts. Cases arise in which people infer that concepts map to weakly associated colors, even when there are more

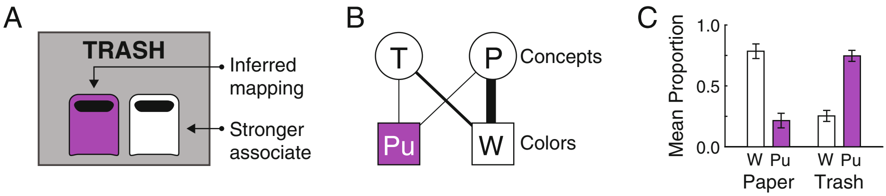

Fig. 1.5 Illustration of the distinction between color–concept associations and inferred mapping from [39]. (**a**) Trial in which participants inferred which color mapped to the target concept trash (arrows and labels to the right are for illustration only and were not part of the trial). (**b**) Bipartite graph showing color–concept association strengths for concepts trash (T) and paper (P) with colors purple (Pu) and white (W). Thicker edges connecting each concept with each color indicate stronger associations. (**c**) Mean proportion of times participants chose each color when the target was paper or trash (error bars represent standard errors of the means). Participants inferred purple mapped to trash, even though white was more strongly associated with trash

strongly associated options. To illustrate this point, we will walk through an example from Schloss et al. [39] in which participants inferred the meanings of colors on trash/recycling bins (see Fig. 1.5).

During the experiment, participants were presented with pairs of unlabeled colored bins and were asked which bin was for disposing a target item named at the top of the screen. In some trials, the target item was trash (Fig. 1.5a), and in other trials, the target item was paper. For each target, participants saw all pairwise combinations of four colored bins (left/right balanced), including white (strongest associate with paper), dark yellow (strongest associate with trash), and red and purple (both weakly associated with trash and paper). The association strengths had been obtained from color–concept association ratings from a different set of participants [39] and are shown in Fig. 1.6. The association strengths for the example trial shown in Fig. 1.5a are represented as a bipartite graph in Fig. 1.5b. In a bipartite graph, edges connect each item in one set (such as colors) to all the items in another set (such as concepts). In this bipartite graph, the circles represent the concepts trash (T) and paper (P), the squares represent the colors purple (Pu) and white (W), and the edge connecting each concept to each color represents the color–concept association strength (thicker indicates stronger associations).

Schloss et al. [39] considered two possible ways observers might approach this task. The first approach, **local assignment**, is simply to choose the color that is most strongly associated with the target. Local assignment would lead to inferring that white is for trash in Fig. 1.5. The second approach, **global assignment**, is to choose the color that optimizes assignments between all colors and concepts in the encoding system. To determine the optimal assignments in Fig. 1.5, we can compare the total goodness or "merit" of one possible assignment (e.g., T–Pu/P–W) to the alternative assignment (e.g., T–W/P–Pu) and determine which assignment has greater merit. For now, assume merit is simply color–concept association strength, but we will return to other definitions of merit below. The assignment that pairs trash with purple and paper with white has greater total merit. Thus, the global assignment approach would lead to inferring purple is for trash.

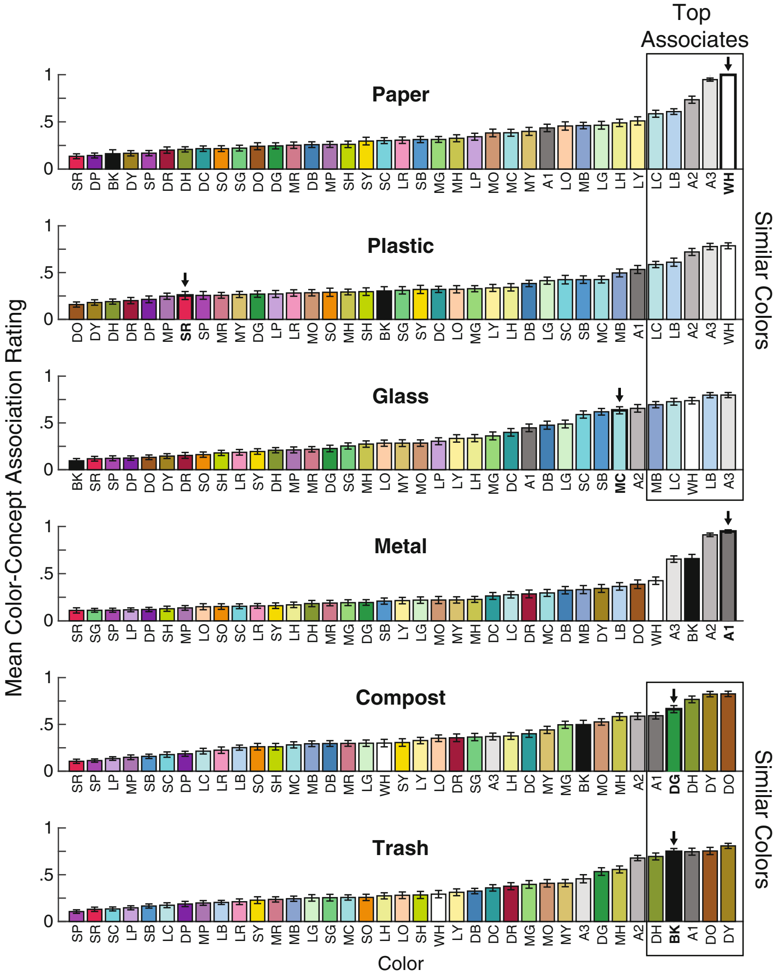

Fig. 1.6 Mean color–concept association ratings for the Berkeley Color Project 37 (BCP-37) colors and the concepts paper, plastic, glass, metal, compost, and trash (data from [39]). Colors are sorted along the x-axis from weak to strong association. Error bars indicate standard errors of the means. The top associated colors are shared among paper, plastic, and glass and shared among compost and trash. Arrows point to the optimal colors for each concept using balanced merit

Consistent with global assignment, participants reliably inferred that the purple bin was for trash (Fig. 1.5c), even though trash was more strongly associated with white. This example illustrates the distinction between inferred mappings and associations. Associations serve as the input to global assignment, but further processing leads to people's inferences about the meanings of colors in visualizations. This process is called assignment inference.

1.2.2 Assignment Inference

Assignment inference is the process by which people infer mappings between visual features and concepts in an encoding system [39]. The process was given this name because it is analogous to an *assignment problem* in the field of optimization. Assignment problems are models for assigning items in one set to items in another set in a manner that optimizes **merit**, or the "goodness," of the assignment [5, 17, 23]. Goodness is defined with respect to a given goal. For example, if the goal is to assign swimmers to strokes in a relay race to minimize time to complete the race, merit is the time it takes for each swimmer to complete each stroke. Solving an assignment problem involves finding the *best* pairings such that the overall merit across all pairs is as good as possible (i.e., the total time is as short as possible). The question is what determines merit in assignment inference?

In our discussion of the trash/paper recycling experiment illustrated in Fig. 1.5, we alluded to the notion of merit in assignment inference as association strength between each color and concept. We explained that global assignment maximizes association strength over possible assignments, even if that means assigning concepts to weakly associated colors when there are more strongly associated options. However, association strength is only one possible way to specify merit, and it is not necessarily the type of merit that humans use in assignment inference.

To study merit in assignment inference, Schloss et al. [39] assumed the role of the designer and created two different color sets (a.k.a. palettes) for trash/recycling bins (Fig. 1.3c). To create the palettes, they used two methods of defining merit and solved an assignment problem to determine the optimal color set within each definition. The logic of their experiment was that observers would be better at interpreting palettes created using a definition of merit that more closely matched merit in assignment inference. Thus, identifying which palette was easier to interpret would provide insight into the type of merit in assignment inference.

The two color palettes were designed for six types of trash/recyclables (paper, plastic, glass, metal, compost, and trash), using two definitions of merit: isolated merit and balanced merit. Both types of merit were specified as follows, using the color–concept association data shown in Fig. 1.6.

Isolated Merit Isolated merit for a given color–concept pair is simply the association strength between that color and that concept. It is called "isolated" merit because it is determined by the association between each color and concept in isolation, without accounting for other colors or concepts in the encoding system. When an assignment problem determines the optimal pairings under isolated merit, it selects color–concept pairs such that the total association strength across all pairings is as large as possible.

The color palette generated using isolated merit is shown in Fig. 1.7a. Note that paper, plastic, and glass share similar top associated colors, and compost and trash share similar top associated colors (Fig. 1.6). As a result, the colors assigned to those concepts were strongly associated with more than one concept in the encoding system. For example, plastic was associated with its assigned color, light gray (A3),

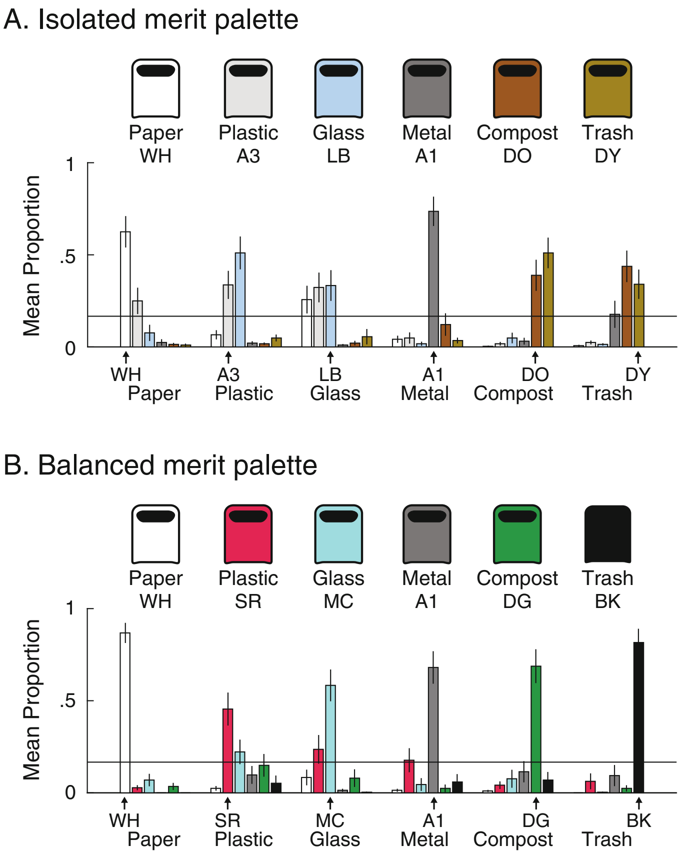

Fig. 1.7 Color palettes and corresponding plots showing the mean proportion of times participants chose each color for each concept when palettes were generated using (**a**) isolated merit or (**b**) balanced merit. Arrows point up to the correct color, specified by the optimal solution to the assignment problem using each definition of merit. Error bars represent standard errors of the means. Data are from [39]

but also was strongly associated with white (WH), the color assigned to paper, and light blue (LB), the color assigned to glass. These observations highlight a potential problem with simply maximizing association strength: it may introduce confusablity when multiple colors are associated with multiple concepts in the encoding system.

Balanced Merit Balanced merit for a given color–concept pair is computed as the association strength for that color–concept pair, minus the association strength for the color with the next most strongly associated concept. This definition of merit is called "balanced merit" because it balances prioritizing color–concept association strength while avoiding confusability that can arise when a color is strongly associated with multiple concepts in the encoding system. When an assignment problem determines the optimal pairings under balanced merit, it makes the association difference across all color–concept pairs as large as possible.

This method of defining merit can lead to assigning a concept to a weakly associated color, which can occur if the color is more associated with that concept than with all other concepts in the encoding system. For example, the color palette generated using balanced merit is shown in Fig. 1.7b. In this palette, plastic was assigned to red (SR), even though plastic is weakly associated with red, because red is more associated with plastic than with the other concepts (Fig. 1.6).

We note that isolated merit and balanced merit result in the same assignments when there are only two colors and two concepts in the encoding system. However, they can diverge when the number of colors and concepts is larger than two, as in the present experiment.

During the experiment, participants were presented with bins from each palette (between subjects) along with the list of six concept labels. They were asked to drag the label to the appropriate bin color. Accuracy was specified as the optimal assignment between colors and concepts according to the assignment problem within each source of merit. Figure 1.7 shows the mean proportion of times participants chose each color for each concept for the isolated merit palette (Fig. 1.7a) and the balanced merit palette (Fig. 1.7b). The optimal color for each concept is indicated by an arrow pointing up to the corresponding bar.

Participants were significantly more accurate for the balanced merit palette than the isolated merit palette, even though some of the associations were weaker in the balanced merit palette. For the isolated merit palette, they showed confusion, especially among white, light gray, and light blue for glass and among dark orange and dark yellow for compost and trash. For the balanced merit palette, participants consistently identified the correct assignments.

These results suggest that merit in assignment inference is closer to balanced merit than isolated merit. That is, during assignment inference, observers account for the difference in associations, and not just maximal associations when inferring mappings between colors and concepts. These results imply that if a designer aims to create color palettes that are easy for people to interpret, it is better to prioritize association difference rather than association strength.

1.2.3 Semantic Discriminability

Examining the data in Fig. 1.7, it can be seen that participants were more likely to infer "unique mappings" between colors and concepts in Fig. 1.7b than in Fig. 1.7a. That is, for each concept, there was one color that was chosen more often than all the other colors in Fig. 1.7b, but multiple colors were chosen similarly often in Fig. 1.7a. This ability to infer unique mappings is called **semantic discriminability** [22, 38]. This idea can be understood by analogy with perceptual discriminability. Perceptual discriminability concerns the degree to which observers can see the difference between different colors, whereas semantic discriminability concerns the

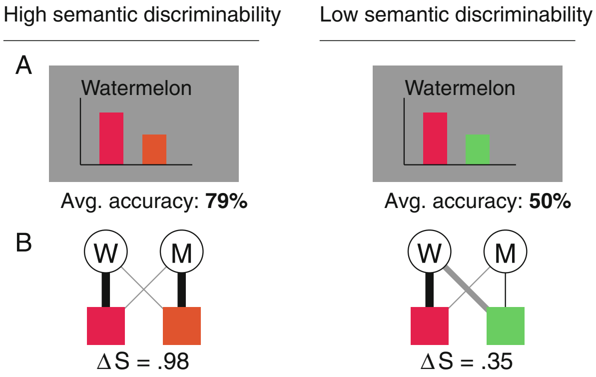

Fig. 1.8 Color palettes with high vs. low semantic discriminability. (**a**) Example trials from [38] in which participants inferred which color corresponded to target concept indicated at the top of the screen. Here, the target was watermelon; on other trials, the target was mango. The average accuracy is indicated below each example trial. (**b**) Bipartite graphs showing merit between watermelon (W) and mango (M) and each color, corresponding to the trials above in (**a**). Black edges correspond to the optimal assignment. Semantic distance (ΔS) for each color pair is indicated below the corresponding bipartite graph

degree to which observers can discern the difference in meaning between different colors. For a set of colors to be semantically discriminable, they must first be sufficiently perceptually discriminable. That is, if colors appear the same, they cannot communicate different meanings [38].

One may presume that semantic discriminability is the same thing as interpretability, but there is an important distinction. Semantic discriminability concerns an observer's inferred mapping, regardless of the encoded mapping specified by the designer. In contrast, interpretability concerns how well observers can discern the encoded mapping specified by the designer. To understand this distinction, consider the bar chart in Fig. 1.8a (left). The chart represents data about the concepts watermelon and mango using two different bar colors, red and orange. Given these two colors and concepts, one would readily infer the mapping that watermelon goes with red and mango goes with orange, not the opposite mapping. As such, these two colors have high semantic discriminability in the context of the concepts watermelon and mango. Now, a designer may choose to encode watermelon using red and mango using orange (matching the observer's inferred mapping), or they may encode watermelon with orange and mango with red (opposite of the observer's inferred mapping). In both cases, semantic discriminability of the colors is the same, but interpretability will be greater for the pairing that matches the observer's inferred mapping (watermelon-red/mango-orange).

Schloss et al. [38] developed a method to quantify semantic discriminability using a metric called **semantic distance** (ΔS). Semantic distance is a measure of how likely observers are to infer one assignment over other potential assignment(s) in an encoding system. Figure 1.8 illustrates examples of color pairs with large and small semantic distance, in the context of the concepts watermelon (W) and mango (M). The bipartite graphs in Fig. 1.8b represent the association strength between each of the two concepts with each of the two colors, corresponding to the visualizations directly above (Fig. 1.8a). The colors in Fig. 1.8 (left) have large semantic distance ($\Delta S = 0.98$) because the W-red/M-orange assignment has far greater merit than the W-orange/M-red assignment. Even if these associations vary due to noise, the difference in merit between the two assignments is sufficiently large such that W-red/M-orange will remain the optimal assignment (assuming a magnitude of variability that is typical of this kind of association data). The colors in Fig. 1.8 (right) have small semantic distance because the W-red/M-green assignment has only slightly greater total merit than the alternative assignment. If the associations varied somewhat due to noise, the outcome could reverse—the W-green/M-red assignment could have greater merit. For formal details on how semantic distance is computed, see [38].

Having defined semantic distance, the next question is whether semantic distance predicts observers' ability to interpret the meanings of colors in information visualizations. To address this question, Schloss et al. [38] asked participants to interpret the meanings of colors in bar charts with two colored bars, such as those in Fig. 1.8a. Each trial had a chart, along with a target concept named above, and participants indicated which bar (left/right) they thought corresponded to the target concept. Participants judged many color pairs, which varied in semantic distance and in perceptual distance (i.e., the difference in appearance of the two colors). Responses were scored as "accurate" if they matched the encoded mapping, which was determined as the optimal assignment using balanced/isolated merit (both are the same when there are two colors and two concepts). The charts did not have a legend, so participants did not know which response was correct during the task.

Overall, participants were able to infer optimal mappings, but accuracy increased with increased semantic distance. This effect of semantic distance was independent of effects of perceptual distance. When perceptual and semantic distance conflicted (e.g., high semantic distance, low perceptual distance), semantic discriminability better predicted accuracy. These results suggest that semantic distance does indeed predict observers' ability to interpret the meanings of colors in information visualizations.

1.2.4 Assignment Inference for Abstract Concepts?

We have established that observers can use assignment inference to interpret optimal mappings between colors and concrete concepts with directly observable colors (e.g., watermelon and mango) [38]. But, is this ability limited to concrete concepts, or might it extend to abstract concepts without directly observable colors (e.g., driving and sleeping)?

Earlier work proposed that some concepts may be "non-colorable," suggesting that such concepts cannot be meaningfully encoded using color [18, 43]. "Colorability" was defined with respect to individual pairs of colors and concepts. Concrete concepts, such as banana, celery, grape, and eggplant, were considered colorable because they had strong, specifically associated colors (i.e., high specificity), whereas abstract concepts, such as sleeping, driving, safety, and comfort, were considered non-colorable because they lacked strong, specific associated colors (i.e., low specificity).

However, this notion of colorability concerns individual concepts alone, and we know from studies on assignment inference that context plays an important role. That is, when inferring mappings between colors and concepts, observers account for all concepts and colors in the scope of an encoding system, not each concept alone (global assignment, see Sect. 1.2.1.2). And, their ability to perform assignment inference depends on semantic discriminability of the colors, which concerns the relative associations between all colors and concepts in an encoding system, not just each concept alone. These previous findings imply that observers should be able to use assignment inference to interpret optimal mappings for abstract concepts, insofar as the colors used to encode those abstract concepts are semantically discriminable.

Mukherjee et al. [22] tested this hypothesis in an experiment in which participants interpreted the meanings of colors in visualizations representing data about abstract or concrete concepts.[1] During the experiment, participants were presented with bar charts along with a set of four concept labels, as shown in Fig. 1.9a. Their task was to drag the labels from the top of the chart and place them under the colored bar that they thought corresponded to each concept. Figure 1.9a shows two example trials, one in which the concepts were all abstract, and the other in which the concepts were all concrete (in other trials abstract and concrete concepts were sometimes mixed).

Each concept appeared in four different concept sets. For example, the concept sleeping appeared with driving, safety, and comfort (set 1), with driving, grape, and banana (set 2), with driving, peach, and cherry (set 3), and with driving, efficiency, and speed (set 4) (Fig. 1.10). For each concept set, the colors of the bars were determined based on the optimal assignment using balanced merit, which selected the four best colors from the UW-71 color library to assign to each of the four concepts.

Overall, participants were able to interpret the optimal mapping between colors and concepts. For example, Fig. 1.9b shows the responses for the stimuli from Fig. 1.9a, plotting the proportion of times participants chose each color for each concept. The arrows below the x-axis point up at the correct color for each concept. Participants consistently chose the correct color for both the abstract and concrete concept sets.

[1] The abstract concepts had relatively low specificity (close to uniform color–concept association distributions), and the concrete concepts had high specificity (peaky color–concept association distributions), but that correspondence is not always the case (e.g., anger is an abstract concept but has high specificity).

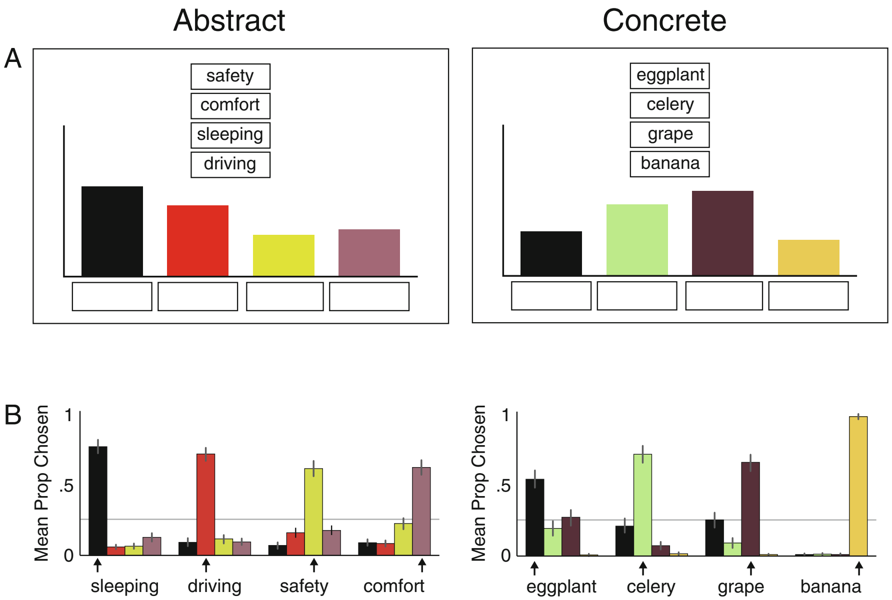

Fig. 1.9 Examples of (**a**) experiment stimuli and (**b**) corresponding data for abstract concepts (left) and concrete concepts (right) from Mukherjee et al. [22]. During the task, participants dragged each concept name to the colored bar that they thought corresponded to each concept. The mean proportion of times participants chose each color for each concept is shown in (**b**) with arrows pointing up to the correct color for each concept. Error bars represent standard errors of the means, and the horizontal gray line represents chance

However, the ability to interpret the correct color for a given concept varied depending on semantic discriminability. This relationship is shown in Fig. 1.10. The plots show the proportion of times that participants chose the correct color for the target concepts sleeping (left) and banana (right). Each plot has four points, corresponding to each of the four concept sets in which the target concept appeared. The x-axis represents the semantic discriminability between the correct color and the other colors in the palette.[2] The positive slope of the best-fit lines through the points illustrates that accuracy increased with increased semantic discriminability. For example, in set 1, participants were highly accurate at assigning yellow to banana because the other concepts in the set (eggplant, celery, and grape) did not compete with banana for yellow. Yet, in set 4, they were less accurate at assigning yellow to banana because corn competed with banana for yellow. This competition led to yellow being less semantically discriminable from the other colors in set 4

[2] Here, when we are discussing semantic discriminability, we are referring to a metric called "semantic contrast." Unlike semantic distance, which quantifies the semantic discriminability of a color palette as a whole, semantic contrast quantifies the distance between a single color and all other colors in the palette. Computational details of these two metrics can be found at [22].

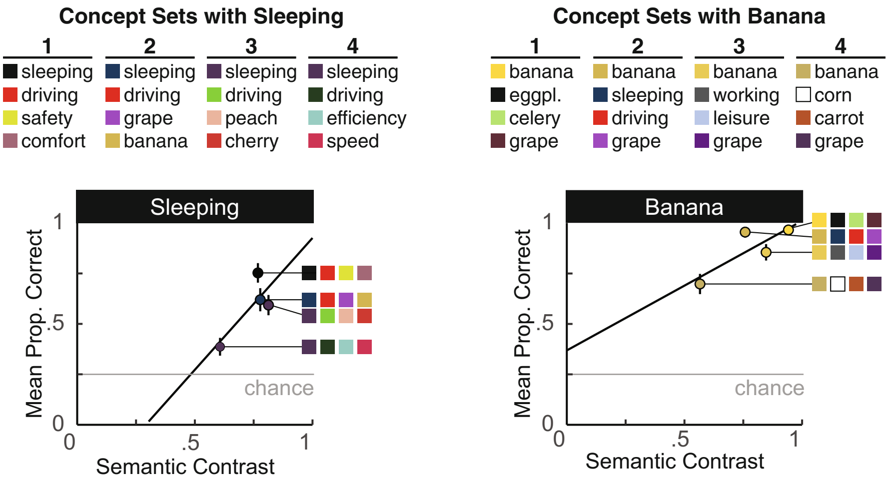

Fig. 1.10 Top: The four concept sets and color palettes for the concepts sleeping (left) and banana (right) in Mukherjee et al. [22]. Bottom: The proportion of correct responses for the target concepts sleeping (left) and banana (right) as a function of semantic discriminability of the colors in the color palettes. Each point corresponds to each of the four concept sets in which the target concepts appeared. Error bars represent standard errors of the means, and the black lines represent the best-fit lines through the data points

compared to set 1. Figure 1.10 shows the data for only two out of the 16 concepts tested in the experiment, but the pattern is representative of the full datasets (see [22]).

The results of this experiment suggest that people can use assignment inference to infer optimal mappings for both concrete and abstract concepts. Yet, the ability to do so depends on the semantic discriminability of the colors, which is determined based on all of the colors and concepts in an encoding system. In short, context matters.

1.2.5 Semantic Discriminability Theory

Thus far, we have presented evidence that semantic discriminability is important for interpreting the meanings of colors in visualizations. The next question is, what determines whether it is possible to produce semantically discriminable colors for a set of concepts? To address this question, Mukherjee et al. [22] proposed a theory called **semantic discriminability theory**. The theory states that the ability to produce semantically discriminable colors for a set of concepts depends on the difference between the color–concept association distributions for those concepts.

This theory is illustrated in Fig. 1.11, which shows three pairs of concept sets, one set with very different associations (peach and celery), one with moderately

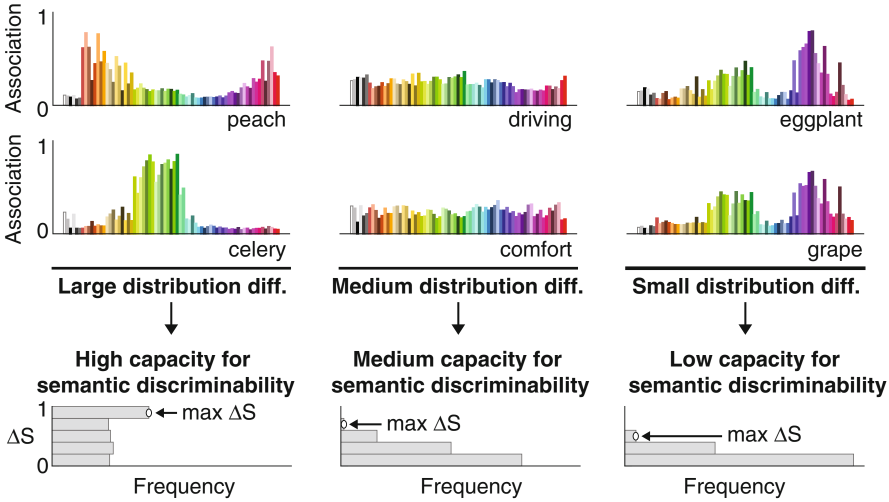

Fig. 1.11 Color–concept association distributions for concept sets with large, medium, and small distribution differences, which result in high, medium, and low capacities for semantic discriminability, respectively. The top two rows show color–concept association distributions for the 71 colors in the UW-71 color library. The bottom row shows the frequency of color pairs at varying degrees of semantic distance (ΔS). An arrow points to the maximum semantic distance (max ΔS) for each concept set. Figure adapted from [22]

different associations (driving and comfort), and one with very similar associations (eggplant and grape). Below each set of color–concept association distributions is a histogram showing the semantic distance between all pairs of colors in the UW-71 color library for that concept set. Peach and celery, which have a large distribution difference, have many color pairs with high semantic discriminability, and the maximum semantic distance (max ΔS) was a perfect semantic distance of 1. This maximum semantic distance is called the "capacity" for semantic discriminability. Examining the other two concept pairs, driving and comfort (medium distribution difference) have medium capacity for semantic discriminability, and eggplant and grape (small distribution difference) have low capacity for semantic discriminability. Note that eggplant and grape have far higher specificity (peakier distributions) than driving and sleeping, but the capacity for semantic discriminability is lower for the pair eggplant and grape because the association distributions for eggplant and grape are too similar to produce semantically discriminable colors.

The relation between capacity for semantic discriminability and distribution difference shown in Fig. 1.11 highlights only three concept pairs, but Mukherjee et al. [22] conducted a systematic study of this relationship for all pairwise combinations of 20 concepts (190 concept pairs in total). The concepts included fruits (peach, cherry, grape, banana, apple), vegetables (corn, carrot, eggplant, celery, mushroom), activities (working, leisure, sleeping, driving, eating), and properties (efficiency, speed, safety, comfort, reliability). In this full dataset, capacity was strongly correlated with distribution difference ($r = 0.93$). Capacity was also correlated with mean specificity of the individual concepts ($r = 0.82$), but significantly less so than with distribution difference. When effects of distribution difference and specificity were evaluated in a single model, only distribution difference was a significant predictor of capacity (see [22] for computational details). Aspects of these results for sets of two colors and concepts also extended to sets of four colors and four concepts. These results support semantic discriminability theory, emphasizing the importance of considering the difference between color–concept association distributions, independent of the specificity of each concept's distribution alone.

Semantic discriminability theory was originally formulated and studied with respect to color. However, Mukherjee et al. [22] suggested it as a general theory with potential to extend beyond color to other visual features (e.g., size, shape, texture) and perceptual features in other modalities (e.g., sound, odor, touch).

1.2.6 Summary and Open Questions for Visualizations of Categorical Information

We began Sect. 1.2 by explaining that the notion of inferred mappings is distinct from color–concept associations. Using assignment inference, observers infer globally optimal assignments between colors and concepts, even if that means assigning a color to a weakly associated concept. We then provided evidence that the ability to perform assignment inference to interpret optimal assignments depends on the semantic discriminability of the colors. Observers can successfully perform assignment inference to interpret optimal assignments for abstract and concrete concepts, as long as the colors representing those concepts are semantically discriminable. Finally, we discussed semantic discriminability, a theory on the constraints for producing semantically discriminable colors for a given set of concepts. Supporting the theory, capacity for semantic discriminability increases with increased differences between the color–concept association distributions for the set of concepts. The series of studies in this section emphasize that people's inference about the meanings of colors is highly context-specific, depending on the other colors and concepts in the scope of the encoding system.

Although much has been learned from research on color semantics for categorical information, many open questions are yet to be answered. Here, we highlight two such questions.

Cultural Effects? Color–concept associations serve as input to assignment inference, which result in interpretations of the meanings of colors in visualizations [36]. If this input differs due to cultural differences in color–concept associations [13, 14, 49], then the output (interpretation of the meanings of colors) should also differ. However, if the process underlying assignment inference is a general cognitive mechanism, and the input is known, then it should be possible to predict cultural differences in the output. Cross-cultural experiments are needed to test if assignment inference is actually a culturally general cognitive mechanism, and if the current model of assignment inference [22, 39] can predict cultural similarities/differences in inference about the meanings of colors in information visualizations.

This logic extends to semantic discriminability theory. The theory implies that distribution difference will predict capacity for semantic discriminability in any culture, as long as the association distribution data reflect the associations held by a given culture. But, if the color–concept associations collected from one culture are used to predict capacity for another culture that has different color–concept associations, then the predictions might be misleading and the palettes generated might not be semantically discriminable for those who are a part of that second culture. Future research is needed to test whether cultural variations in color–concept association distribution differences predict cultural variations in capacity for semantic discriminability.

Extension to Other Perceptual Features? The work described in this section focused on color, but semantic discriminability theory is broadly defined to apply to other perceptual features in vision (e.g., shape, visual texture, orientation, size) and features in other modalities (e.g., sounds, odors, tactile textures) [22]. However, questions remain as to how to effectively sample perceptual features in these other domains to test this hypothesis, and which other kinds of perceptual features will have systematic and distinct enough associations with concepts to support semantic discriminability.

1.3 Color Semantics for Continuous Data

In Sect. 1.2, we focused on color semantics for visualizations representing categorical information. In Sect. 1.3, we turn to factors that contribute to color semantics for visualizations of continuous data, such as the colormap data visualizations ("colormaps" for short) from Fig. 1.2b. In colormaps, gradations of color are mapped to gradations of quantities across a spatial representation [12]. The spatial representation could take a variety of forms depending on the type of data, including geographical maps to show climate data across regions of the world, a brain map to show neuroimaging data across different regions of the human brain, or a matrix to show gene expression co-occurrences in different samples of organisms.

Traditionally, the literature has drawn a distinction between the kinds of factors that influence inferred mappings for categorical information and continuous data. For categorical information, the emphasis has been on "direct" color–concept associations (Sect. 1.2), whereas for continuous data, the emphasis has been on "relational" associations. Direct color–concept associations (or direct associations for short) are just the color–concept associations we discussed in Sect. 1.2, but here we call them "direct" associations to distinguish them from "relational associations." Unlike direct associations, which are the degree to which an individual color is associated with an individual concept, relational associations are correspondences between relational properties of visual features and relational properties of concepts [40]. For example, observers have a dark-is-more bias, inferring that darker colors map to larger quantities [4, 8, 21, 37, 40, 47]. The dark-is-more bias is relational because it concerns the relative lightness within a sequence of colors, rather than the lightness of any individual color alone.

Although previous work distinguished factors relevant for visualizations of categorical information and continuous data, recent work by Schoenlein et al. [40] suggests that inferred mappings for continuous data visualized in colormaps are influenced by both direct and relational associations. The relative contribution of these different factors can be understood as different sources of merit in assignment inference. In the following sections, we will first discuss different kinds of relational associations for colormaps and then explain how relational and direct associations can all be considered as sources of merit in assignment inference for colormap data visualizations.

1.3.1 Relational Associations for Colormaps

Several types of relational associations can contribute to inferred mappings for colormap data visualizations (Table 1.1). The effects of relational associations on inferred mappings are governed by at least two main principles:

1. **Applicability principle:** A relational association can only be activated if it is applicable to the visualization, given the perceptual properties of the visualization.
2. **Combination principle:** If multiple relational associations are activated, they will combine to produce the inferred mapping. Sometimes relational associations work together and sometimes they conflict. When they conflict, they may cancel each other or some relational associations may dominate others, depending on their relative strength.

In the following sections, we will discuss empirical evidence for each type of relational association listed in Table 1.1. In doing so, we will consider perceptual properties that determine whether each relational association applies to a given visualization, and how relational associations combine when multiple are activated at the same time.

1.3.1.1 Structure Preservation

Structure preservation is a relational association in which structure among perceptual features corresponds to structural properties among concepts to which they are mapped [3, 11, 12, 20, 27, 46, 50]. One example of such structure is the progression of lightness (light to dark) and the progression of quantity (small to

Table 1.1 Types of relational associations between visual features and quantity

Association type	Description	Related references
Structure preservation	Structure among perceptual features corresponds to structural properties among concepts to which they are mapped.	[3, 11, 12, 20, 27, 46, 50]
Dark-is-more bias	Regions that appear darker map to larger quantities.	[4, 8, 21, 37, 40, 47]
Opaque-is-more bias	Regions that appear more opaque map to larger quantities.	[1, 35, 37]
Hotspot-is-more bias	Regions closer to the center of "hotspots" map to larger quantities.	[42, 47]
High-is-more bias	Colors higher up on vertically oriented legends map to larger quantities.	[12, 37, 47, 50]

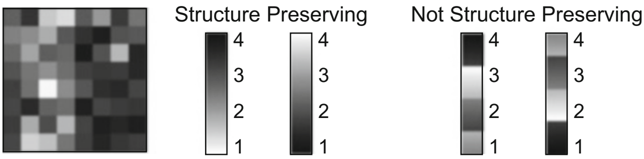

Fig. 1.12 Example colormap assigning lightness (light to dark) to quantities (1–4) with legends that maintain structure preservation (left) and legends that do not maintain structure preservation (right). Figure adapted from [40]

large). For example, Fig. 1.12 shows a colormap and four accompanying legends specifying encoded mappings that could correspond to the colormap. The left two encoded mappings are structure-preserving because gradations of lightness align with gradations of quantity. From the perspective of structure preservation, both of these encoded mappings (dark-more and light-more encodings) are equally good. However, the right two encoded mappings are not structure-preserving because lightness is scrambled with respect to quantity.

Structure preservation is applicable whenever there is structure among the concepts that can be preserved by the visual features that represent those concepts. Structure preservation is always applicable when discussing continuous data because the data have graded structure. Structure preservation is assumed in all of the rest of the relational associations we will discuss next.

1.3.1.2 Dark-is-More Bias

The dark-is-more bias leads to the inference that darker colors map to larger quantities [4, 8, 21, 37, 40, 47]. It is applicable when colors in the color scale vary in lightness. When we say "lightness," we mean the perceptual dimension of lightness, going from dark to light (e.g., L* in CIELAB space). We note that in HSB color space, both the "saturation" (S) and "brightness" (B) dimensions vary in perceptual lightness, so when some discuss color scales defined by saturation variation, there is still lightness variation. Although it is possible for color scales to have no lightness variation (e.g., vary only in hue or perceptual saturation), in practice, color scales tend to vary in lightness, which helps perceive spatial structure in data [15, 34, 52]. Thus, the dark-is-more bias is almost always applicable to inferred mappings for colormaps.

Early evidence for the dark-is-more bias comes from studies in which participants were shown colormaps without legends and were asked to indicate which regions represented "more" (Fig. 1.13a) [8, 21]. Participants systematically chose the darker regions, suggesting they inferred that darker colors mapped to larger quantities.

More recent evidence comes from studies in which participants were shown colormaps with legends specifying the encoded mapping. Participants were asked to

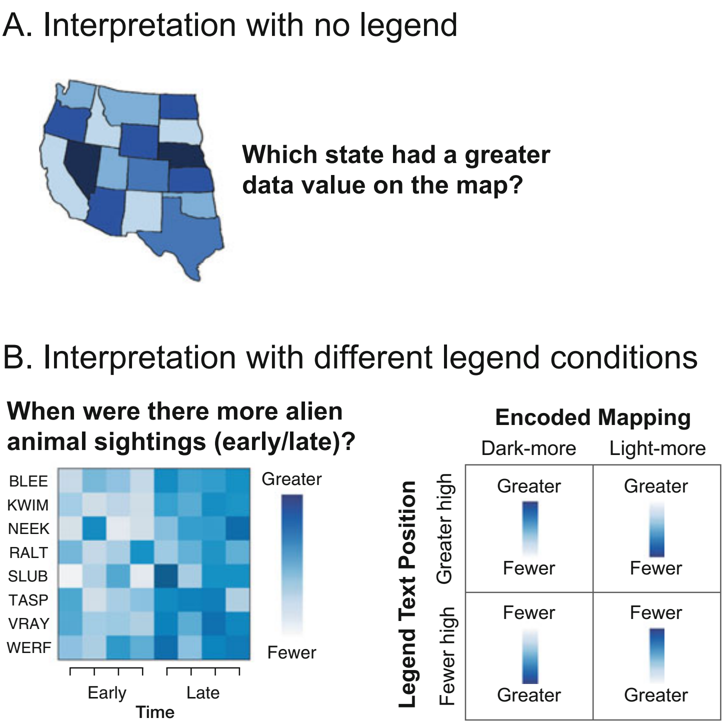

Fig. 1.13 Types of tasks for assessing inferred mappings for colormaps. (**a**) Interpretations are made based on the colormap alone, with no legend to specify the encoded mapping (as in [21]). Inferred mappings are assessed by examining the proportion of times each option is chosen. (**b**) Interpretations are made by reporting the correct answer based on the legend (as in [37]). Inferred mappings are assessed by determining which encoded mappings facilitate faster response times (RTs) to make accurate responses (i.e., encoded mappings facilitate faster RTs if they better match inferred mappings)

correctly interpret the colormap according to the legend [37]. On half of the trials, the legend specified dark-more encoding, and on the other half, the legend specified light-more encoding (Fig. 1.13b). Also, on half of the trials, “greater” was at the top of the legend, and on half of the trials, it was at the bottom. Participants therefore had to read the legend on every trial to determine the encoded mapping. Participants were faster at correctly interpreting the colormap when the legend specified dark-more encoding, providing further evidence for the dark-is-more bias.

We will discuss what happens when the dark-is-more bias combines with each of three other relational associations in the following sections.

1.3.1.3 Opaque-is-More Bias

The opaque-is-more bias leads to the inference that regions appearing more opaque map to larger quantities. This bias is only applicable when regions of the colormap appear to vary in opacity. The percept of opacity variation can be achieved by starting with a colored region and decreasing its alpha in a series of steps so that more and more of the background becomes visible through the region’s surface [35]. Functionally, this amounts to interpolating between the color of that region and the color of the background (Fig. 1.14). This interpolation can vary along the perceptual dimensions of lightness, as described above in Sect. 1.3.1.2, hue, chroma, or any combination therein.

Apparent opacity variation therefore depends not only on properties of the color scale used to create the colormap, but also properties of the background. Schloss et al. [37] developed a metric for quantifying apparent opacity variation, called the *opacity variation index*. It is computed for a given color scale and background by: (1) identifying the endpoint of the color scale that contrasts most with the background, (2) drawing a line between the color of that region and the color of the background region in CIELAB space, (3) calculating the distance between each color on the color scale and its projection onto the line, and (4) computing the root

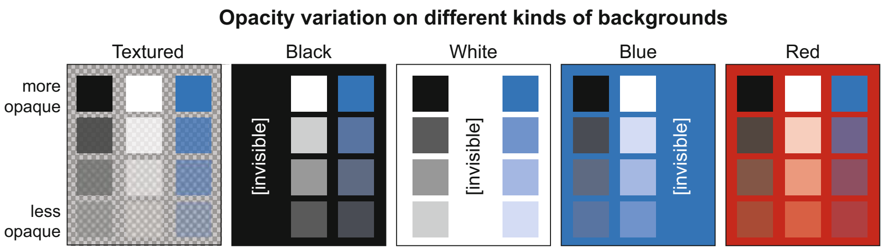

Fig. 1.14 Black, white, and blue squares are displayed on different backgrounds to show how their appearance changes with opacity variation. The squares in the top row are opaque, and they decrease in opacity in each sequential row below. Colored squares are rendered invisible when they match the color of the background, but they are included in the diagram for completeness

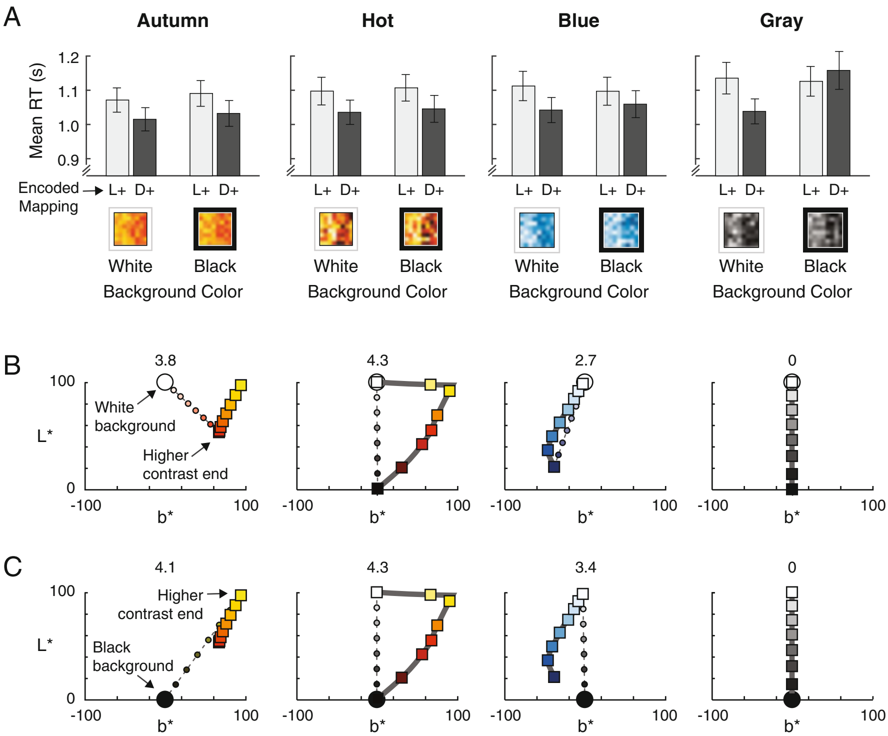

Fig. 1.15 Opacity variation in colormap visualizations. (**a**) Mean response times (RTs) to correctly interpret dark-more vs. light-more encodings of colorscales varying in opacity when presented on a white vs. black background. Error bars represent standard error of the means. (**b**) Plots in CIELAB space, showing the colors from each color scale (squares) and the interpolation between the highest-contrast color and the white background (circles). Plots are shown on the plane of L* (lightness) and b* (yellowness/blueness), and the axis for a* (redness/greenness) is not shown. The number above each plot is the opacity variation index. (**c**) The same as (**b**), but for a black background. Figure adapted from [37]

mean-squared error of those distances[3] (Fig. 1.15b and c). This method is only an initial approach to quantifying apparent opacity variation in colormaps and likely can be improved upon in future work. Nonetheless, it was effective at predicting human performance, as we will discuss next.

Researchers have long considered that the background could have an effect on people's inferred mappings for colormaps, but this notion was framed in terms of contrast with the background [21]. For example, McGranaghan [21] presented

[3] As specified in [37], the *opacity variation index* is defined as $log(z + 1)$, where z is the root mean-squared error between each point on the color scale (square markers in Fig. 1.15b and c) and the line between the highest-contrast color and the background (circle markers in Fig. 1.15b and c). Smaller values correspond to greater perceptual evidence for opacity variation.

participants with partial maps of the United States, with states colored in various shades of blue (Fig. 1.13a). Maps were shown on a white, gray, or black background. McGranaghan hypothesized that participants would infer dark meant more on a light background, but light meant more on a dark background, in a *contrast-is-more* bias. The results showed that participants inferred dark meant more on all three backgrounds, though the effect was weaker on the black background. This was taken as evidence against the existence of a potential contrast-is-more bias.

In a subsequent study examining the effects of the background, Schloss et al. [37] presented participants with colormaps of fictitious data about alien animal sightings on white or black backgrounds. The color scales were standard scales used in visualization (Autumn, Hot, and Gray from MATLAB, and ColorBrewer Blue). As described in Sect. 1.3.1.2, each colormap had a legend, and participants were asked to interpret the colormap by reading the legend and indicating whether there were more alien animal sightings early or late in the day (Fig. 1.13b).

The effect of the background lightness depended on the color scale (Fig. 1.15a). For Autumn and Hot, the background had no effect, and responses were consistent with a dark-is-more bias on both white and black backgrounds. For ColorBrewer Blue, the background had a moderate effect, but responses were still consistent with a dark-is-more bias on both the black and white background (similar to what McGranaghan [21] reported). For Gray, the background had a larger effect that trended toward inferences that lighter colors meant more. The authors were initially puzzled by why the background mattered for some color scales and not others, until they realized that the colormaps differed in how much the regions appeared to vary in opacity. Thus, they developed the opacity variation index described above to test whether these effects could be predicted by apparent opacity variation. Overall, there was a bias for participants to be faster when the legend specified dark-more encoding than light-more encoding (dark-is-more bias), but this was modulated by opacity variation in a manner consistent with an opaque-is-more bias.

This brings us to our first consideration of the combination principle. On a white background, the dark-is-more bias and opaque-is-more bias work together—the darker region is also the more opaque region, so response times were especially fast for dark-more encoding than light-more encoding. On a black background, these two biases conflict—the darker region is the less opaque, more transparent region. Under such conflicts, if the opacity variation index was strong (Gray color scale), the opaque-is-more bias tended to override the dark-is-more bias when combining to produce the inferred mapping. When the index was moderate (ColorBrewer Blue color scale), the opaque-is-more bias dampened the effect of the dark-is-more bias but did not cancel it out. This finding aligns with the results reported by McGranaghan [21]. Finally, when the index was weak (Autumn and Hot), and therefore not applicable, there was no opaque-is-more bias activated to influence the inferred mapping. One can avoid conflicts between the dark-is-more bias and opaque-is-more bias by either: (1) presenting colormaps on light backgrounds, such that the two biases work together, or (2) avoiding colormaps that appear to vary in opacity when displayed on a dark background.

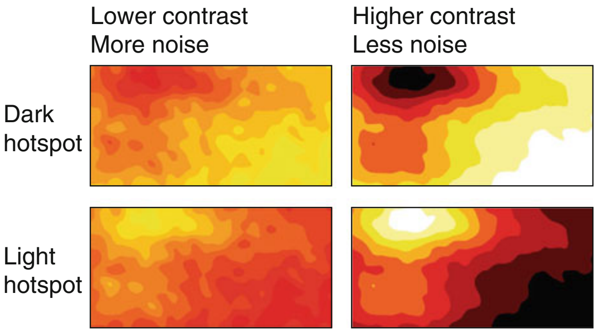

Fig. 1.16 Colormaps with dark (top) and light (bottom) hotspots. Colormaps on the right have higher lightness contrast and less noise in the underlying dataset than colormaps on the left

1.3.1.4 Hotspot-is-More Bias

The hotspot-is-more bias leads to the inference that regions closer to the center of "hotspots" map to larger quantities [42, 47]. It is applicable when there is spatial structure in the data that looks like a hotspot (e.g., concentric rings), such as in Fig. 1.16.

Until now, in this section, we have discussed colormaps in which there was little spatial structure in the data to provide a cue to the locus of larger quantities (e.g., grids of randomly colored squares [37]). However, Schott [42] raised the possibility that color-based biases (e.g., dark-is-more bias) may not influence interpretations of colormaps when there are strong spatial cues to the locus of large quantities, such as hotspots. Hotspots are properties of datasets in which the region with extreme values (very high or very low values) is surrounded by roughly concentric regions with less and less extreme values. These patterns are characteristic of fMRI and EEG signals from neuroimaging data and storm patterns in meteorological data.

Sibrel et al. [47] tested whether a hotspot-is-more bias exists, and if so, whether it overrides the influence of the dark-is-more bias. They asked participants to interpret colormaps containing hotspots, such as those in Fig. 1.16, left. The participants were told the colormaps represented data about alien animal sightings in different regions of a planet, and their task was to press the left/right arrow key to indicate whether there were more sightings on the left or right of the region based on the legend. On one half of the trials, the hotspot was light, and on the other half, the hotspot was dark (hotspot location and darker region location were left/right balanced across trials). In this initial experiment, participants were faster at responding when the legend indicated dark was more (dark-is-more bias), with no effect of whether the hotspot was light or dark (no hotspot-is-more bias).

Surprised by this result, Sibrel et al. [47] conducted a series of subsequent experiments to see if they could find evidence for a hotspot-is-more bias and to see if they could make it strong enough to override the dark-is-more bias. First they modified the trial structure such that the hotspot was a reliable cue to the locus of the larger quantity. That is, rather than the legend specifying that the colors in the hotspot mapped to more on 50% of the trials, the legend was biased to indicate that the hotspot mapped to more on 77% of the trials. Here, they found evidence for a hotspot-is-more bias. When the hotspot was dark, RTs were faster for dark-more encoding than light-more encoding, consistent with both the dark-is-more bias and the hotspot-is-more bias. However, when the hotspot was light, causing a conflict between the dark-is-more bias and hotspot-is-more bias, the difference in RTs was significantly weaker. Still the hotspot-is-more bias did not override the dark-is-more bias. To get the hotspot-is-more bias to slightly, but significantly, override the dark-is-more bias, it was necessary to not only have the hotspot be a reliable cue, but also to make it even more perceptually salient through increasing lightness contrast and reducing visual noise in the image (Fig. 1.16, right).

These results suggest that color-based biases are powerful contributors to inferred mappings, which cannot be merely dismissed when there is strong spatial structure in the data.

1.3.1.5 High-is-More Bias

The high-is-more bias leads to the inference that colors positioned higher up on a vertically oriented legend map to larger quantities. The high-is-more bias is only applicable when colormaps have vertically oriented legends, which is not always the case in experiments [21] or in practice, as documented by Christen et al. [6]. The high-is-more bias is part of a more general expectation that larger amounts will be displayed higher in space [12, 50].

Evidence supporting the high-is-more bias comes from studies showing that response times to correctly interpret colormaps are faster when “greater” is at the top of the legend than at the bottom [37, 47] (Fig. 1.13b). Moreover, the dark-is-more bias has a larger influence when “greater” is at the top of the legend than at the bottom. One way to view this finding is that when these two biases work together (i.e., the darker region encodes “more” and “more” is represented at the top of the legend), inferences are clearer and interpretation is especially easy, but once these biases conflict, inferences become muddled and interpretation is generally harder.

In Sect. 1.3.1, we have highlighted several kinds of relational associations that can contribute to inferred mappings, when they are applicable. We also described what can happen to inferred mappings when different sources of relational associations combine and which types of relational associations tend to dominate when different types conflict. Ultimately, a goal in this line of work is to construct a comprehensive model to predict people’s inferred mappings for information visualizations, while accounting for all applicable factors for a given type of visualization. Next, we discuss an initial step toward such a model.

1.3.2 Assignment Inference for Visualizations of Continuous Data

Until now in this chapter, we have discussed distinct factors that contribute to inferred mappings for different kinds of visualizations: direct color–concept associations for visualizations about categorical information and relational associations for visualizations of continuous data. However, recent work by Schoenlein et al. [40] has bridged these areas by extending the framework of assignment inference previously established with visualizations of categorical information (Sect. 1.2.2) to visualizations of continuous data. Their approach is illustrated in Fig. 1.17.

During their study, participants were presented with colormaps such as those in Fig. 1.17 (left). The colormaps represented fictitious data about environmental concepts, such as the amount of ocean water in different counties. The task was to indicate where there was more of the concept, on the left or right side of the map. There was no legend, so participants responded according to their inferred mappings. For both colormaps in Fig. 1.17, the dark-is-more bias implies participants should infer the darker side represents more ocean water. However, direct associations imply different responses for the top and bottom colormaps. For the top colormap, direct associations imply they will choose the darker side because ocean water is more associated with dark blue than with light brown (congruent with the dark-is-more bias). For the bottom colormap, direct associations imply that participants will choose the *lighter* side because ocean water is more associated with light blue than with dark yellow (incongruent with the dark-is-more bias). How will participants respond?

This problem can be considered through the framework of assignment inference. Direct and relational associations are distinct sources of merit, and inferred mappings are computed over the weighted combination of these two sources of

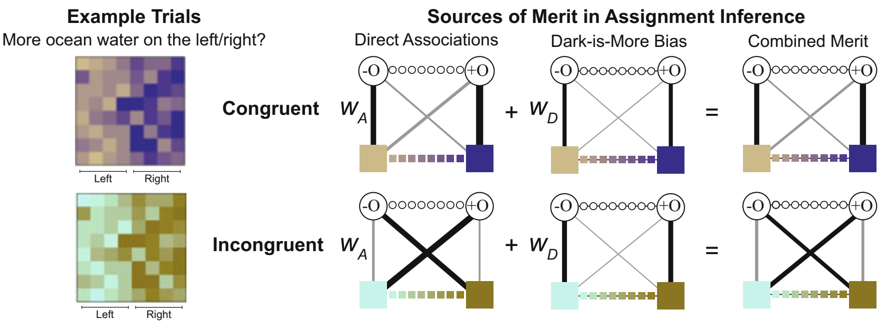

Fig. 1.17 Example trials from Schoenlein et al. [40] in which participants inferred which region of colormaps (left/right) represented more of the domain concept ocean water. Inferences can be predicted by simulating assignment inference using a weighted combination of multiple sources of merit (direct associations and dark-is-more bias), in cases when they are congruent (top row) and incongruent (bottom row). Figure reproduced from [40]

merit. Figure 1.17 (right) illustrates this scenario using separate bipartite graphs to represent merit from direct associations and the dark-is-more bias. The concepts are the two endpoints of the conceptual dimension (a lot of ocean water, $+O$, and no ocean water, $-O$). The two colors are the two endpoint colors from the color scale used to create the colormap. Although the colormaps included gradations of colors and quantities, the problem was reduced to the two endpoint colors and concepts. This simplification was possible because in their stimuli, association strength and lightness both varied monotonically between the two endpoint colors. Given that there were only two colors and two concepts, merit from direct associations could be treated as association strength between each endpoint color and each endpoint concept (as described for categorical data in Sect. 1.2.2). Merit for the dark-is-more bias puts greater value on dark-more/light-less edges than light-more/dark-less edges (see [40] for details). The question was, how much weight should be put on direct associations (W_A) vs. the dark-is-more bias (W_D) when combining these sources of merit?

Schoenlein et al. [40] addressed this question by systematically varying the amount of weight put on each source while simulating assignment inference, and determined which weighting best predicted participant's inferred mappings. They found that the best combination of weights placed a 0.7 weight on direct associations and a 0.3 weight on dark-is-more bias. This combined weighting was better for predicting participant judgments than weighting on each source of merit alone. With greater weight on direct associations, direct associations overrode the dark-is-more bias when they were in strong conflict. As such, participants inferred that lighter colors mapped to more ocean water in the incongruent example in Fig. 1.17.

This study has set up a method for combining multiple sources of merit to predict inferred mappings in assignment inference. Of course, direct associations and the dark-is-more bias are only two potential sources of merit in assignment inference. But, Schoenlein et al.'s [40] approach can be extended to account for all known direct and relational sources of merit, plus new sources of merit that are yet to be discovered.

1.3.3 Summary and Open Questions for Visualizations of Continuous Data

In Sect. 1.3, we have discussed multiple factors that influence inferred mappings for colormap data visualizations: structure preservation, dark-is-more bias, opaque-is-more bias, hotspot-is-more bias, high-is-more bias, and direct associations. We have also presented a framework of understanding how to combine multiple (sometimes competing) sources of merit to predict inferred mappings from assignment inference.

Still, many questions remain about the nature of inferred mappings for continuous data, especially with regard to the kind of data being represented and the observers' knowledge about the domain. These questions have been raised in previous work [6, 37, 40, 47], and we summarize them here.

More of What? When colormaps use color to encode quantities, "more" could refer to more of the concept being represented, or more of the numerical values used to measure the concept. For example, when discussing data about response time, researchers often refer to instances in which people were *faster* (i.e., when there was more speed), which corresponds to smaller numbers (i.e., amount of milliseconds). Under such instances, people may infer that darker colors are mapped to faster response times, which correspond to smaller numbers. The question is whether the relational associations reported above, all focusing on what maps to "more," operate at the conceptual or numeric level.

Effects of Domain Expertise? Some people have expertise working with colormaps in particular domains (e.g., neuroscientists, climate scientists, epidemiologists). Within these domains, conventions arise, which sometimes violate the biases reported above. For example, in neuroimaging, there is a convention to use light-more encodings [6], violating the dark-is-more bias. Questions remain concerning whether domain experts have qualitatively different inferred mappings from novices, and if so, whether those differences are constrained to colormaps in their domain, or generalize to other colormaps on data outside their area of expertise.

Relative Contributions of Different Sources of Merit? Schoenlein et al. [40] established the relative weighting to be placed on direct associations and the dark-is-more bias when simulating assignment inference when considering only those two sources of merit. Open questions concern how to construct a comprehensive model that places appropriate weight on each source of merit that is applicable for any given kind of visualization.

Addressing these questions will deepen our understanding of inferred mappings for colormaps, and this knowledge will help design colormaps that facilitate interpretability.

1.4 Conclusion

A central goal in the psychology of information visualization is understanding people's inferences about the meanings of visual features in visualizations. If visualizations are designed in a manner that aligns with people's expectations, then people can spend less cognitive resources on figuring out what the visual features mean and focus their effort on figuring out how to use the information presented in visualizations to think about and act on the world around them.

It may be tempting to seek out prescriptive rules for how to use color to convey meaning (e.g., use color x to always mean y). However, as discussed in this chapter, inferences about the meanings of colors are context-dependent, contingent on the other colors and concepts in the encoding system, as well as spatial properties (e.g., hotspot structure, height in space). Thus, fully anticipating people's expectations about the meanings of colors in visualizations will require a comprehensive model that accounts for all factors influencing inferred mappings. Initial steps toward this end are showing promising results, but there is much more exciting work to be done.

Although we do not yet have a comprehensive model, designers can still use the results discussed in this chapter to inform their designs. For example, evidence suggests that for visualizations of categorical information, it is better to use color palettes that maximize association difference rather than association strength. Ultimately, when selecting colors for visualizations, we advocate for learning as much as possible about the various factors that can influence people's expectations about the meanings of colors. Then, use critical thought to consider which factors are most relevant for a particular visualization, and how to leverage them in a manner that makes sense for the design as a whole.

By deepening the understanding of color semantics, this field of research is providing insight into the human ability to translate perceptual input into knowledge about the world, while providing insight into how to design visualizations that facilitate visual communication.

References

1. A. N. Bartel, K. J. Lande, J. Roos, and K. B. Schloss. A holey perspective on Venn diagrams. *Cognitive Science*, 46(1):e13073, 2021.
2. J. Bertin. *Semiology of graphics: diagrams, networks, maps*. University of Wisconsin Press, Madison, 1983.
3. J. Blachowicz. Analog representation beyond mental imagery. *The Journal of Philosophy*, 94(2):55–84, 1997.
4. C. A. Brewer. Color use guidelines for mapping and visualization. In A. M. MacEachren and D. R. F. Taylor, editors, *Visualization in Modern Cartography*, pages 123–148. Elsevier Science Inc., Tarrytown, 1994.
5. R. Burkard, M. Dell'Amico, and S. Martello. *Assignment problems: revised reprint*. SIAM, 2012.
6. M. Christen, D. A. Vitacco, L. Huber, J. Harboe, S. I. Fabrikant, and P. Brugger. Colorful brains: 14 years of display practice in functional neuroimaging. *NeuroImage*, 73:30–39, 2013.
7. W. S. Cleveland and R. McGill. Graphical perception and graphical methods for analyzing scientific data. *Science*, 229(4716):828–833, 1985.
8. D. J. Cuff. Colour on temperature maps. *The Cartographic Journal*, 10(1):17–21, 1973.
9. A. R. Fox. A psychology of visualization or (external) representation? *arXiv preprint arXiv:2009.13646*, 2020.
10. R. L. Goldstone, F. Pestilli, and K. Börner. Self-portraits of the brain: cognitive science, data visualization, and communicating brain structure and function. *Trends in Cognitive Sciences*, 19(8):462–474, 2015.
11. G. P. Goodwin and P. Johnson-Laird. Reasoning about relations. *Psychological Review*, 112(2):468, 2005.
12. M. Hegarty. The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, 3(3):446–474, 2011.

13. D. Jonauskaite, A. M. Abdel-Khalek, A. Abu-Akel, A. S. Al-Rasheed, J.-P. Antonietti, Á. G. Ásgeirsson, K. A. Atitsogbe, M. Barma, D. Barratt, V. Bogushevskaya, et al. The sun is no fun without rain: Physical environments affect how we feel about yellow across 55 countries. *Journal of Environmental Psychology*, 66:101350, 2019.
14. D. Jonauskaite, J. Wicker, C. Mohr, N. Dael, J. Havelka, M. Papadatou-Pastou, M. Zhang, and D. Oberfeld. A machine learning approach to quantify the specificity of colour–emotion associations and their cultural differences. *Royal Society Open Science*, 6(9):190741, 2019.
15. G. Kindlmann, E. Reinhard, and S. Creem. Face-based luminance matching for perceptual colormap generation. In *IEEE Visualization, 2002. VIS 2002.*, pages 299–306. IEEE, 2002.
16. S. M. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3):185–225, 1989.
17. H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955.
18. S. Lin, J. Fortuna, C. Kulkarni, M. Stone, and J. Heer. Selecting semantically-resonant colors for data visualization. In *Computer Graphics Forum*, volume 32, pages 401–410. Wiley Online Library, 2013.
19. A. Lindner, N. Bonnier, and S. Süsstrunk. What is the color of chocolate?–extracting color values of semantic expressions. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2012, pages 355–361. Society for Imaging Science and Technology, 2012.
20. C. J. Maley. Analog and digital, continuous and discrete. *Philosophical Studies*, 155(1):117–131, 2011.
21. M. McGranaghan. Ordering choropleth map symbols: The effect of background. *The American Cartographer*, 16(4):279–285, 1989.
22. K. Mukherjee, B. Yin, B. E. Sherman, L. Lessard, and K. B. Schloss. Context matters: A theory of semantic discriminability for perceptual encoding systems, 2021.
23. J. Munkres. Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics*, 5(1):32–38, 1957.
24. S. Murthy, R. Hawkins, and T. L. Griffiths. Shades of confusion: Lexical uncertainty modulates ad hoc coordination in an interactive communication task. *Cognition*, in press.
25. D. Norman. *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books (AZ), 2013.
26. L.-C. Ou, M. R. Luo, A. Woodcock, and A. Wright. A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *Color Research & Application*, 29(3):232–240, 2004.
27. S. Palmer. Fundamental aspects of cognitive representation. In E. Roach and B. B. Lloyd, editors, *Cognition and Categorization*, pages 259–303. Lawrence Erlbaum Associates, Hillsdale, NJ., 1978.
28. S. Pinker. A theory of graph comprehension. *Artificial Intelligence and the Future of Testing*, 73:126, 1990.
29. R. Rathore, Z. Leggon, L. Lessard, and K. B. Schloss. Estimating color-concept associations from image statistics. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1226–1235, 2020.
30. H. Ritchie and M. Roser. Biodiversity. *Our World in Data*, 2021. https://ourworldindata.org/biodiversity.
31. H. Ritchie and M. Roser. Clean water and sanitation. *Our World in Data*, 2021. https://ourworldindata.org/clean-water-sanitation.
32. H. Ritchie and M. Roser. Farm size. *Our World in Data*, 2021. https://ourworldindata.org/farm-size.
33. H. Ritchie, M. Roser, and P. Rosado. CO_2 and greenhouse gas emissions. *Our World in Data*, 2020. https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions.
34. B. E. Rogowitz, L. A. Treinish, and S. Bryson. How not to lie with visualization. *Computers in Physics*, 10(3):268–273, 1996.
35. R. E. Roth, A. W. Woodruff, and Z. F. Johnson. Value-by-alpha maps: An alternative technique to the cartogram. *The Cartographic Journal*, 47(2):130–140, 2010.

36. K. B. Schloss. A color inference framework. In . G. V. P. L. MacDonald, C. P. Biggam, editor, *Progress in Colour Studies: Cognition, Language, and Beyond*. John Benjamins, Amsterdam, 2018.
37. K. B. Schloss, C. C. Gramazio, A. T. Silverman, M. L. Parker, and A. S. Wang. Mapping color to meaning in colormap data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):810–819, 2019.
38. K. B. Schloss, Z. Leggon, and L. Lessard. Semantic discriminability for visual communication. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1022–1031, 2021.
39. K. B. Schloss, L. Lessard, C. S. Walmsley, and K. Foley. Color inference in visual communication: the meaning of colors in recycling. *Cognitive Research: Principles and Implications*, 3(1):5, 2018.
40. M. A. Schoenlein, L. Campos, J, K. J, Lessard, L, Schloss, and K. B. Unifying effects of direct and relational associations for visual communication. *IEEE Transactions on Visualization and Computer Graphics*, in press.
41. M. A. Schoenlein and K. B. Schloss. Colour-concept association formation for novel concepts. *Visual Cognition*, pages 1–23, 2022.
42. G. D. Schott. Colored illustrations of the brain: some conceptual and contextual issues. *The Neuroscientist*, 16(5):508–518, 2010.
43. V. Setlur and M. C. Stone. A linguistic approach to categorical color assignment for data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):698–707, 2016.
44. P. Shah and P. A. Carpenter. Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, 124(1):43, 1995.
45. P. Shah and J. Hoeffner. Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1):47–69, 2002.
46. R. N. Shepard and S. Chipman. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1(1):1–17, 1970.
47. S. C. Sibrel, R. Rathore, L. Lessard, and K. B. Schloss. The relation between color and spatial structure for interpreting colormap data visualizations. *Journal of Vision*, 20(12):7–7, 2020.
48. C. Soriano and J. Valenzuela. Emotion and colour across languages: implicit associations in Spanish colour terms. *Social Science Information*, 48(3):421–445, 2009.
49. D. S. Y. Tham, P. T. Sowden, A. Grandison, A. Franklin, A. K. W. Lee, M. Ng, J. Park, W. Pang, and J. Zhao. A systematic investigation of conceptual color associations. *Journal of Experimental Psychology: General*, 149(7):1311, 2020.
50. B. Tversky. Visualizing thought. *Topics in Cognitive Science*, pages 499–535, 2011.
51. B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, 2002.
52. C. Ware. Color sequences for univariate maps: Theory, experiments and principles. *IEEE Computer Graphics and Applications*, 8(5):41–49, 1988.
53. C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, 2nd edition, 2004.

Chapter 2
Theories and Models in Graph Comprehension

Amy Rae Fox

Abstract Graph comprehension is the act of deriving meaning from graphs, an activity grounded in visuospatial reasoning that develops through a combination of instruction and practice. What we know about the mechanisms of graph comprehension stems from interleaving lines of inquiry in statistics, computer science, education, and psychology dating back to the 1980s. In this integrative review, I describe how models of graph comprehension evolved in response to developments in cognitive theory, offering a critical commentary on how foundational theories build upon each other, extending rather than replacing theoretical claims at different levels of analysis. I illuminate the landscape of contemporary research, before concluding with an argument for the role of visualization psychology in supporting theoretical integration across disciplinary boundaries.

2.1 Introduction

There is a conceptual paradox at the center of research on graph comprehension. The reason we employ graphical displays is that—in relation to text or tables of numbers—they seem effortless. Deriving meaning from a graph is described as "seeing" the information, equated with the facile fluency of perception. But this effortless access obscures a murky, error-ridden reality. Correctly reading a graph is much harder than we think. After 40 years of empirical research and theory building, we have learned that our ability to interpret a graph is influenced by a multitude of interacting factors affecting the display, the individual, and the situation.

In this chapter I offer a historical commentary on the development of graph comprehension research. I describe how theory in graph comprehension arose out of empirical research across disciplines and propose a role for visualization psychology in facilitating theoretical integration. This chapter will be useful for visualization

A. R. Fox (✉)
University of California San Diego, La Jolla, CA, USA

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_2

researchers looking to navigate the interdisciplinary milieu of graph comprehension, and students of behavioral and social sciences seeking a primer on this essential area of research.

2.1.1 What Kind of Graph Is a Graph?

The term external representation is used to indicate things in the world—subject to experience by human perception—that purposefully refer to other things. External representations can be constructed for any sensory modality and medium, though the visualization researcher is particularly interested in those employing *graphics* that can be *seen* on some *surface*. The text on this page is a visual external representation, with the letters of the alphabet functioning as symbols referring to sounds that you have learned to assemble into words from which you construct a certain understanding of what I intend to communicate. Similarly, a photograph is a visual external representation, referring via resemblance and analogy to the scene it depicts. A rich spectrum lies between these symbolic texts (describing the world) and analogous pictures (depicting the world). The design and interpretation of external representations belongs to the interdisciplinary realm of *semiotics*: the study of meaning-making (see Chap. 9). The focus of this chapter is a subset of external representations colloquially referred to as graphs (from the Greek graphē "writing, drawing"), charts, or plots: diagrams that convey relationships between sets of information via visual-spatial variables in a coordinate system (see [6, 62]). These are not to be confused with another set of representations referred to as "graphs": collections of edges that join pairs of vertices (à la "graph theory; node-link diagrams). Graphs are typically distinguished from *maps* which use scaled space to represent geographic relations. Both kinds of graphs belong to the larger class of diagrams: external representations that use space and simplified visual forms to convey relationships between their referents. Importantly, the use of these terms in empirical research is as fluid as the taxonomies that seek to structure them (see [25, 34, 53]). While the models and theories of comprehension reviewed in this chapter reference graphs specifically, it is reasonable to infer that the general purpose mechanisms of graph comprehension may also apply to the larger class of external representations.

2.2 An Abridged History of Theory in Graph Comprehension

As is often the case with interdisciplinary research, the study of graph comprehension arose from the needs of practice, rather than an invariable march of basic theory. The pioneering graphical inventions of Playfair, Minard, and Galton in the "golden age" of visualization were only made mainstream through inclusion in textbooks (e.g., [11]) and standards reports (e.g., [2]), through championing in professional texts (e.g., [78]) and essays in scholarly journals (e.g., [21, 45]). As the use of such "statistical graphics" spread, guidelines were needed for when and how they could be used to communicate effectively: a call for science to explain the art.

The earliest empirical investigations were published in statistics [22, 24, 82] and consisted of discrete comparisons between bar and pie charts, testing a viewer's performance in judging proportions. Concurrent work in educational psychology [85] tested secondary school students on their memory of facts learned from bar and line charts, pictographs, and tables. Studies of this kind were framed as empirical tests of guidelines offered in textbooks like that of Brinton [11] but were subject to methodological critiques of construct validity. In contextualizing their results, the authors tended to frame outcomes as properties of the representations themselves: *a bar chart is more effective at [X] than a pie chart*, while contemporary scholars would identify performance as arising from the *interaction between* the individual and representation. This subtle but important difference betrays that the focus of early efforts was on understanding the nature of the representations and their properties.

These types of point-to-point and application-grounded studies would continue for decades, in the absence of frameworks, theories, or models to guide causal or mechanistic investigation. The work was published in statistics, educational psychology, computer graphics, and the burgeoning field of HCI. This would be the case until three developments in the 1980s paved the way for a more coherent, additive body of research to unfold. First, Jaques Bertin's seminal work *A Semiology of Graphics* was translated from French to English by WJ Berg (under the supervision of Howard Wainer) in 1983. Bertin was the first to offer a concise language and structure for decomposing the questions we might ask about what a graphic is and how it might work. Second, post-cognitive revolution, substantial theories connecting visual perception to higher order cognition had been published in cognitive science—notably Marr [52] and Ullman [80]. Finally, the "mental imagery debate" was well underway, which saw leading cognitive scientists debating the nature of mental representation. This focuses on representation spurred interest in *external* representation and in particular how graphics are leveraged for problem solving and communication (e.g., [46]).

In the sections that follow, I describe a progression of theoretical development that has shaped the trajectory of graph comprehension research—work that directly addresses the fundamental question: *how are humans able to read graphs?* Our focus will be on the elaboration of general *theory*—accounts of the mechanisms through which our interaction with statistical graphics unfold—rather than individual empirical contributions. We will see examples of theory reasoned from personal experience, appeal to logic, and theory reasoned from experimental evidence. A substantial body of theory has been developed in information visualization and education that addresses the application of visualization and diagrammatic representations more broadly, though (cognitive) theory in graph comprehension can be construed as its foundation, the backbone of investigations exploring specific phenomena observed within those interactions. Questions like *what kind of graph is most effective for decision-making?* or *how can we help learners correctly interpret a graph?* rely on general purpose mechanisms of graph comprehension, just as questions of effective linguistic communication rely on the underlying mechanisms of reading and speech comprehension. Figure 2.1 summarizes early theoretical contributions, including a number of general taxonomic grammars and computational efforts that are not discussed in further detail.

The reader will notice that our understanding of graph comprehension did *not* progress via development of *competing* models and theories. Rather, research has unfolded as a progressive elaboration of a vast problem space, with works that shed light on disparate aspects or tasks, and others that expand on prior theory at different levels of detail, iterating rather than refuting. Half of the challenge is deciding what questions need to be answered, and here lies the power and difficulty of such interdisciplinary inquiry.

2.2.1 A Semiology of Graphics: Bertin

> To utilize graphic representation is to relate the visual variables to the components of the information. With its eight independent variables, graphics offers an unlimited choice of constructions for any given information. (...) The basic problem in graphics is thus to choose the most appropriate graphic for representing a given set of information. — Bertin [6, p. 100]

Jacques Bertin (1918–2010) was a French cartographer, born in the suburbs of Paris and educated in the School of Cartography at the Sorbonne. An esteemed map-maker, he contributed to new methods of cartographic projection as the head of research at France's National Center for Scientific Research (CNRS) [58]. Yet his most widespread legacy would be the first and most far-reaching effort to provide a theoretical foundation to the design of information graphics, first offered in the text *Sémiologie Graphique* [5].

Bertin's volume resists concise summary,[1] though its most oft-cited concepts in contemporary writing, are the *visual variables* and *levels of organization*, which taken together form a table of perceptual properties: a heuristic for information-visual mapping (Fig. 2.2a). Bertin organized the tools at our (external) representational disposal in terms of space (two *planar dimensions*: location on a surface) and the visual (*retinal*) properties along with marks positioned within the space can vary: size, value, texture, color, orientation, and shape. In short, the visual variables

[1] Any attempt to summarize the 400 page volume would be too brief, and this author is convinced that although widely cited, the depth of Bertin's intellectual contributions is underestimated on account of opaque linguistic constructions. Bertin also contributed theory on *levels of reading* [p. 141], *stages of processing*[140], *functions of graphics*[p. 160], and *information processing*[p. 166]. The motivated reader is strongly encouraged to give "Part 1. Semiology of the Graphic Sign-System" a close reading [6].

Early Theoretical Contributions to Graph Comprehension

Year	Author	Key Contributions
1967	Bertin	**visual variables; levels of organization**
1981/2	Pinker	early version of Pinker 1990, as MIT report
1983	Bertin	english translation by WJ Berg
1984	Cleveland & McGill	**ordering of elementary perceptual tasks** (codes); (*re-articulates Bertin's visual variables with partial accuracy rankings*)
1985	Kosslyn	*Book review in the J. Amer. Statistics Assoc contained thorough but accessible primer of contemporary information processing psych as applied to graphics*
1986	Mackinlay	codification of graphic design criteria in a form that can be used by the presentation tool, including expanded (theoretical) ranking of elementary codes
1987	Cleveland & McGill	expanded set of **elementary codes** with refined accuracy rankings
1987	Simkin & Hastie	**judgement tasks; elementary mental processes** *(demonstrates interaction of encoding & task; positions Cleveland & McGill in context of Pinker & information procesing)*
1989	Kosslyn	**analytic scheme for deconstructing graphs; acceptability principles** *(thorough treatment, framing common graphical intuitions in terms of information processing)*
1990	Pinker	***first general process account***; *(schema-theoretic account from information processing perspective)*
1993	Lohse	*computational (symbolic, GOMS) production-system model predicting scanpath & response time from question & graph*
1994	Gillan & Lewis	*computational Mixed Arithmetic-Perceptual (MA-P) model derived from task analyses*
2002	Shah & Freedman	**construction-integration model of graph comprehension** (*builds upon Pinker 1990 to integrate iteration & prior-knowledge driven processing)*
2002/3	Peebles & Cheng	*ACT-R/PM based computational model capable of predicting scanpaths on cartesian graphs under questions*
2008	Trafton, et. al	*argues for explicit inclusion of 'spatial processing' and 'cognitive integration' in existing models*

Fig. 2.1 Early influential theories, frameworks, and models in Graph Comprehension [32, 49, 59, 60, 76]

(A) Bertin (1967, 1983)

LEVEL OF THE VARIABLE

VISUAL VARIABLES	ASSOCIATIVE *(similar)*	SELECTIVE *(different, groups)*	ORDERED *(ordered)*	QUANTITATIVE *(proportional)*
	Position	Position	Position	Position
	Size	Size	Size	Size
	Color (value)	Color (value)	Color (value)	
	Texture	Texture	Texture	
	Color (hue)	Color (hue)		
	Orientation	Orientation		
	Shape			

(B) Cleveland & McGill (1984, 1987)

DATA TYPE

ELEMENTARY PERCEPTUAL TASKS: QUANTITATIVE *(1984)*	ELEMENTARY CODE: QUANTITATIVE *(1987)*
Position (*along a common scale)*	Position (*along a common scale)*
Position (*along a non-aligned scale)*	Position (*along a non-aligned scale)*
Length, Direction, Angle	Length
Area	Angles
Volume, Curvature	Slopes*
Shading, Color (saturation)	Areas
	Volumes
	Densities
	Color (saturation)
	Color (hue)

(C) Mackinlay (1986)

DATA TYPE

PERCEPTUAL TASKS	NOMINAL	ORDINAL	QUANTITATIVE
	Position	Position	Position
	Color (hue)	Density	Length
	Texture	Color (saturation)	Angle
	Connection	Color (hue)	Slope
	Containment	Texture	Area
	Density	Connection	Volume
	Color (saturation)	Containment	Density
	Shape	Length	Color (saturation)
	Length	Angle	Color (hue)
	Angle	Slope	
	Slope	Area	
	Area	Volume	
	Volume		

Fig. 2.2 Four contributions ranking perceptual accuracy of visual-spatial encodings. Bertin (**a**) was reasoned phenomenologically, Cleveland and McGill (**b**) derived from experimental studies with quantitative proportion judgments, which (**c**) Macklinlay [51] extended for nominal and ordinal data reasoning from existing psychophysics studies, not empirically validated in the context of graph comprehension

offer eight channels into which information can be mapped. Bertin argued these channels have varying capacities for adequately representing different aspects of information: a correspondence between the nature of the information and perceptual requirements for discerning it in graphical form. In an orthogonal scheme, he posited four *levels of organization* that govern what *about* some information we might seek to perceive. Selective perception involves discerning categorical belonging; associated perception grouping like instances; and ordered perception discerning step-wise order and quantitative perception discerning the absolute value of an instance or numeric ratio between instances. Bertin asserted that to map data to a visual variable, the level of organization of the data must correspond to the capacity of the visual variable (Fig. 2.2a). Any mismatch is a source of "graphic error" [6, p. 64].

Bertin envisioned a unifying framework that could govern the design of all kinds of graphics. A CNRS colleague reflected that it was the exposure to hundreds of representations from different scientific domains—brought to Bertin for advice—that endowed him with the sort of global perspective required to write a text as comprehensive as Sémiologie Graphique [7]. In modern parlance, we would say Bertin offered a structured design space for mapping information-to-graphical marks. Though it is important to note that these ordered mappings were inferred from a combination of logical reasoning and perceptual experience rather than experimental evidence. Bertin's treatise is partially descriptive: structuring his observation of the components of graphical communication, and prescriptive: offering guidelines for how and when certain mappings should be made. In justification of the levels of organization assigned to each variable, Bertin offers a test, a sort of phenomenological self-check (or to the researcher, suggested experimental task) that should convince the reader. In this way, the classification of visual variables can be read as a set of hypotheses for controlled psychophysics experiments. The continued influence of Bertin's work should remind us of the value of the kind a priori theorizing required to construct such a theoretical framework. He did not conduct experiments or build models to explain data, but rather imposed a coherent logical structure on a disorganized set of phenomena growing rapidly in importance. Though perceptual experiments would follow, Bertin's visual variables still stand as the most common starting point for information-graphic mapping in visualization design. His work is widely cited in the pioneering research in computer graphics and information visualization, as well as the psychological studies of graphical perception that would begin in earnest in the 1980s.

2.2.2 *Elementary Structures in Graphical Perception: From Cleveland and McGill to Simkin and Hastie*

> We do not pretend that the items on our list are completely distinct tasks; for example, judging angle and direction are clearly related. We do not pretend that our list is exhaustive;

> for example, color hue and texture (Bertin 1973) are two elementary tasks excluded from the list because they do not have an unambiguous single method of ordering from small to large and thus might be regarded as better for encoding categories rather than real variables. Nevertheless the list … is a reasonable first try and will lead to some useful results on graph construction. — Cleveland and McGill [16, p. 532]

The Semiology of Graphics would not be published in English until 1983, and as graphic displays of information became prevalent in American statistical journals in the early 1970s, calls were made for more systematic inquiry. A "theory of graphical methods" was needed [21, p. 5] in order to overcome the state of "dogmatic and arbitrary" design guidance of the time [45, p. 29]. William Cleveland and Robert McGill were statisticians at Bell Labs when they answered this call, publishing a series of empirical studies in the Journal of the American Statistical Association (JASA) which they described as theory for the relative accuracy for a set of *elementary perceptual tasks* readers perform to extract the values of real variables from statistical graphs [16]. In subsequent years, Cleveland and McGill refined their terminology, replacing *perceptual tasks* [16] with *graphical-perceptual tasks* [17], *basic graphical judgments* [18], and finally, *elementary codes* [19], with influential publications spanning venues of statistics, HCI, and popular science. Claims made in their initial 1984 work were tested by additional experiments and deeper engagement with contemporaneous theories of vision, resulting in the much refined 1987 publication ranking accuracy of an expanded set of *elementary codes* (Fig. 2.2b).[2] These codes describe channels available for mapping quantitative information to graphic form. In this sense, the authors re-articulated the visual variables described by Bertin [5, 6] and further ordered them according to human accuracy in making quantitative relational judgments. Cleveland and McGill's variables do not match those of Bertin and, however, are admittedly neither exhaustive nor mutually exclusive [16, p. 532]. One explanation for this discrepancy is their having conceived of the codes on the basis of their personal experience with statistical graphs, while Bertin set out to theorize a structure that could account for the visual-spatial properties of all graphic marks on 2D surfaces.

Cleveland and McGill's approach was partially deductive—structured a posteriori from personal experience and perceptual theory (e.g., [74]) and inductive, generalizing from reviews of psychophysical experiments (e.g., [4]), and their own original studies. It is perhaps most accurate to characterize their studies as tests of Bertin's hypotheses for the appropriate visual variables for quantitative perception. The experimental task asked participants—presented with two marked graphic components—to indicate "what percentage the smaller is of the larger" (p. 539), an operationalization of Bertin's test for quantitative perception: "ask the reader the value of the larger sign if a value of one is attributed to the smaller sign" [6, p. 69].

[2] Nonetheless, the more preliminary 1984 publication remains the most widely cited of their works, with nearly eight times as many citations as the 1987 elaboration [as reported by Google Scholar and Web of Science, January 2021]. This observation reinforces the importance of tracing the intellectual history of theoretical works to find their most mature form and should serve as a warning against cherry-picking references.

While Bertin reasoned that only the planar dimensions (spatial location) and size can adequately communicate quantitative information, Cleveland and McGill give us the relative accuracy of ten encodings for the same task. Their experimental data support Bertin's hypothesis that spatial location (e.g., position along common scale, position along non-aligned scales) can carry this information most accurately. If *length* is imputed as the size variation of a line [6, p. 71] and area the size variation of a point, then the data support Bertin's conclusions about the size variable, but not in relation to direction (Bertin's orientation for line) or angle (potentially construed as shape). There is enough discrepancy suggested in the empirical results to warrant further scrutiny of Bertin's criteria for judging a variable as applicable to a particular level and of the experimental tasks themselves.

Four years later, Northwestern University psychologists David Simkin and Reid Hastie offered JASA a contextualization of Cleveland and McGill's elementary codes, under a framework of information processing psychology [72]. Simkin and Hastie emphasized that performance of graphical perception depends not only on the way information is encoded but also on the judgment tasks performed by the human beings for whom the graphs are intended. Building upon Follettie [26], they differentiated between measurement, discrimination, proportion, and comparison judgments (Fig. 2.3a). It is important to note that all of Cleveland and McGill's studies used proportion judgments. Follettie, and later Simkin and Hastie, brought awareness to a whole new range of judgment tasks for which statistical graphs are used. Most importantly, they demonstrated that choosing a graphic mapping for a variable of data should not only depend on the data type (Bertin's level of organization) but also on the judgment task the designer wants the reader to perform. They offered empirical demonstrations of the interaction between elementary codes and judgment tasks (e.g., comparison judgments were most accurate with simple bar charts (position along common scale) while proportional judgments were most

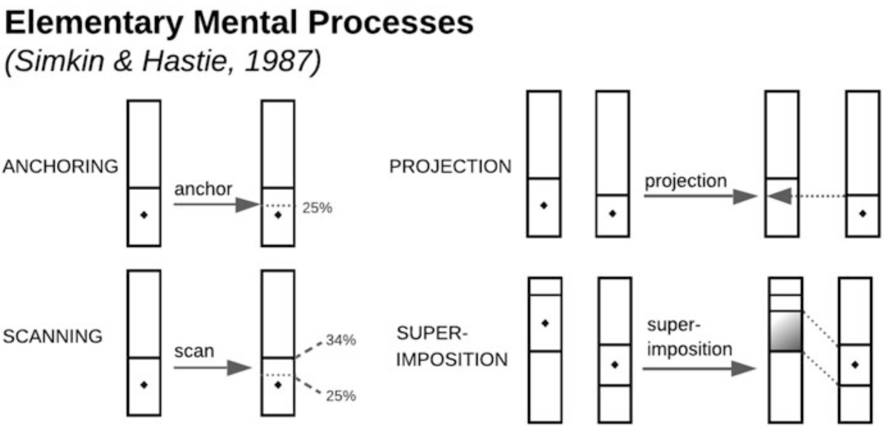

Fig. 2.3 Schematic diagram of Simkin and Hastie's theorized Elementary Mental Processes, adapted from (1987)

accurate with simple pie charts (angles)). Moving beyond encoding, they theorized four *elementary mental processes* that could—in an algorithmic sense—explain relative error and response rates across tasks (Fig. 2.3b). The elementary mental processes can be construed as visual data extraction steps: ordered in procedures that are executed by the perceptual system in order to accomplish a judgment task.

Over the course of the 1980s, the use of statistical graphics in publishing and data analysis surged with the development of software packages that made simple visualizations accessible for personal computer users. The cross-fertilization of empirical research between perceptual psychology and statistics demonstrated how demand for design recommendations can drive applied research questions that in turn inspire basic science research. Though the decade began with a focus on mapping information to visual forms, it would end with sophisticated hypotheses about how such mappings would interact with tasks, governed by perceptual rules, to elicit comprehension.

2.2.3 The Rise of Process Theories

Prior to 1980, there had been very little systematic research on the psychology of graph comprehension [84]. Over the course of the 1980s, methods and theories from cognitive psychology began to permeate the community in statistics concerned with graphical perception. Simkin and Hastie, notably, were psychologists, though they published their seminal work in the Journal of the American Statistical Association (JASA) rather than a journal of applied cognition or perception. Their contribution stood in direct conversation with the earlier work of Cleveland and McGill in the same venue. In [43], psychologist Stephen Kosslyn published in JASA a review of five books on charts and graphs, including Bertin [6], Tufte [77], and Chambers [13]. Rather than a straightforward critique however, Kosslyn offered a thorough primer on relevant concepts from cognitive psychology contextualized with respect to graph reading. He provided a sketch of contemporary visual information processing [52] and the distinction between short- and long-term memory [3, 47] before addressing the extent to which the practical guidance offered by each book comported with aspects of cognitive theory. Although its citation count pales in comparison to the aforementioned works, the importance of Kosslyn's contribution cannot be overstated. In this cross-disciplinary fertilization, he offered—like Bertin—a structure for thinking about the scope of what questions might be asked of graphical performance. He shared a simple (conceptual, process) model of visual information processing (Fig. 2.4) in which graph perception would be situated. To an application-focused community of statisticians *using* graphics, he brought a concise summary of relevant psychological constructs. While previous efforts focused on structural questions of encodings and tasks, Kosslyn drew attention to the way that graph reading unfolds as a *process*.

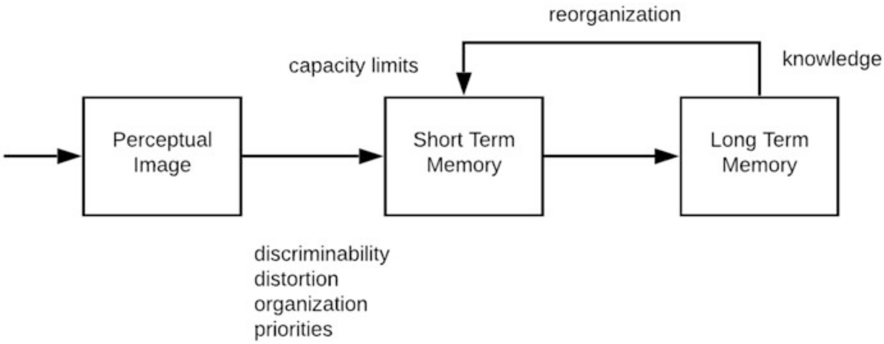

Fig. 2.4 A process description of visual information processing, adapted from [44]. The same figure appeared (without linguistic annotation of the important characteristics) in [43]

But Kosslyn's influence would not end there. In [44] he published an analytic scheme for deconstructing graphs[3] into constituent parts, which could then be analyzed at the levels of: syntactics (configuration of marks), semantics (the meaning that arises from configurations), and pragmatics (conveyance beyond direct interpretation of symbols). This contribution was more structural than procedural, offering a schema for evaluating graphs with respect to acceptability principles reasoned from cognitive theory. But in doing so, he would make reference to a forthcoming publication from his former graduate student Steven Pinker, one that would go on to stand as the most widely cited theory of graph comprehension.

2.2.3.1 A Theory of Graph Comprehension: Steven Pinker

While experimental psychologist Steven Pinker is most widely recognized for his popular science books on language and human nature, he got his start in the late 1970s as a doctoral student studying visual cognition with Stephen Kosslyn at Harvard. His chapter "A Theory of Graph Comprehension" in the book *Artificial Intelligence and the Future of Testing* would influence research on the design and function of visual-spatial displays across psychology, education, and computer science for decades [62]. In fact, the ideas were influential *before* publication, with earlier versions of the theory cited via MIT technical reports from the early 1980s.

Pinker's theory consists of a series of computational processes that propagate representations of information across components of a theorized human cognitive architecture (Fig. 2.5). He proposes that graph interpretation begins with construction of a *visual array*: a relatively raw, minimally processed representation of the

[3] Kosslyn makes a distinction between charts (specifying discrete relations between discrete entities) and graphs (a more constrained form, requiring at least two scales associated via a "paired with" relation).

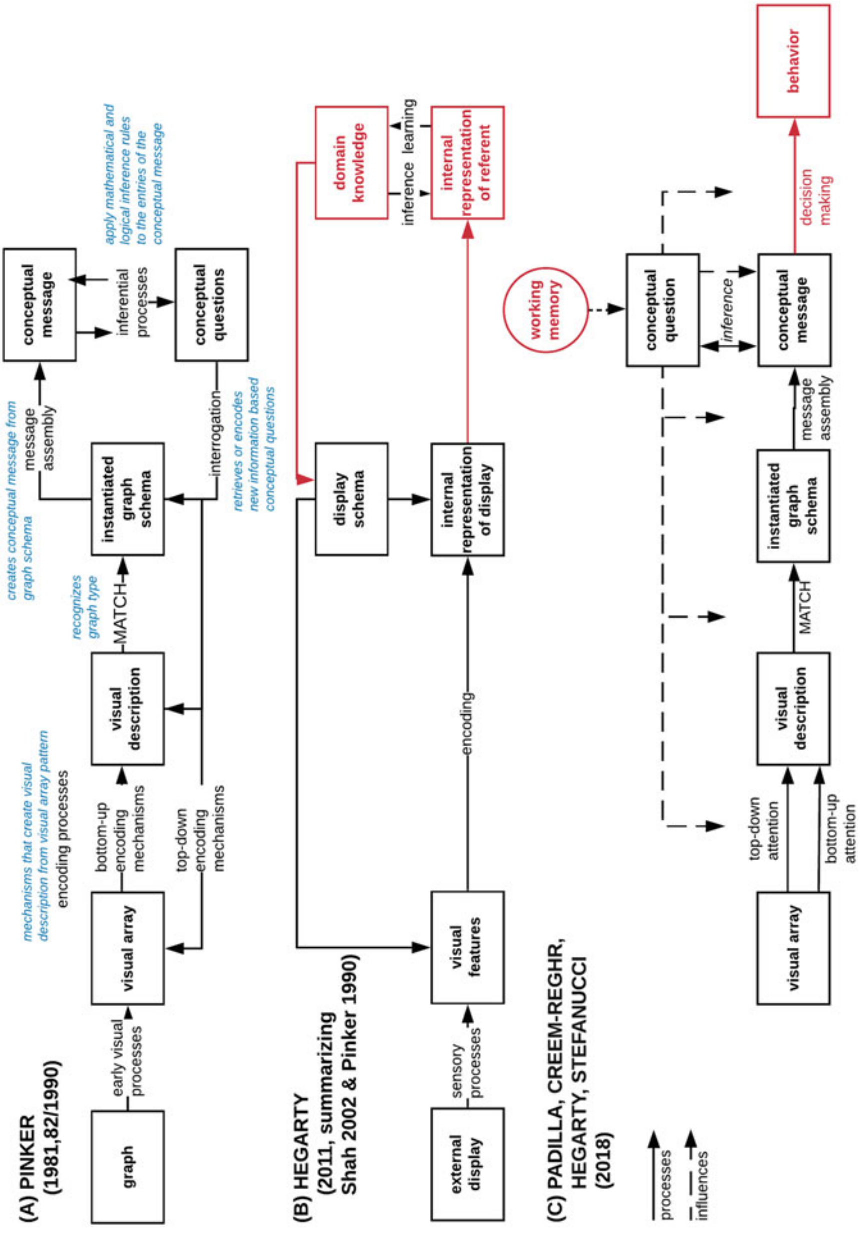

Fig. 2.5 Three versions of Information Processing accounts of Graph Comprehension. Italic annotations in blue indicate clarifications, and red indicates changes from prior models. In reading these diagrams, it is important to recognize they represent processes, not components. The boxes in Pinker, for example, indicate representations of information, not theorized cognitive structures, like working memory or executive control. The diagrams are not schematics for the structure of a cognitive system, but schematics of how information is processed, and care must be taken to avoid inadvertently reifying them into component structures, which might serve an *implementation* level of analysis

information made available to the nervous system via patterns of intensity on the retinas. The visual array is then *encoded* into a *visual description*: a symbolic, structural representation of the scene in a form more efficient for computation with knowledge in memory. A *MATCH process* then compares the visual description with the contents of memory in order to select the correct *graph schema*—a sort of placeholder indicating the structural relation of information for that particular class of graph. Once *instantiated*, information from the visual description is structured according to the relations of the selected schema. By this point, the external representation of the graph has been transformed into an internal representation in some structured, symbolic form that can be interrogated (searched) in order to extract information. Pinker uses the term *conceptual question* to refer to the information the reader wishes to derive from the graph and *conceptual message* the information that is actually extracted. A *message assembly process* searches the instantiated graph schema for information to translate to the form of the *conceptual message*. But processing capacity limitations prevent all the information from being automatically translated to messages. Rather, the *interrogation* process searches the graph schema for information matching the conceptual question. If it is found, message assembly takes over. But if not, *interrogation* can traverse the prior stages of representation (the visual description, then visual array) until the desired information is found, a top-down search that may require re-encoding the visual array. Finally, Pinker appeals to a general class of (logical, mathematical, and qualitative) *inferential processes* that operate on the conceptual message in service of answering the conceptual question.

Pinker's approach was deeply situated in the tradition of information processing, expressing an orientation toward a computational theory of mind. His explanation functions at Marr's *algorithmic* level of analysis—specifying representations and procedures for transforming them [52]. He offers an exceptionally detailed account of the properties of the representations he proposes (especially the visual description) and how they comport with cognitive theory in vision, memory, and attention. The 1990 publication is not an easy read, and it is my personal opinion that its scope is often misunderstood and contribution inadvertently reified *as* its diagrammatic representation of information processing.[4] Figure 2.5a is adapted from Pinker's Figures 4.14 and 4.19 which he characterizes as "representing the flow of information specified by the current theory" [62, p. 104]. The diagram depicts the order of representations and names of processes that transform them but fails to adequately describe re-encoding of the visual array (by re-attending to the graph) or the timecourse of decay of any representation based on the capacity limits of short (i.e., working) memory (e.g., [62, p. 89]). This leads to the misconception that Pinker does not address the role of working memory or proposes that an entire

[4] Just as we are drawn to graphs of empirical results, we are drawn to diagrams of theoretical offerings. The readers are warned against assuming that a diagram *entirely represents* a theoretical account, and writers encouraged to explicitly describe the representational role of diagrams in the scope of their theory.

graph is encoded in a single linear process. Rather, it is more appropriate to construe the diagrammatic representation as a snapshot of the flow of information through a single iteration of a bottom-up (perceptually driven) loop. We are similarly left wondering “where” in the mind his representations exist. This is not explicitly defined in the process diagram nor the text, but it can be reasonably inferred that all posited internal representations exist in short term (i.e., working) memory, as this is where processing would occur in the context of the cognitive theories he references (with the exception of the uninstantiated graph schema, likely in long-term memory).

Most importantly, justification for the theory rests on a single proposition: that graph comprehension exploits general purpose cognitive and perceptual mechanisms. Pinker’s chapter was not the culmination of decades of empirical experimentation with graphs, but rather, the application of contemporaneous theories of vision, memory, and attention to the phenomenon of graph comprehension. This statement is not offered in critique, but in observation of the variety of ways that theory is developed. In this case, refutation rests on change to theories of vision, attention, and memory or evidence that graph comprehension is sufficiently different from the phenomena used to construct those theories to warrant special purpose cognitive mechanisms.

2.2.3.2 A Construction-Integration Model: Shah and Colleagues

An alternative to refuting a theory is refining it, by elaboration (specifying detail) or contextualization (situating in larger scope). In the late 1990s and early 2000s, Priti Shah and colleagues arguably did both: zooming out to describe the iterations of information processing when comprehending a graph and zooming in to elaborate the influence of “top-down” factors.

While prior experimental work focused on the perceptual aspects of graph comprehension, Cognitive Psychologist Priti Shah’s mid-1990s dissertation work emphasized the role of *cognitive processes* in graph comprehension. Though contemporary Cognitive Science resists a precise delineation between perception and cognition, in graph comprehension a distinction is typically drawn between sources of information. Perception—information arriving via the senses—is referred to as “bottom-up” processing, while prior knowledge and computation over internal representations is referred to as “top-down” processing. Like Pinker, Shah, and her colleagues reasoned that graph comprehension would make use of general purpose cognitive processes rather than some special graphics engine in the mind. Drawing inspiration from Walter Kintsch’s well-regarded Construction-Integration Theory [41], Shah elaborated how the processes of constructing meaning with a graph might proceed in the same fashion as constructing meaning from text or linguistic discourse.

Along with Patricia Carpenter, Shah first drew attention to the timecourse of information processing when reading a graph [12, 67]. Prior perceptual accounts tended to emphasize holistic pattern recognition processes that allow the readers to make the sort of quick proportional judgments used in studies of graphical perception. Carpenter and Shah employed more complex tasks, asking the readers to describe graphs and answer comprehension questions. Performance on these tasks, accompanied by measurements of eye fixations, revealed a more iterative procedure was taking place: one that involved a serial identification of visual chunks, followed by inferences and reasoning, repeated until the task goal had been accomplished. Along with evidence of differential task performance based on prior knowledge of semantic content, their studies provided support for the claims that (1) successful graph interpretation depends not only on appropriate information-to-graphical encoding but also on prior knowledge and skill of the graph interpreter and (2) graph comprehension is an iterative, multi-stage process. Publications in 2002 drew more strongly from CI Theory, characterizing the timecourse of processing in terms of two phases: an initial *construction phase*, where visual chunks activate relevant prior knowledge and are integrated into a coherent representation, and an *integration phase*, where inferences are made over the (coherent) representation (Fig. 2.6a) [30, 68]. The phases follow in order, though can be repeated, and integration can be followed by further construction, as necessary (Fig. 2.6b).

The astute reader will ask how Shah's Construction-Integration Model relates to Pinker's [62] Theory of Graph Comprehension. The answer depends on one's interpretation of each text. In a 2005 review, Shah and colleagues describe their model as differing from Pinker's in that it specifies that prior knowledge (and in turn, expectations) is activated by the encoding of visual chunks, which serve as a top-down constraint on inferential processing [69]. Pinker also describes the activation of prior knowledge, though in slightly different terms. Specifically, the MATCH process "searches" prior knowledge in order to instantiate an appropriate schema (prior knowledge structure) for the type of graph being perceived [62, p. 101]. In this way, the prior knowledge of graph type is activated by the (symbolic) visual description of the graph (the encoded visual chunk). Since inferential processes act on the instantiated graph schema, this prior knowledge serves to constrain interpretation. What Pinker does not explicitly describe is the activation of prior *domain knowledge*, or any understanding the reader has about the information being represented by the graph, though a generous interpretation would be that he includes this constraining influence under the scope of *inferential processes* (p. 103), a catch-all term to describe all of the higher order processing (logical, mathematical, judgments, and decisions) that one performs *on* the instantiated graph schema. If Shah's *coherent representation* is equated with Pinker's *instantiated graph schema*, then the two accounts are congruous. They are consistent in appealing to general purpose mechanisms, to describing a serial process of encoding, some form of integration with prior knowledge, and inferential processing. They both posit the existence of internal representations: Pinker gives a specific account of a plausible form of these representations, Shah requires only that they exist, leaving the CI model with less explanatory power for mechanisms, but greater robustness to change in the perennial debate on the nature of internal representation. It is *this* author's reading that these two accounts of graph comprehension are highly compatible, serving to elaborate different aspects of graphical processing at different levels of

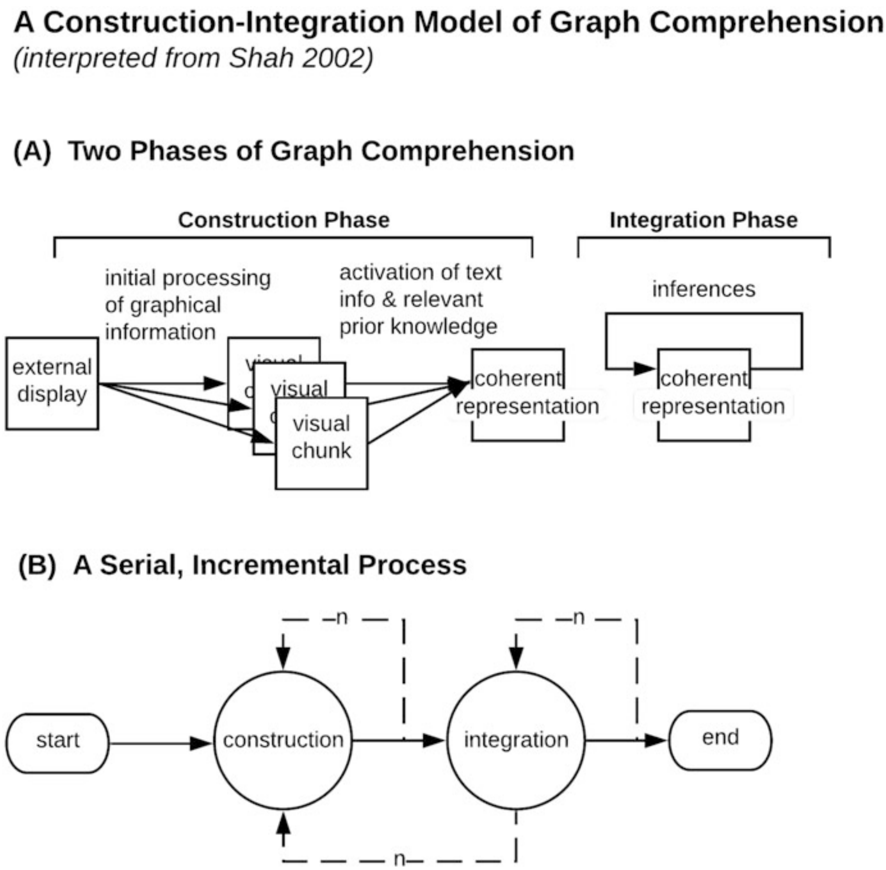

Fig. 2.6 A Construction-Integration Model of Graph Comprehension, derived from the text description in [30, 68]. (**a**) describes two distinct phases of comprehension: the first involves encoding visual chunks, while the second involves higher order cognitive processing over the working internal representation. (**b**) describes how integration follows some number of iterations of construction, before processing is either complete and ready for integration

specificity. While Pinker attends to a computationally plausible encoding structure for graphical information, Shah attends to the more global timecourse of processing and iterations of "perceptual" and "cognitive" efforts. They both offer testable predictions about how factors of the graphical display *and* the graph reader should differentially influence task performance.

2.3 The Landscape of Contemporary Research

Statistical graphics have never been more prevalent than they are today in scientific inquiry, business operations, or popular media. With such a wealth of applications, it is a good time to be a Visualization Psychologist but is not easy to *study* the psychology of visualization because as an applied area of inquiry, both students and scholars alike must navigate an opaque disciplinary milieu. The readers can find relevant empirical research in venues as distinct as journals and conferences of science or math education, learning science, information and library science, cognitive, educational, perceptual or (general) experimental psychology, vision science, cognitive science, and of course computer science—where the conference triad *InfoVIS, SciVIS, and VAST* claim some epistemic authority of the subject matter by virtue of naming rights.

In the two decades since Shah's Construction-Integration model, we have not seen similar overarching, general process accounts of comprehension. Rather, the researchers across these fields have progressively elaborated a complex ecosystem of factors that influence performance on graph comprehension tasks. We can organize these factors into three groups: those pertaining to the display, the individual, and the situation.

Display Factors The research on display characteristics tends to center on determining the most ideal encoding of information, a question of design. Bertin offered the first experientially deduced guidelines for mapping data to graphic marks [5, 6],[5] some of which were experimentally tested using relational judgment tasks and ranked by Cleveland and McGill [16, 19] and further extended by Mackinlay [51] who ranked encodings according to theorized perceptual accuracy for communicating quantitative, versus ordered, verses categorical data (see Fig. 2.2c). If humans were perceptual computers, this might be the crux of visualization psychology. But we are, of course, more delightfully nuanced creatures. Contemporary research has demonstrated that effectiveness of encodings depends not only on the capacity of a particular type of mark to carry a certain type of information but also on what *about* that information the designer wants the reader to perceive most effortlessly. Ensemble encoding, for example, relies on characteristic performance of the visual system to inform encoding choice when the goal is to facilitate, for example, identification of an outlier, versus recognition of a statistical mean, or apprehension of clusters within the data [75]. Design choices within a particular encoding strategy are nuanced as well, as evidenced by research on the use of color. Color hue has been shown to be particularly effective for encoding data for nominal or absolute value judgments, while color brightness is superior to hue when encoding the same data for *relative* judgments [10, 55]. The plot thickens—design choices become more complex—when visualizing more than one variable and the interactions between

[5] The oft-overlooked footnote to these heuristics is that the rankings are meant to apply when the reader's *task* is an "elementary reading" (extracting a specific value).

encoding strategies need be considered. Smart and Szafir recently demonstrated that the shape of a graphic mark significantly influences perception of color and size [73]; whatever the designer's most informed intentions, their efforts can be thwarted by interactions between decisions they make. Similarly, visual saliency (how "attractive" an area is to the eye) has been shown to influence how humans attend to visual stimuli [38]; though recent efforts to computationally reconcile bottom-up saliency models top-down "cognitive" models have proven ineffective at predicting gaze behavior [48]. While display characteristics were the focus of the earliest research in graph comprehension, they receive no less attention in modern research efforts. Designers need practical guidance on when and how to use animation [8, 79] and 3D [68], how to use signals or instructions to augment a display and scaffold comprehension [1, 28, 42, 54], and how to use interaction most effectively [61, 66]. Since the time of Cleveland and McGill, research on display characteristics has become increasingly nuanced, revealing more factors that influence how a display should be designed and the interactions between them.

Individual Factors Research on individual differences, or factors that give rise to differential performance with the same graphic display, is most common in cognitive and educational psychology and learning science. As Carpenter and Shah argued, "individual differences in graphic knowledge should play as large a role in the comprehension process as does variation in the properties of the graph itself" [12, p. 97]. But what is meant by *graphic knowledge*? In empirical work, graph knowledge is tightly entwined with graph reading abilities and expertise. The terms graphicacy, graphical literacy, graph sense, graphical competence, and representational competence are used throughout the literature in psychology and education to refer to a reader's ability to understand (and potentially create) information displayed graphically. If graph comprehension is the act of deriving meaning from a graph, then *graphicacy* is its educational flip side: the ability to perform a graph comprehension task. Some have treated this ability as a foundational step in cognitive development, akin to numeracy and literacy [31]. Others treat the ability as a practice, implicating the importance of experience and socio-cultural influences [64, 65]. In education in particular, the researchers have pursued general learner characteristics that might serve as pre-requisites or predictors of these graphing abilities, including mathematical ability [23], working memory [12], and spatial reasoning [81]. Ulrich Ludewig's recent doctoral dissertation offers a thorough reconciliation between perspectives of graph comprehension and graphicacy [50]. It is slightly easier to differentiate between ability and knowledge with respect to specific graphs, for example, domain knowledge of the information represented in a particular graph, and knowledge of that particular representation's graphical formalisms. The act of graph reading requires that we use our knowledge of a graph's formalisms to perform some task (e.g., extract a value, detect a trend), thereby "learning" something about the domain. In my own research, I have demonstrated that this procedure is not reciprocal. It is much more difficult to use prior knowledge of a domain to "reverse engineer" understanding of a graphical formalism, such as may be required to understand an unfamiliar or unconventional type of graph [28, 29]. A reader's understanding of the concepts represented in a graph has been shown to guide not only the reader's interpretation of the display [63] but early perceptual processing as well [68]. In some cases, a reader's expectations seem to "inoculate" them from true relations presented in the data or lead them to over or underestimate the magnitude of relations. Conversely, domain knowledge has been shown to support comprehension by making the readers more likely to ignore "noise" in data [86]. More recently, Jessica Hullman and colleagues have explored the role of prior beliefs [37, 40] and even judgments of expectations of others [36] on graph interpretation. Taken together, the research on characteristics of individuals has provided strong evidence for "top-down" influences on graph comprehension.

Situational Factors Factors that change comprehension performance of an individual with a particular display depending on the *situation* are the least structured, thus least understood pieces of this factorial puzzle. Affect (emotion) and motivation clearly influence human performance of any task, and although these are characteristics of an individual, we classify them as situational because they are more situationally variable—in the context of a repeated measures study, for example—than the relatively stable[6] factors like prior knowledge or ability. *Task* is the most studied situational factor, though it is at present a hierarchical concept poorly operationalized across the literature. The term "task demand" is used to indicate a variety of contextual factors, from a relatively low-level step of information extraction (i.e., a micro-step in a larger process, such as identifying a location of interest in a graph), to a specific task or goal provided to a reader in an experiment (e.g., extract a value, compare two points, characterize a trend), to the context of some cognitive activity (e.g., analyzing data, making a decision, forecasting, solving a problem), and to the communicative intent of the designer (e.g., to inform, educate, entertain, persuade, etc.). In the beginning, there was but a single task: Cleveland and McGill's proportional judgments [16, 19]. Folettie, followed by Simkin and Hastie, elaborated further judgments (measurement, discrimination, and (non-proportional) comparison) [26, 72]. Bertin also addressed tasks, proposing three "levels of reading" [6, p. 141]. Other tripartite classifications have been proposed in the same vein, all structuring how much of the depicted information the reader need attend to, and how explicit or precise their response should be [5, 6, 23, 31, 83]. In their application of ensemble encoding theories to visualization, Szafir and colleagues offer a parallel taxonomy of four tasks-types that require visual aggregation [75]. These can be partially but not entirely mapped onto the extant tripartite structures. The most complete deconstruction of the concept of task can be found in Brehmer and Munzner's, "Multi-Level Typology of Abstract Visualization Tasks," which surveyed an impressive volume of prior task frameworks in computer graphics and visualization, visual analytics, human–computer interaction, cartography, and information retrieval [9]. A fruitful undertaking for

[6] Variability, of course, depends on the scope of time under consideration.

visualization psychology would be to extend this typology to include the tripartite classifications that grew out of education, the lower level tasks elaborated in vision science, and higher level "communicative context" that is evident in the structure of the field of visualization itself [27]. A strong underlying assumption of much research in graph comprehension (and visualization writ-large) is that the graph designer's goal is to clearly communicate, "the truth" of some data to the reader. Thus, the graph should be maximally informative and minimally difficult—the graphical equivalent of Grice's maxims for communication. But research in learning science has taught us that sometimes difficulty is *desirable*. Perhaps if my graph is for *learning*, I might encode data differently so as to scaffold a reader's process of discovery and more deeply engage with the data. Alternatively, if the context of my communication is *persuasion*, I might use more signals to direct reader's attention than I would if the context were exploratory analysis. The role of communicative context is seen structurally through the emergence of specialized workshops at the IEEE VIS conference but has not yet been systematically investigated across a full range of communicative tasks. My own theoretical intuition—reasoned from design experience and engagement with the literature—is that situational factors are those that present mediating or moderating influences on other individual and display characteristics, at either the time of design or comprehension.

A primary challenge facing designers and researchers alike is the sheer number of factors found to influence comprehension and the fact that they are typically studied in limited clusters, inconsistently operationalized between studies and across disciplines. This makes it difficult to conceive of the complex interactions that may exist between factors and how to go about constructing nuanced guidelines for designers. The most comprehensive summaries of factors can be found in [31, 33, 70] and [35], which features a concise set of empirically grounded principles for display design that would make a useful addition to the wall of any graph designer.

2.4 What Remains to Be Discovered

The good news is that "the state of our (sub) discipline is strong." The bad news is that it is difficult to navigate and even more difficult to *integrate*. In the two decades since the last publication of a general process theory of graph comprehension [68], the march of empirical research has only quickened, offering insight into factors that affect graph comprehension, but in forms too piecemeal to be fruitfully and consistently applied. There are myriad open questions to be answered, from how exactly factors interact to influence performance to how performance is expressed in different forms of cognitive activity: decision-making vs. problem solving, forecasting, learning, or creative construction. We need to explore our boundaries: how does interaction with the narrowly defined class of "graphs" compared to the broader class of diagrams or external representations, in general? (see [14, 15] for thorough treatments). And our field too must address the challenge of traversing "lower levels" of explanatory analysis: there is a tremendous gulf of explanation between conceptual models of graph comprehension and understanding of how these processes are enacted in the body.

Hegarty [71] and more recently Padilla [56] have convincingly argued for the importance of *cognitive models* in guiding visualization research. Hegarty suggests they are useful for predicting the effectiveness of designs and informing design decisions. Padilla argues that cognitive models can be used to promote innovation and evaluate validity of empirical research designs. In sum, they can bridge an important gap and presuming they are communicated in an appropriate venue, well-articulated models can help ensure that the "state of the art" in basic research is available to guide applied efforts in design and instruction. But what kinds of models do we need, and what makes a model *cognitive*?

Those seeking easy answers to these questions will fall quickly down a philosophical rabbit hole. Models in science come in all shapes and sizes, with differing levels of analysis and varieties of explanation. In the social and behavioral sciences alone, one finds component and structural models, conceptual models, computational models, and task-analytic and mathematical models. Models differ in what *aspect* of a phenomenon they explain (e.g., structures, relationships, processes), how they are justified (e.g., by phenomenological, experimental or task-analytic empirical evidence, by logic or appeal to reason), and the way they are represented (conceptually: typically via words and diagrams or computationally: via math and/or computer programs). The importance of clearly conceptualizing and subsequently articulating the scope and purpose and form of a model cannot be overestimated, as the failure to do so can have tragic consequences for the intellectual trajectory of a field.

Take, for example, [62] Theory of Graph Comprehension. Setting aside for the moment that it is characterized as a theory and not a model,[7] a quick inspection of its diagrammatic representation (Fig. 2.5a) will reveal no mention of memory. Does this mean that Pinker believed memory was not involved in graph comprehension? *No*, it means that the reader needs clarification on what aspect of the phenomenon Pinker's model explains: a propagation of representations and the processes that transform them. Close reading of the accompanying text reveals what was likely obvious to readers at the time: all of the representations and processing take place *in* some form of memory. Pinker might have chosen to represent this in the diagrams by locating the representations (boxes) inside other graphics representing memory structures. This would have been advantageous for subsequent theorists looking to position their own ideas in relation to his but would also have changed the type of model, from the flow information processing to the flow of information processing *and* component structures—taking on an additional Marrian level of analysis [52]. In applying Pinker's model to a specific cognitive activity (decision-making), Padilla and colleagues have done well to clearly articulate the role of memory,

[7] Theories are typically treated as superordinate to models, though their exact relation is a topic of debate in philosophy of science.

as well their interpretation of the construct of memory itself [57], implicating a multiple component conception where "a multicomponent system (...) holds information temporarily and mediates its use in ongoing mental activities" [20, p. 1160]. While these details may be superfluous for those keen to *apply* the model, they are absolutely essential for the ongoing intellectual dialogue expressed via works of scholarship that move our science forward. Imagine next year a groundbreaking study is published in a journal of experimental psychology that questions the multicomponent conception of working memory, supporting a rival account with implications for how visual attention is directed. Changes to the underlying constructs on which a model or theory rests should necessitate its re-evaluation, no different from the need for testing and upgrading software when the libraries on which they are built mature.

The obvious difficulty is that constructs are transient, under-specified, and certainly not versioned like packages of code. Too often the precise conceptualization of constructs is held as tacit knowledge instantiated in encapsulated research labs, propagated through limited networks via the exchange of students and postdoctoral scholars.[8] Too little space is allocated in our written scholarship to descriptions of what we *specifically mean* by the terms we use, a symptom of a drive toward innovation and novelty over depth of explanation. I propose that in theoretical scholarship we should strive to be a little more like academic philosophy, where precision and justification in language is not only valued but demanded. We should be novel in our applications, but religiously rigorous in our theory. Models and theories should exist in direct dialogue with those that come before, explaining *exactly* how and why they differ and offer sufficiently impactful differences to be worthy of inclusion in the scientific canon.

In this onerous challenge stands a role for visualization psychology: as a mediator between disciplines (computer science, psychology, and education) and between professions (basic and applied research, design, and instruction). As a community, visualization psychology can position itself at the intersection of these goal-driven efforts and moderate the construction of *reference models*, intended to integrate theory across disciplines and levels of analysis that is specifically related to our phenomena of interest. We need not be concerned with explaining precisely how memory or attention are instantiated by the body but should take responsibility for maintaining enough awareness of the progression of those basic theories, so we can apply and as needed update our own models of how such cognitive phenomena drive the performance of graph (and visualization) comprehension.

[8] see Kaiser [39] for a fascinating intellectual history of this phenomenon with respect to dialects of Feynman diagrams.

References

1. C. Acarturk, C. Habel, and K. Cagiltay. Multimodal Comprehension of Graphics with Textual Annotations: The Role of Graphical Means Relating Annotations and Graph Lines. In G. Stapleton, J. Howse, and J. Lee, editors, *Diagrammatic Representation and Inference*, Lecture Notes in Computer Science, pages 335–343. Springer Berlin Heidelberg, 2008.
2. American Statistical Association. Joint Committee on Standards for Graphic Presentation. *Publications of the American Statistical Association*, 14(112):790–797, 1915.
3. J. R. Anderson and G. H. Bower. *Human Associative Memory*. Human associative memory. V. H. Winston & Sons, Oxford, England, 1973.
4. J. C. Baird. *Psychophysical Analysis of Visual Space*. Pergamon Press, 1970.
5. J. Bertin. *Sémiologie Graphique. Les diagrammes—les reseaux—les cartes.* Mouton, The Hague, 1967.
6. J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, Madison, WI, 1983.
7. S. Bonin. Le développement de la graphique de 1967 à 1997. *Cybergeo : European Journal of Geography*, Nov. 2000.
8. J.-M. Boucheix and E. Schneider. Static and animated presentations in learning dynamic mechanical systems. *Learning and Instruction*, 19(2):112–127, 2009.
9. M. Brehmer and T. Munzner. A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.
10. L. A. Breslow, J. G. Trafton, and R. M. Ratwani. A perceptual process approach to selecting color scales for complex visualizations. *Journal of Experimental Psychology: Applied*, 15(1):25–34, Mar. 2009.
11. W. C. Brinton. *Graphic methods for presenting facts*. The Engineering magazine company, 1914.
12. P. A. Carpenter and P. Shah. A Model of the Cognitive and Perceptual Processes in Graph Comprehension. *Journal of Experimental Psychology: Applied*, 4(2):75–100, June 1998.
13. J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey. *Graphical Methods for Data Analysis*. Wadsworth Publ. Co, Belmont, CA, USA, 1983.
14. P. C.-H. Cheng. What Constitutes an Effective Representation? In M. Jamnik, Y. Uesaka, and S. Elzer Schwartz, editors, *Diagrammatic Representation and Inference*, Lecture Notes in Computer Science, pages 17–31, Cham, 2016. Springer International Publishing.
15. P. C.-H. Cheng. A Sketch of a Theory and Modelling Notation for Elucidating the Structure of Representations. In A.-V. Pietarinen, P. Chapman, L. Bosveld-de Smet, V. Giardino, J. Corter, and S. Linker, editors, *Diagrammatic Representation and Inference*, volume 12169, pages 93–109. Springer International Publishing, Cham, 2020.
16. W. S. Cleveland and R. McGill. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.
17. W. S. Cleveland and R. McGill. Graphical Perception and Graphical Methods for Analyzing Scientific Data. *Science*, 229(4716):828–33, Aug. 1985.
18. W. S. Cleveland and R. McGill. An Experiment In Graphical Perception. *International Journal of Man-Machine Studies*, 25(5):491–500, Nov. 1986.
19. W. S. Cleveland and R. McGill. Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data. *Journal of the Royal Statistical Society Series a-Statistics in Society*, 150:192–229, 1987.
20. N. Cowan. The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, 24(4):1158–1170, Aug. 2017.
21. D. R. Cox. Some Remarks on the Role in Statistics of Graphical Methods. *Applied Statistics*, 27(1):4, 1978.
22. F. E. Croxton and R. E. Stryker. Bar Charts versus Circle Diagrams. *Journal of the American Statistical Association*, 22(160):473–482, Dec. 1927.

23. F. R. Curcio. Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, 18(5):382–393, Nov. 1987. Publisher: National Council of Teachers of Mathematics.
24. W. C. Eells. The Relative Merits of Circles and Bars for Representing Component Parts. *Journal of the American Statistical Association*, 21(154):119–132, June 1926.
25. Y. Engelhardt. *The Language of Graphics*. PhD thesis, University of Amsterdam, 2002.
26. J. Follettie. Real-World Tasks of Statistical Graph-Using and Analytic Tasks of Graphics Research. *unpublished paper presented at the annual meeting of the National Computer Graphics Association, Anaheim, CA.*, 1986.
27. A. R. Fox. A Psychology of Visualization or (External) Representation? In *Proceedings of the Workshop on Visualization Psychology @ IEEE VIS*, 2020.
28. A. R. Fox and J. D. Hollan. Read It This Way: Scaffolding Comprehension for Unconventional Statistical Graphs. In P. Chapman, G. Stapleton, A. Moktefi, S. Perez-Kriz, and F. Bellucci, editors, *Diagrammatic Representation and Inference*, Lecture Notes in Computer Science, pages 441–457. Springer International Publishing, 2018.
29. A. R. Fox, J. D. Hollan, and C. M. Walker. When Graph Comprehension Is An Insight Problem. In *Proceedings of the Annual Conference of the Cognitive Science Society*, Montreal, Canada, 2019.
30. E. G. Freedman and P. Shah. Toward a model of knowledge-based graph comprehension. In *Diagrammatic Representation and Inference*, pages 18–30, 2002.
31. S. N. Friel, F. R. Curcio, and G. W. Bright. Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education*, 32(2):124–158, 2001.
32. D. J. Gillan and R. Lewis. A Componential Model of Human Interaction with Graphs: 1. Linear Regression Modeling. *Human Factors*, 36(3):419–440, Sept. 1994.
33. N. Glazer. Challenges with graph interpretation: a review of the literature. *Studies in Science Education*, 47(2):183–210, Sept. 2011.
34. R. L. Harris. *Information Graphics: A Comprehensive Illustrated Reference*. Oxford University Press, 1999.
35. M. Hegarty. The Cognitive Science of Visual-Spatial Displays: Implications for Design. *Topics in Cognitive Science*, 3(3):446–474, 2011.
36. J. Hullman, E. Adar, and P. Shah. The Impact of Social Information on Visual Judgments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1461–1470, New York, NY, USA, 2011. Association for Computing Machinery.
37. Hullman, J. , Kay, M., Kim, Y. , and Shrestha, S. Imagining Replications: Graphical Prediction & Discrete Visualizations Improve Recall & Estimation of Effect Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):446–456, Jan. 2018.
38. L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews. Neuroscience*, 2(3):194–203, Mar. 2001.
39. D. Kaiser. *Drawing theories apart: the dispersion of Feynman diagrams in postwar physics*. University of Chicago Press, Chicago, 2005.
40. Y.-S. Kim, L. A. Walls, P. Krafft, and J. Hullman. A Bayesian Cognition Approach to Improve Data Visualization. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, Glasgow Scotland Uk, May 2019. ACM.
41. W. Kintsch. Modeling comprehension processes: The construction-integration model'. In *Comprehension. A paradigm for cognition.*, pages 93–120. Cambridge University Press, Cambridge, 1998.
42. N. Kong and M. Agrawala. Graphical overlays: Using layered elements to aid chart reading. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2631–2638, 2012.
43. S. M. Kosslyn. Graphics and Human Information Processing: A Review of Five Books. *Journal of the American Statistical Association*, 25(4):R134–R136, 1985.
44. S. M. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3):185–225, 1989.

45. W. Kruskal. Visions of Maps and Graph. *Auto- Carto II, Proceedings of the International Symposium on Computer Assisted Cartography, ed. J. Kavaliunas, Washington, D.C.: U.S. Bureau of the Census and American Congress on Survey and Mapping.*, pages 27–36, 1975.
46. J. Larkin and H. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 99:65–99, 1987.
47. P. H. Lindsay and D. A. Norman. *Human Information Processing*. Academic Press, New York, second edition, 1977.
48. M. A. Livingston, L. E. Matzen, A. Harrison, A. Lulushi, M. Daniel, M. Dass, D. Brock, and J. W. Decker. A Study of Perceptual and Cognitive Models Applied to Prediction of Eye Gaze within Statistical Graphs. In *ACM Symposium on Applied Perception 2020*, pages 1–9, Virtual Event USA, Sept. 2020. ACM.
49. G. L. Lohse. A cognitive model for understanding graphical perception. *Human-Computer Interaction*, 8(4):353–388, 1993.
50. U. Ludewig. *Understanding Graphs: Modeling Processes, Prerequisites and Influencing Factors of Graphicacy*. PhD thesis, Universität Tübingen, Tübingen, Germany, 2018.
51. J. Mackinlay. Automating the Design of Graphical Presentations of Relational Information. *ACM Transactions on Graphics*, 5:110–141, 1986.
52. D. Marr. *Vision*. The MIT Press, July 1982.
53. M. Massironi. *The Psychology of Graphic Images*. Psychology Press, 2001.
54. P. D. Mautone and R. E. Mayer. Cognitive aids for guiding graph comprehension. *Journal of Educational Psychology*, 99(3):640–652, 2007.
55. D. H. Merwin and C. D. Wickens. Comparison of Eight Color and Gray Scales for Displaying Continuous 2D Data. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 37(19):1330–1334, Oct. 1993.
56. L. M. Padilla. A Case for Cognitive Models in Visualization Research : Position paper. In *2018 IEEE Evaluation and Beyond—Methodological Approaches for Visualization (BELIV)*, pages 69–77, Oct. 2018.
57. L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, 3(1):29, July 2018.
58. G. Palsky. Jacques Bertin, from classical training to systematic thinking of graphic signs. *Cartography and Geographic Information Science*, 46(2):189–193, Mar. 2019.
59. D. Peebles and P. C. H. Cheng. Extending task analytic models of graph-based reasoning: A cognitive model of problem solving with Cartesian graphs in ACT-R/PM. *Cognitive Systems Research*, 3(1):77–86, Mar. 2002.
60. D. Peebles and P. C. H. Cheng. Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors*, 45(1):28–46, 2003.
61. W. A. Pike, J. Stasko, R. Chang, and T. A. O'Connell. The Science of Interaction. *Information Visualization*, 8(4):263–274, 2009.
62. S. Pinker. Theory of Graph Comprehension. In R. Freedle, editor, *Artificial Intelligence and the Future of Testing*, pages 73–126. Erlbaum, Hillsdale, NJ, 1990.
63. Y. Postigo and J. I. Pozo. On the Road to Graphicacy: The Learning of Graphical Representation Systems. *Educational Psychology*, 24(5):623–644, Oct. 2004.
64. W.-M. Roth. Toward an Anthropology of Graphing. In *Toward an Anthropology of Graphing: Semiotic and Activity-Theoretic Perspectives*, W.-M. Roth (Ed), pp. 1–21. Springer, Netherlands, 2003. https://doi.org/10.1007/978-94-010-0223-3_1
65. W.-M. Roth, L. Pozzer-Ardenghi and J. Y. Han. *Critical Graphicacy: Understanding Visual Representation Practices in School Science*. Springer, Netherlands, 2005. https://www.springer.com/gp/book/9781402033759
66. K. Sedig and P. Parsons. Interaction Design for Complex Cognitive Activities with Visual Representations: A Pattern-Based Approach. *AIS Transactions on Human-Computer Interaction*, 5(2):84–133, 2013.

67. P. Shah. A Model of the Cognitive and Perceptual Processes in Graphical Display Comprehension. In *Reasoning with diagrammatic representations II*, M. Anderson (Eds), pages 94–101. AAI Press, Menlo Park, CA, 1997.
68. P. Shah. Graph Comprehension: The Role of Format, Content and Individual Differences. In M. Anderson, B. Meyer, and P. Olivier, editors, *Diagrammatic Representation and Reasoning*, pages 173–185. Springer London, London, 2002.
69. P. Shah, E. G. Freedman, and I. Vekiri. The Comprehension of Quantitative Information in Graphical Displays. In P. Shah, editor, *The Cambridge Handbook of Visuospatial Thinking*, pages 426–476. Cambridge University Press, New York, NY, 2005.
70. P. Shah and J. Hoeffner. Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1):47–69, Mar. 2002.
71. P. Shah, R. Mayer, and M. Hegarty. Graphs as Aids to Knowledge Construction: Signaling Techniques for Guiding the Process of Graph Comprehension. *Journal of Educational Psychology*, 91(4):690–702, 1999.
72. D. Simkin and R. Hastie. An Information-Processing Analysis of Graph Perception. *Source Journal of the American Statistical Association*, 82(398):454–465, 1987.
73. S. Smart and D. A. Szafir. Measuring the Separability of Shape, Size, and Color in Scatterplots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–14, New York, NY, USA, 2019. Association for Computing Machinery.
74. S. S. Stevens. *Psychophysics*. John Wiley, New York, 1975.
75. D. A. Szafir, S. Haroz, M. Gleicher, and S. Franconeri. Four types of ensemble coding in data visualizations. *Journal of Vision*, 16(5):11–11, Mar. 2016.
76. S. B. Trickett and J. G. Trafton. Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. *Lecture Notes in Computer Science*, 4045 LNAI:286–300, 2006.
77. E. Tufte. *Visual Display of Quantitative Information*. Graphics Paper Press LLC, Cheshire, CT, first edition, 1983.
78. J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.
79. B. Tversky, J. B. Morrison, and M. Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, Oct. 2002.
80. S. Ullman. Visual routines. *Cognition*, 18(1):97–159, Dec. 1984.
81. Velez, M.C. , Silver, D. , and Tremaine, M. Understanding visualization through spatial ability differences. In *VIS 05. IEEE Visualization, 2005.*, pages 511–518, Oct. 2005.
82. R. von Huhn. Further Studies in the Graphic Use of Circles and Bars: I: A Discussion of the Eells' Experiment. *Journal of the American Statistical Association*, 22(157):31–36, Mar. 1927.
83. H. Wainer. Understanding Graphs and Tables. *Educational Researcher*, 21(1):14–23, Jan. 1992.
84. H. Wainer and D. Thissen. Graphical Data Analysis. *Annual Review of Psychology*, 32(1):191–241, 1981.
85. J. N. Washburne. An experimental study of various graphic, tabular, and textual methods of presenting quantitative material. *Journal of Educational Psychology*, 18(6):361–376, Sept. 1927.
86. J. C. Wright and G. L. Murphy. The Utility of Theories in Intuitive Statistics: The Robustness of Theory-Based Judgments. *Journal of Experimental Psychology: General*, 113(2):301–322, 1984.

Chapter 3
Mental Models and Visualization

Florian Windhager and Eva Mayr

Abstract Mental models are internal representations of external phenomena. During their interaction with visualizations, the users construct mental models to represent these visualizations internally, to visually reason on them and solve problems with them. This chapter synthesizes existing theories on mental models and visualization to discuss their role and relevance for the design of visualization systems. From a mental models perspective, we discuss two challenges of visualization design: (a) supporting the initial construction of mental models and (b) supporting the integration of information from multiple views by synchronous or sequential coherence techniques. We argue that the theory of mental models allows to understand visualization research and practice in a more unified fashion as an advanced model-building endeavor, operating on human computer ensembles engaged in "distributed cognition."

3.1 Introduction

Visualizations aim to amplify and augment human cognition and action in face of the challenges posed by complex data and information [7, 32, 43]. Accordingly, visualization research investigates the cognitive effects of interaction with visualizations to prove the value of novel techniques. Theoretical reflections of this practice build on different conceptualizations of cognitive entities and processes—from insights theory [34] to sensemaking approaches [38]—and more elaborate cognitive science perspectives [18, 26, 36]. Yet, looking at the state of research, visualization experts work with rather sketchy conceptions of the cognitive apparatus and its operational entities and did not extensively explore the question how users build up internal representations of data, reason with them, and assemble local insights into bigger internal pictures. This chapter aims to contribute to a better understanding of internal

E. Mayr · F. Windhager (✉)
Danube University for Continuing Education Krems, Krems an der Donau, Austria
e-mail: eva.mayr@donau-uni.ac.at; florian.windhager@donau-uni.ac.at

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_3

representations in visualizations by synthesizing relevant theories and models and by discussing their relevance for the design of visualizations.

To take first steps in this direction, we turn to the theory of mental models from a distributed cognition perspective [26] (Sect. 3.2). A mental model is an analog "small scale" internal representation of an external phenomenon [11]. During their interaction with visualizations—which arguably are models of (data about) external phenomena themselves [12, 30]—the users construct, adapt, and manipulate their own mental models of the external visualizations. They then use the external *and* the internal representation in a distributed fashion to reason with them, to take different perspectives on them, to organize and integrate detailed information, and to derive insights and infer hypotheses for the interpretation of the original, external phenomena.

Aside from providing an elegant cognitive-theoretical foundation for this kind of practice, we will argue that a mental model perspective also has implications for the future design of more complex visualization systems. One reason for this might be the knowledge about the mental efforts and costs of model building: the human working memory can only "maintain a limited amount of information (their capacity) for a finite period" [36] and aims to keep the required working memory resources, the *cognitive load*, to a necessary minimum. Even as visualizations allow to "offload" certain cognitive operations from the working memory to the external representation, their understanding requires their internal cognitive reconstruction and connection to existing knowledge—at least to a certain degree. Yet, construction efforts are known to be highly demanding in terms of working memory capacity and cognitive load and visualization design is well advised to support such construction processes.

Thus, we discuss related challenges for the design of visualization systems which can support the construction and elaboration of mental models in face of rising topic and data complexity (Sect. 3.3). This chapter aims to establish and consolidate a mental model perspective in visualization, to outline related design challenges, and to initiate a more systematic discussion and implementation of such techniques and studies.

3.2 Internal Representations

Visualization research develops external representations of data to support the internal efforts of human cognition, including decisions about behavioral responses and actions. Visualization systems thus are designed to serve as "amplifiers," "mediators," or "prostheses" for human cognition and action in face of challenging (e.g., abstract, multidimensional, or massive) constellations of data [2]. Given this widely accepted functional stance, visualization designers can benefit from existing knowledge about internal representations [27]: how users construct them when they interact with visualizations, how they manipulate them to reason on them, and how they offload cognition to the external representations. In the following, we will build

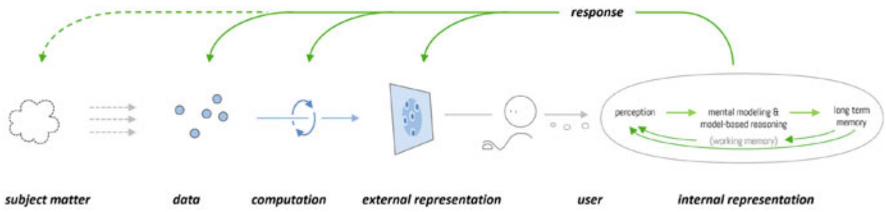

Fig. 3.1 Schematic line up of an extended cognitive system, showing the translation from complex data and subject matters (left) to an external representation (blue), which is (re)constructed as an internal representation by a user in working memory and stored in long-term memory (right)

on the theories of distributed cognition (Sect. 3.2.1) and mental models (Sect. 3.2.2) to discuss future challenges for both—visualization and theory development.

3.2.1 Distributed Cognition

According to the theories of situated and distributed cognition [26], human cognition cannot be understood without its ubiquitous amplification and augmentation by cultural artifacts (like tools, data carriers, calendars, or computers) and its constant social and cultural interaction and resonance with other processing units [21].

Interactive visualizations—as an advanced species of cultural artifacts—thus allow cognitive systems to expand into human–computer ensembles mediated by visual displays: in such an extended cognitive system (see Fig. 3.1), humans visually analyze complex objects of study by means of visual displays (as external representation on the computer) and continuously build up and manipulate corresponding mental models (as internal representations in their working memory).

On these (internally–externally) coupled representations, reasoning operations can take place as visual and cognitive manipulations. As the human working memory is limited in terms of storage and processing capacity, an extension with the external representation allows to process more complex information with less cognitive load. Results then can be stored as individual instances or generalized structures in long-term memory. Across the distributed architecture of such extended cognitive systems, basic transaction processes between internal representations on the user side and external representations on the visualization side play out in different combinations [27, p.1002]:

- If there is no prior knowledge or existing internal representation, internalization equals the construction of a new mental model based on a given external representation (i.e., learning). These processes require a bottom-up synthesis of visual patterns into various forms of internal representations—and a high amount of mental effort.

- As acquired (and generalized) structures, already existing internal representations (e.g., schemata, scripts) guide the processing and interpretation of external representations [38].
- Existing internal representations can not only be activated and simulated in working memory for visual reasoning operations but can also be augmented by external representations to work in joint as coupled and distributed representations, which equals the standard constellation of an visualization-mediated human–computer ensemble.

For visualization systems, numerous effects have been described how they support human working memory and cognitive processes [7, 26, 27, 37, 43] and have been grouped into four functions [17]. (1) External storage of detailed information on visual displays unburdens the working memory (from imagination, integration, and memorization) and allows the corresponding internal representation to remain lean and lightweight. (2) The visual-spatial arrangement of information unburdens cognition from decoding abstract, alpha-numerical symbols, and sequences of language-like representations. It enables a more natural interpretation (e.g., of data items' relations via "display proximity" [56]) and facilitates visual search and information integration. (3) Complex analytical operations can be offloaded to the swift workings of visual pattern recognition and pre-attentive processing [16]. (4) Strenuous symbol-based reasoning operations with abstract data can be offloaded to interactions with visual-spatial models of the data. The users then can explore these visual models perceptually and read off conclusions "without presupposing mental logics and formal rules" [27, p. 1000].

3.2.2 Mental Models

Cognitive science has developed a variety of concepts to describe and understand internal representations of external data.[1] One specifically interesting approach comes with the theory of mental models: observations and explorations—in physical [54], as well as in abstract and artificial environments [22]—instruct "the creation and interpretation of an internal mental model" [50, p.921].

While humans explore physical surroundings or information spaces—from caves to cities and from libraries to complex datasets—they continuously build up mental models as analog representations of observed objects, systems, or their environments and use these mental models to reason on them. In this regard, mental models are similar to the concept of cognitive "frames" [24], which also integrate, connect, and organize data from external observations (see Fig. 3.2, left-hand side). The practical relevance of both, frames and mental models, is their flexibility and

[1] Prominent concepts include cognitive schemata, cognitive scripts, cognitive images, cognitive maps, prototypes, or cognitive frames [23, 27].

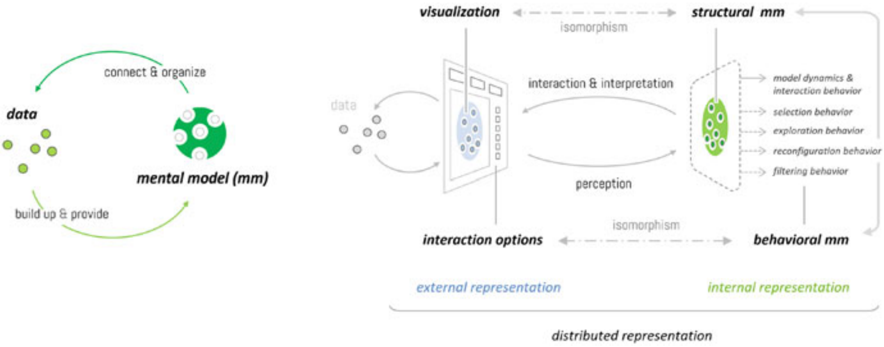

Fig. 3.2 Like cognitive frames, mental models connect and organize data in an interactive loop (left, adapted from [24]). Visualizations provide external templates (center, blue) for the construction and augmentation of internal mental models (green). Visual reasoning thus equals the hybrid interplay of such coupled, but distributed representations (right)

their potential to "describe, explain and predict a system's purpose, form, function, and state" [17, p. 46].

For more complex external objects or systems like machines or interfaces, two different conceptions of mental models were developed, which lay their emphasis either on the structural or on the behavioral and functional aspects of a system [19]: while a structural model focuses on the spatial or topological arrangement of a system and its parts [22], a behavioral model represents its dynamic and causal processes, including the conditional and functional behavior of a system—like how a device works when used in specific ways [33].

With specific regard to visualizations, Liu and Stasko declare that a mental model is a functional analog representation to an external interactive visualization system, which preserves structural and behavioral properties of external systems [27, p. 1001]. On the *structural* side, the mental model internalizes the spatial layout of a visualization, but it also integrates other visual cues (color, hue, size, etc.), overlaid images, and texts [27, p. 1000]. When seeing data of a subject matter encoded into a visualization on a display, the analyst's structural model mirrors the "spatial, temporal or distributional relations between the data items" [27, p. 1000]. On the *behavioral* or *functional* side, mental models include dynamic information about a structure's performance and how it reacts to interactions (see Fig. 3.2, right).[2]

For users of visualizations, the quality of their internal representations and the related distributed reasoning processes depends to a great extent on the design of

[2] As we will argue later on, visualization theory would also benefit from integrating narrative sequences and stories into the second category of behavioral models (Sect. 3.3.2.2), so that the mental model concept can cover representations of static structures *and* time-oriented sequences in an equal fashion—similar to the distinction of cognitive schemata and cognitive scripts [44].

the external representation, as their internal models are modeled on the base of the latter.[3]

However, traditional descriptions of mental models frequently agree on their non-veridic character: they commonly do not mirror external representations in an accurate or detailed manner, but they have to be good enough to ensure (task-specific) functionality or viability. Whether they are serving for descriptive, explanatory or predictive cognitive operations—internal representations mostly do so without even coming close to the complexity and full details of its corresponding external representations or subject matters. They can remain parsimonious, sketchy, and lightweight but can still be functional, viable, or "runnable" for the achievement of certain tasks [33]. This is possible only if a mental model is isomorphic to certain aspects of the external representation, which again has to show isomorphic aspects with regard to the available data about an external phenomenon [20].

As a full correspondence or richness of detail is no important indicator for the quality of mental models, other quality indicators have been discussed. Among them, a model's inherent *coherence* and *consistency* have been emphasized by cognitive science research as indicator for the integration and connection of relevant aspects into the internal representation [40, 46, 54]. In the following, we will elaborate on the role of visualization design for the construction of mental models and for the coherent integration of information.

3.3 Designing Visualizations from a Mental Models Perspective

The outlined theory of distributed cognition and model-based reasoning allows to reflect on the interplay of visualization and cognition in a synoptic fashion and to reframe known challenges for visualization design from a generic model-building perspective. In the following, we will focus on two major visualization challenges and show how they can be theoretically unified and understood as challenges of model development. Firstly, we will look at the challenge to *initially construct* a mental model, when users work with a dataset and/or tool for the first time (Sect. 3.3.1). Secondly, we will reflect on the challenge to integrate information from multiple views into larger mental macro models (Sect. 3.3.2). This can be done by spatial or synchronic coherence techniques (Sect. 3.3.2.1) or by temporal coherence techniques, commonly referred to as "narrative visualization design" (Sect. 3.3.2.2).

[3] Due to the prevalence of user-oriented design, the quality of visualizations as external representations is tied back to the quality of the internal representations that they generate (e.g., the utility, efficiency, correctness, esthetic appeal, etc.). Arguably, it is this circle, which makes it relevant for visualization designers to know about cognitive principles (i.e., from Gestalt and color perception to more complex model construction and reasoning processes) to design for the effective amplification of perceptual and cognitive processes.

3.3.1 Supporting the Initial Construction of Mental Models

Based on their first impression of a visualization, the users form a tentative mental model, which can be further manipulated and elaborated in working memory. How much effort has to be put into this construction, and which structure and function this initial mental model includes, depends on a number of factors—including the user's existing internal representations (i.e., visualization literacy, prior knowledge, domain knowledge), situational factors (i.e., data, tasks, motivation), and the external representation (i.e., the visualization and system design) itself.[4] If these factors do not align with each other, there is a risk that users ignore a visualization and forego the mental efforts to build up an internal representation. To reduce this risk and to facilitate the initial construction of a mental model, users can be actively supported, e.g., by introducing the basic structure of the data and its visual representation (see Sect. 3.3.1.1) or by functional and behavioral onboarding support, e.g., by providing transitions from known concepts and visualizations (see Sect. 3.3.1.2).

3.3.1.1 Structural Construction Support: Advance Organizer

In cognitive science, *advance organizers* have been introduced as effective means to facilitate the construction of mental models [3]. In general, an advance organizer is a structural pre-sketch of the information to be learned which is administered in advance, so as to better organize and integrate subsequent details and information into this structure. Within the framework of mental models, advance organizers serve as external sketches or construction plans for the tentative buildup of internal structures, which then are further elaborated in working memory. Thus, the effort of model construction is significantly reduced by introducing a simplified model of a more complex external representation first. The construction of an internal representation thereby becomes an incremental endeavor.

[4] A large part of the basic research on mental models has been done in the context of text comprehension and with regard to subject matters, where a spatial layout of environmental data is given. In such a context, understanding an external representation (e.g., the description of a built environment) requires the construction of a mental model, for which a visual-spatial isomorphy between relevant aspects of internal and external representations should be achieved—and is relatively easy to verify. Despite the fact that (the rules of construction for) external representations preserving a spatial layout are widely known and universally established (e.g., by naturalistic images, miniature models, or instances of "scientific visualization"), it is known that the initial build-up of an internal model (i.e., internalization) is cognitively and energetically demanding. This holds even more true for the internalization of pictures which spatialize abstract or conceptual data due to the rules of a diagrammatic syntax (often summarized as techniques of "information visualization"). Especially, if the users are not familiar with the rules of construction, they face higher barriers as they have to build up both: a (structurally and behaviorally) isomorphic model from the external representation *and* a basic understanding of the principles or rules of image construction (visualization literacy [6]).

Advance organizers are known to support mastery of content in nonlinear, unstructured environments like hypermedia [31] or multidimensional information environments [57]: they not only provide a conceptual overview and facilitate navigation in more complex information spaces but can also raise curiosity and interest. An advance organizer can be graphical or textual, but graphical representations have been claimed to be less ambiguous and more concise than textual ones [9].

Advance organizers for visualization systems can take different forms: (1) they give a simplified overview of the data in a selected single view (e.g., a simplified visualization, a picture of a visualization anatomy, or a thumbnail preview) or (2) they present structural information on a dataset [46]. Also, the widely known visual information seeking mantra "overview first, zoom and filter, then details-on-demand" [48, p. 337] equals an argument for incremental model construction, so that overviews internalized in advance allow to organize the subsequent intake of more detailed information.

For more complex datasets, which encompass more data dimensions than can be displayed in one single view, a structural data model can facilitate the understanding of the corresponding more complex visualizations, e.g., as given by multi-view systems (see Fig. 3.3, right): after introducing multiple data dimensions, they can support the initial construction of internal representations for individual views and illustrate their integration into compound visualizations (and internally into mental macro models) later on.

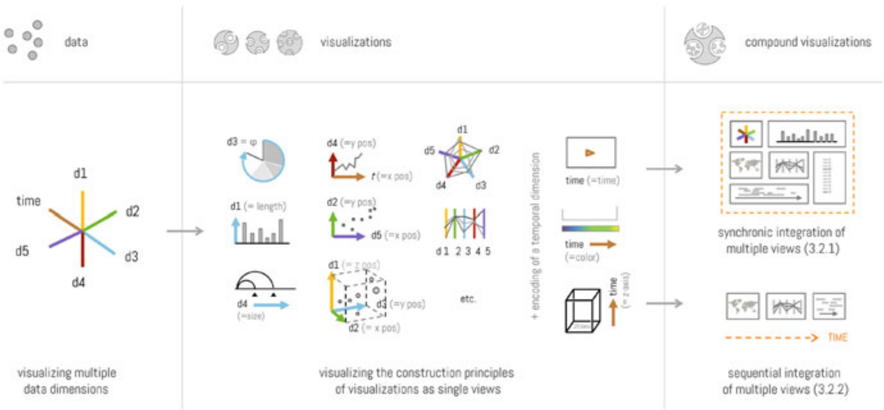

Fig. 3.3 Advance organizer (left) depicting the structure of a multidimensional dataset, whose encoding into specific visualizations within single views (center) and compound visualizations (right) could be visually traced

3.3.1.2 Behavioral Construction Support: Onboarding Techniques

Visualizations are interactive, artificial images, for which users do not only have to learn their visual structures (i.e., their construction principles and their visual-spatial Gestalt) but also their behavior via tool functions—especially for the first time users. From a mental model perspective, this has been addressed as the need to also build up *functional* or *behavioral* mental models (Fig. 3.2, right-hand side). While many visualization tools convey such behavioral knowledge with help functions and text or video tutorials, some of this knowledge can also be introduced by visualization *onboarding techniques*, which help "users in reading, interpreting, and extracting information from visual representations of data" [52, p.2]. Especially novel visualization techniques or complex visualization systems require a certain amount of training and learning.

Onboarding support for web-based visualization tools is often provided with *guided tours*, *step-by-step wizards*, or initial *overlays*, which highlight the most important parts of the graphical user interface and interaction options for the first time visitors [52]. As such, the users learn what they can do with the visual representations on screens and how they behave due to user interaction.

Another form of functional onboarding support is to provide *training via examples* [15]: the users are walked through visualizations by means of an exemplary dataset. During this walk-through, they construct a prototypical functional mental model incrementally and can elaborate it later on for their own datasets.

An interesting onboarding technique for the structural aspects of visualizations (especially with regard to the origins of mental model research in the field of spatial cognition [54]) is the use of *seamless transitions*, which allow to trace the re-arrangement of familiar spatial constellations into abstract (information) visualization layouts. These techniques help to transfer structural knowledge (and context) from existing mental models into novel diagrammatic constellations [42] (see also Sect. 3.3.2.1).

3.3.2 Supporting the Integration of Information from Multiple Views

Complex visualization systems frequently operate with multiple views [4]. Such systems require sensemaking and model building on multiple levels of information integration: their users do not only have to build up internal representations for the single views but also an integrated compound representation, which we refer to as mental *macro model* (Fig. 3.4, top right).

Information integration on a compound or macro level requires additional cognitive effort. In the following, we will discuss two approaches to support the construction of internal macro models: (A) either multiple distinct mental models of individual views are connected in a parallel or *synchronous* fashion (see Sect. 3.3.2.1

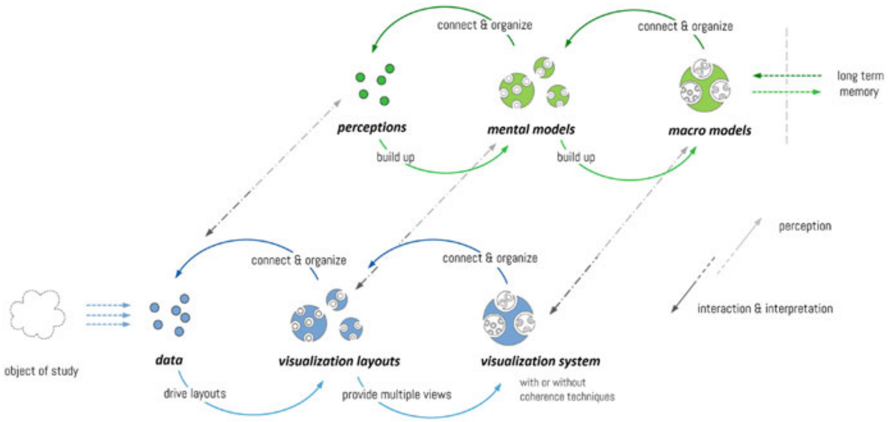

Fig. 3.4 Depiction of a complex visualization system from a distributed cognition perspective, with external representation components in blue and internal representation components in green. Both layers contain basic model-building cycles on the left-hand side, which are coupled by perception and interaction dynamics (gray). These basic cycles are then extended by macro modeling cycles (right-hand side)

or (B) the mental model of an individual view is *sequentially* connected to novel views (commonly referred to as visual narration or storytelling, see Sect. 3.3.2.2).

3.3.2.1 Synchronous Integration: Coordination and Linkage of Views

When confronted with multiple views in parallel, users have to mentally connect and integrate information from all views to assemble a bigger picture and to achieve a more comprehensive understanding. Without further support, users will build up unconnected mental models first, which have to be interconnected with significant cognitive effort later on to allow reasoning on their integrated data [22]. In this context, the design of *coordinated* visualizations supports coherent information integration and macro modeling from the beginning.

By offering multiple views in parallel, their diverse encodings are brought into a spatially *adjacent* compound constellation, which offers complementary analytical perspectives for synchronous contemplation [41]. To better connect information from these views, the visual encoding of data should be handled *consistently* across different views (e.g., consistent use of colors, labels, directions of axes, or other design decisions) [40].

Further integration support is usually provided by *coordinated interaction techniques*, which enrich a visualization system with further coherence cues to establish synchronous connections between juxtaposed views. Among the most common techniques are *coordinated selection and highlighting* or *linking and brushing* or also synchronized panning, scrolling, or zooming [35]. Coordinated interaction methods provide instant visual modifications of the same data elements in different views, which enable the synchronous perceptual integration of parallel views.

An interesting option to literally interconnect multiple views has been proposed by Collins and Carpendale [10]: as a method to explore relations between two different views, VisLink connects the same data items on different canvases with explicit links and thus establishes synchronous perceptual bridges. While this method could also be extended beyond two views, it is expected to become visually complex soon.

3.3.2.2 Sequential Integration: Narration, Storytelling, and Seamless Transitions

A second major principle for connecting multiple views is given by *sequential* coherence techniques which interconnect individual views sequentially, i.e., over time. In this case, it is not spatial, but temporal adjacency—commonly together with a range of other narrative cues and connections—which provide the binding relations. Practically, as a pure sequence, sequential integration of visualizations utilizes an observer's memory to hold transient perceptions of single views present in the working memory and to store them as a compound sequence. As such, pure sequential integration also requires significant mental effort. Therefore, various other time-oriented connection techniques have been suggested to facilitate sequential information integration.

In a visualization context, *narration and storytelling* became a widely used and much-debated design approach over the last decade. Kosara and Mackinlay [25] proposed to use storytelling as a more effective way of communicating and presenting data: "Stories have proven to be not only an incredibly popular way of conserving information and passing it on, they also provide the connective tissue between facts to make them memorable" [25, p. 2]. They define a story as a *causally related* chain of individual visualizations (see Fig. 3.5, right). This ordered sequence often, but not necessarily, corresponds to a chronological course of events and provides the user with a clearly defined path through the data.

From a mental model perspective, the relevance of the growing work on visualization storytelling seems obvious: as a major complement to space-leveraging, synchronous integration techniques, sequential storytelling utilizes time as an "orthogonal" dimension of information integration (see Fig. 3.5, right) and thus provides the second major coherence technique for the construction of mental macro models. Specific examples for such a narrative integration are instantiated by *slideshows* (i.e., multiple visualizations further interwoven by spoken or written language), *magazine* or *scrollytelling* designs (e.g., visualizations interwoven by text), or *animation* and *film-based* approaches [47].

Within such sequential arrangements, *animated transitions* can further strengthen the coherence between different views by offering fluid and traceable incremental changes. Various forms of morphing can support the translation from one spatialization method to another and thus provide more elastic macro designs [5]. By

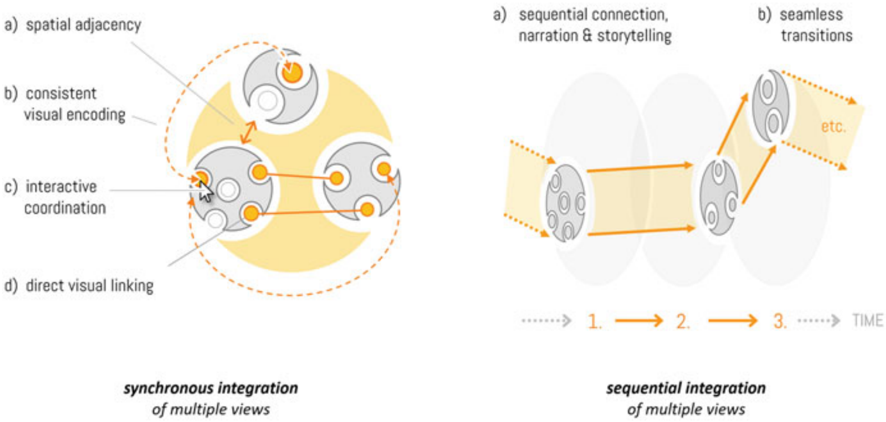

Fig. 3.5 Macro model construction based on visual coherence techniques for the synchronous interconnection of multiple views (left, Sect. 3.2.1) and their sequential interconnection, commonly referred to as narrative visualization or storytelling (right, Sect. 3.2.2)

changing layouts incrementally—as opposed to cutting abruptly—the spatial re-arrangement can be traced and the shifting of relevant well-known elements can be followed smoothly and fluidly [13]. These techniques correspond to the concept of preservation of the mental map [1], which aims at developing algorithms that keep the number of changing elements to a suitable minimum.

Results from cognitive science [22] and visualization [42] provide evidence that it is cognitively more efficient to elaborate an existing mental model and transform it into a new arrangement than to combine two different mental models, which puts sequential integration in favor of a synchronous integration. The unmediated, synchronous presentation of two separate layouts can easily lead to two separate models, whereas their sequential interconnection via storytelling or seamless transitions allow the adaptation and extension of an existing mental model, resulting more likely in one coherently integrated mental model with less cognitive effort [22]. Results of studies with users prove the value of this technique: “transitions could save hours to be spent with reading a manual otherwise” [49, p. 637]. Taking these considerations into account in a recent study, we implemented a seamless transition from a familiar 2D map view to a 3D space-time cube to help laypersons, who are unfamiliar with this visualization technique, to understand the functionality of this visualization. The transition supported users to connect the new visualization with their mental model of a map and was evaluated as very helpful [58]. These empirical results indicate that the direction of a transition is very important: designers should carefully consider which view is presented first—what Tominski et al. refer to as “prioritized views” [53]—and knowledge of the users’ mental models can support this decision.

Obviously, synchronous and sequential information integration can also play together in various ways, which has been discussed as combinations of author-driven and reader-driven approaches [47]. While exploration-oriented, synchronous compound visualizations do not prescribe any specific order or linearity, narrative, or author-driven visualizations guide through their materials in a predefined way. Advanced guidance systems thus often search for a balance of both approaches. Like an advanced organizer (see Sect. 3.3.1.1), they introduce the readers to the story first and thereby support their construction of a first mental model. Later, users can freely explore the visualization but can better integrate further information into their initial mental model.

3.4 Discussion

Current visualization research builds on different cognitive frameworks, producing a fair amount of terminological and theoretical diversity—which becomes also visible throughout discussions about visualization foundations [39]. Different theories in visualization focus on different aspects of the analysis process: whereas sensemaking theories [38] provide a broad framework for the description of the analysis process, insight theory [34] focuses on the outcomes of the analysis process. In contrast, the concept of mental models [27] fills a gap in visualization research: how do users represent a visualization (system) internally to use it for distributed sensemaking processes and to generate single or interconnected insights? A better understanding of internal representations can help us to better understand the contributions of visualization *and* cognition to a mutually connected (i.e., "distributed") modeling endeavor. Such a unified, model-based approach can also help to understand some challenges of visualization (systems) better—and to direct attention to areas where users frequently need additional model-building support. In this chapter, we drew together existing techniques to support model construction (i.e., for the initial construction of single views and the further connection of multiple views) and thus aimed to illustrate how otherwise separated visualization topics and debates could be organized and mediated in a more unified fashion.

The concept of mental models has received a considerable amount of research in cognitive science—but also beyond. When we transfer these findings to the field of visualization, some new questions arise, which remain to be solved in the future.

3.4.1 Macro Models

We consider visualization to be a key competence and practice to provide "bigger pictures" of complex topics—from society, technology, or ecology—to the society. Such subject matters are commonly represented by complex text collections only and remain invisible to the unaided eye. Such bigger pictures—especially in the form of well-designed visualization systems—can be worth a whole text collection. But how large can the resulting mental models actually get and how much information can be memorized? How much visualized information can users hold in their working memory (and later recall from their long-term memory) when a visualization system gets more complex? This size will very likely depend not only on the users' motivation and domain knowledge but also on existing internal representations they can build on. But the design of the visualization (system) and the coherence techniques discussed in this chapter also play a decisive role. In this context, we want to argue for the collective development of a whole catalog of "templates" for mental macro models and corresponding visualization techniques. It stands to reason that such an endeavor should combine a more systematic collection of compound visualization designs (see Sect. 3.3.2) with a collection of visual (macro) metaphors [14], which could help to organize the development of shared mental macro models in education, organizations, or journalism.

3.4.2 Model Quality, Stability, and Depth of Internalization

A certain challenge for the use of the mental model concept comes from its common connotation to describe a relatively autonomous and stable entity. However, it is an open question to which degree of detail, coherence, autonomy, and longevity internal representations of visualizations actually rise—especially in cases where the first time users meet complex visualization systems. Even if mental models are said to form as non-veridic, sketchy, and lightweight frames of a visualization (Sect. 2.2), we can expect significant differences in terms of internal modeling diligence and coherence for different users in different contexts. In this context, Tversky [54] suggested the term of *cognitive collages* for internal representations which do not cross a certain threshold of consistency and coherence, but rather appear as a distorted mix-up of partial information. In terms of modeling depth, it is an open question if (aside from a visualization's structure and behavior) mental models also include rich information about the underlying data, since we can offload these to the visualization, or whether we internally represent also data in our working memory [27].

3.4.3 Advancement of Story Models

Another conceptual challenge comes from the need to evenly cover internal representations of static structures and dynamic processes with the mental model concept. While the differentiation of structural and behavioral mental models already provides a useful distinction, we think that the behavioral model concept should be further elaborated and enriched to also cover all kinds of time-oriented phenomena, including transient sequences, processes, dynamics, or complex stories. While stories in visualization have been robustly defined as *causally related chains of individual visualizations* [25], we consider a rich body of work on narrative mental models ("story world", cf. [55, 60]) in the area of text comprehension and narrative research to provide material for the future refinement of narrative visualization techniques [29]: Recipients usually have a whole repertoire of schemata on how a story is built, how it progresses, and what its constituents are [51]. These schemata reduce cognitive load and allow stories to be processed fast and efficiently—not only in texts but also in narrative visualizations. Narrative cues and coherence indices (information on time, place, protagonist, cause, and goal [59]) direct their attention and help recipients to build up and update a narrative mental model. On a more abstract level, stories guide recipients through a bigger picture by providing various sequential links between events or places, or, generally speaking, between different local data elements.

3.4.4 Modality

Are mental models predominantly visual, verbal, or multimodal structures? Working memory research distinguishes a visual component (visuo-spatial sketchpad) from a verbal component (phonological loop) for the separated processing of incoming information. Mental models have been associated to both modalities but are more often allocated on the visual side, where they can also flexibly integrate multimodal information (such as tags, comments, or (narrative) context information in the case of visualization). Related theories (like the dual coding theory [8] or integrated models of text and picture comprehension [45]) often suggest a two-layered architecture with numerous transmodal connections for visual–verbal and verbal–visual translation. This question is especially relevant for the assessment of mental models (see [28] for a summary of evaluation techniques from cognitive science): many evaluations build on some kind of verbal reporting (for instance from think aloud protocols or interviews), but if a mental model is mainly visual, are such methods valid? Or do they raise additional cognitive load to translate visual information to verbal one?

3.4.5 Sharing Mental Models

Visualization design—and the evaluation of internal representations—is frequently oriented toward the amplification of individuals' performance. By contrast, we see a need to go beyond the individual and strengthen research into visualizations as means for the development of shared mental models—and the need for evaluation methods able to measure these collaborative efforts. Not only to arrive at more agreed-upon models (e.g., in collaborative constellations and teams) but also to become aware of meaningful modeling differences, where the understanding of complementary roles, perspectives, and positions matter for the understanding of a complex subject matter. To that end, we also consider "discursive" visualization practices a desideratum as a means to collectively construct perspective-rich external representations. Such discursive visualizations would make traces and layers of modeling controversies visible—and thus foster the understanding how and to which degree visual and mental models of public, complex subject matters actually converge.

The theory of mental models provides a well-sourced cognitive framework for visualization design with the potential to translate into a unified and instructive model-building framework from a distributed cognition perspective. While this work has been started a while ago [26, 27], it still contains a range of open questions. This chapter introduced the theoretical background, implications for visualization design, and suggestions for their future development. As such, this framework might further unfold its potential to instruct visualization research and teaching—and to draw attention to general questions of (macro) model development support for visualization novices and non-experts—which tend to go unnoticed in an expert and performance-oriented research field.

Acknowledgments This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101004825 and from the Austrian Science Fund (FWF), Project No. P28363-G24.

References

1. D. Archambault and H. C. Purchase. Mental map preservation helps user orientation in dynamic graphs. In *International Symposium on Graph Drawing*, pages 475–486. Springer, 2012.
2. R. Arias-Hernandez, T. M. Green, and B. Fisher. From cognitive amplifiers to cognitive prostheses: Understandings of the material basis of cognition in visual analytics. *Interdisciplinary science reviews*, 37(1):4–18, 2012.
3. D. P. Ausubel. The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of educational psychology*, 51(5):267–272, 1960.
4. M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *AVI '00: Proceedings of the working conference on Advanced visual interfaces*, pages 110–119. ACM Press, 2000.
5. M.-J. Bludau, V. Brügemann, A. Busch, and M. Dörk. Reading traces: Scalable exploration in elastic visualizations of cultural heritage data. *Computer Graphics Forum*, 39(3):77–87, 2020.
6. K. Börner, A. Bueckle, and M. Ginda. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, 116(6):1857–1864, 2019. Publisher: National Acad Sciences.
7. S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers Inc., 1999.
8. J. M. Clark and A. Paivio. Dual coding theory and education. *Educational Psychology Review*, 3(3):149–210, Sept. 1991.
9. J. W. Coffey, R. Hoffman, and A. Cañas. Concept map-based knowledge modeling: perspectives from information and knowledge visualization. *Information Visualization*, 5(3):192–201, 2006.
10. C. Collins and S. Carpendale. Vislink: Revealing relationships amongst visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1192–1199, 2007.

11. K. J. W. Craik. *The nature of explanation*. Cambridge University Press, 1943.
12. A. W. Crapo, L. B. Waisel, W. A. Wallace, and T. R. Willemain. Visualization and the process of modeling: a cognitive-theoretic view. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 218–226, 2000.
13. N. Elmqvist, A. Vande Moere, H.-C. Jetter, D. Cernea, H. Reiterer, and T. J. Jankun-Kelly. Fluid interaction for information visualization. *Information Visualization*, 10(4):327–340, Oct. 2011.
14. M. J. Eppler. The image of insight: The use of visual metaphors in the communication of knowledge. In *Proceedings of I-KNOW*, volume 3, pages 2–4, 2003.
15. J. Gurlitt, S. Dummel, S. Schuster, and M. Nückles. Differently structured advance organizers lead to different initial schemata and learning outcomes. *Instructional Science*, 40(2):351–369, 2012.
16. C. G. Healey, K. S. Booth, and J. T. Enns. High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(2):107–135, 1996.
17. M. Hegarty. Diagrams in the mind and in the world: Relations between internal and external visualizations. In *International Conference on Theory and Application of Diagrams*, pages 1–13. Springer, 2004.
18. M. Hegarty. The cognitive science of visual-spatial displays: Implications for design. *Topics in cognitive science*, 3(3):446–474, 2011.
19. J. Heiser and B. Tversky. Arrows in comprehending and producing mechanical diagrams. *Cognitive science*, 30(3):581–592, 2006.
20. C. Held, M. Gottfried, G. Vosgerau, and M. Knauff. *Mental models and the mind: current developments in Cognitive Psychology, Neuroscience and Philosophy of Mind*. Elsevier, 2006.
21. E. Hutchins. *Cognition in the Wild*. MIT press, 1995.
22. P. N. Johnson-Laird. Mental models in cognitive science. *Cognitive science*, 4(1):71–115, 1980.
23. G. Klein and R. R. Hoffman. Macrocognition, mental models, and cognitive task analysis methodology. *Naturalistic decision making and macrocognition*, pages 57–80, 2008.
24. G. Klein, J. Phillips, E. Rall, and D. Peluso. A Data-frame theory of sensemaking. In R. R. Hoffman, editor, *Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*, pages 113–155, New York, NY, 2007. Lawrence Erlbaum Assoc Inc.
25. R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013.
26. Z. Liu, N. Nersessian, and J. Stasko. Distributed cognition as a theoretical framework for information visualization. *IEEE transactions on visualization and computer graphics*, 14(6):1173–1180, 2008.
27. Z. Liu and J. Stasko. Mental models, visual reasoning and interaction in information visualization: A top-down perspective. *IEEE transactions on visualization and computer graphics*, 16(6):999–1008, 2010.
28. E. Mayr, G. Schreder, M. Smuc, and F. Windhager. Looking at the representations in our mind: Measuring mental models of information visualizations. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, BELIV '16, page 96–103, New York, NY, USA, 2016. Association for Computing Machinery.
29. E. Mayr and F. Windhager. Once upon a spacetime: Visual storytelling in cognitive and geotemporal information spaces. *ISPRS International Journal of Geo-Information*, 7(3), 2018.
30. W. McCarty. Modeling: A Study in Words and Meanings. In S. Schreibman, R. Siemens, and J. Unsworth, editors, *A Companion to Digital Humanities*, page (chapter 19). Blackwell, Oxford, online edition, 2004.
31. T. F. Mcmanus. Individualizing instruction in a web-based hypermedia learning environment: Nonlinearity, advance organizers, and self-regulated learners. *Journal of Interactive Learning Research*, 11(2):219–251, 2000.

32. T. Munzner. *Visualization Analysis and Design*. A K Peters/CRC Press, Boca Raton, 1 edition, Dec. 2014.
33. D. A. Norman. Some observations on mental models. *Mental models*, 7(112):7–14, 1983.
34. C. North. Toward measuring visualization insight. *IEEE computer graphics and applications*, 26(3):6–9, 2006.
35. C. North and B. Shneiderman. Snap-together visualization: can users construct and operate coordinated visualizations? *International Journal of Human-Computer Studies*, 53(5):715–739, 2000.
36. L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. Decision making with visualizations: A cognitive framework across disciplines. *Cognitive research: principles and implications*, 3(1):1–25, 2018.
37. R. E. Patterson, L. M. Blaha, G. G. Grinstein, K. K. Liggett, D. E. Kaveney, K. C. Sheldon, P. R. Havig, and J. A. Moore. A human cognition framework for information visualization. *Computers & Graphics*, 42:42–58, 2014.
38. P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, volume 5, pages 2–4. McLean, VA, USA, 2005.
39. M. Pohl, M. Smuc, and E. Mayr. The user puzzle—explaining the interaction with visual analytics systems. *IEEE transactions on visualization and computer graphics*, 18(12):2908–2916, 2012.
40. Z. Qu and J. Hullman. Keeping Multiple Views Consistent: Constraints, Validations, and Exceptions in Visualization Authoring. *IEEE transactions on visualization and computer graphics*, Aug. 2017.
41. J. C. Roberts. State of the art: Coordinated & multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on*, pages 61–71. IEEE, 2007.
42. P. Ruchikachorn and K. Mueller. Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE transactions on visualization and computer graphics*, 21(9):1028–1044, 2015.
43. M. Scaife and Y. Rogers. External cognition: how do graphical representations work? *International journal of human-computer studies*, 45(2):185–213, 1996.
44. R. C. Schank and R. P. Abelson. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press, 1977.
45. W. Schnotz and C. Kürschner. External and internal representations in the acquisition and use of knowledge: visualization effects on mental model construction. *Instructional Science*, 36(3):175–190, May 2008.
46. G. Schreder, F. Windhager, M. Smuc, and E. Mayr. A mental models perspective on designing information visualizations for political communication. *JeDEM-eJournal of eDemocracy and Open Government*, 8(3):80–99, 2016.
47. E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.
48. B. Shneiderman. The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343, 1996.
49. M. Smuc, P. Federico, F. Windhager, W. Aigner, L. Zenk, and S. Miksch. How do you connect moving dots? insights from user studies on dynamic network visualizations. In *Handbook of human centric visualization*, pages 623–650. Springer, 2014.
50. R. Spence. A framework for navigation. *International Journal of Human-Computer Studies*, 51(5):919–945, 1999.
51. N. L. Stein and V. I. Kissel. Story schemata and causal structure. In *Routledge Encyclopedia of Narrative Theory*, pages 567–568. Taylor & Francis Group, 2010.
52. C. Stoiber, F. Grassinger, M. Pohl, H. Stitz, M. Streit, and W. Aigner. Visualization onboarding: Learning how to read and use visualizations. In *Proceedings of VisComm 2019*. OSF Preprints, 2019.

53. C. Tominski, G. Andrienko, N. Andrienko, S. Bleisch, S. I. Fabrikant, E. Mayr, S. Miksch, M. Pohl, and A. Skupin. Toward flexible visual analytics augmented through smooth display transitions. *Visual Informatics*, 2021.
54. B. Tversky. Cognitive maps, cognitive collages, and spatial mental models. In *European conference on spatial information theory*, pages 14–24. Springer, 1993.
55. T. A. van Dijk and W. Kintsch. *Strategies of Discourse Comprehension*. Academic Press, 1983.
56. C. D. Wickens and C. M. Carswell. The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human factors*, 37(3):473–494, 1995.
57. F. Windhager and E. Mayr. Cultural heritage cube. A conceptual framework for visual exhibition exploration. In *2012 16th International Conference on Information Visualisation*, pages 540–545. IEEE, 2012.
58. F. Windhager, S. Salisu, R. A. Leite, V. Filipov, S. Miksch, G. Schreder, and E. Mayr. Many views are not enough: Designing for synoptic insights in cultural collections. *IEEE Computer Graphics and Applications*, 40(3):58–71, 2020.
59. R. A. Zwaan. Five dimensions of narrative comprehension: The event-indexing model. In S. R. Goldman, A. C. Graesser, and P. van den Broek, editors, *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso*, pages 93–110. Lawrence Erlbaum Associates, 1999.
60. R. A. Zwaan and G. A. Radvansky. Situation models in language comprehension and memory. *Psychological bulletin*, 123(2):162, 1998.

Chapter 4
Improving Evaluation Using Visualization Decision-Making Models: A Practical Guide

Melanie Bancilhon, Lace Padilla, and Alvitta Ottley

Abstract In visualization research, evaluation is a crucial step to assess the impact of visualization on decision-making. Existing work often gauges how good a visualization is by measuring its ability to induce accurate and fast judgment. While those measures provide some insight into the efficacy of a graph, underlying cognitive processes responsible for reasoning and judgment are often overlooked when they can have significant implications for visualization recommendation. Cognitive processes do not need to be a black box. There exists multiple models that describe decision processes, such as theories from behavioral economics and cognitive science. In this chapter, we compare and contrast different models and advocate for the inclusion of cognitive models for visualization evaluation in the context of decision-making. The goal of this work is to show visualization researchers the advantages of adopting a more mechanistic approach to evaluation at the intersection of visualization and cognitive science.

4.1 Introduction

We make decisions based on data every day, ranging from trivial to complex. Such choices could include when to leave the house to catch the bus, take an umbrella given the chance of rain, or invest in the stock market given the historical trends. In many instances, charts and graphs have become an integral part of our decision-making process. Visualization research has provided valuable insight into perceptual science and has led to the amelioration of chart design and visualization recommendations. Charts frequently appear in information communication, data analysis, sensitization campaigns, and even medical diagnostics and can significantly impact

M. Bancilhon (✉) · A. Ottley
Washington University in St. Louis, St. Louis, MO, USA
e-mail: mbancilhon@wustl.edu; alvitta@wustl.edu

L. Padilla
Northeastern University, Boston, USA
e-mail: lace.padilla@ucmerced.edu

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_4

people's lives. But all charts are not equal. When a new graph or chart is designed, it is essential to conduct an evaluation under realistic decision-making conditions to understand and foresee its effect on real-life decisions.

However, it can be hard to know if an evaluation is close enough to natural decision-making conditions to provide meaningful insights into the efficacy of a visualization. One way to conduct rigorously valid evaluations is to understand and simulate the underlying mental mechanisms at work when a viewer completes the real-world task. Fortunately, cognitive scientists have extensively studied cognitive mechanisms responsible for interpreting and misinterpreting visual designs under different modes of reasoning. For example, *dual-process theory* posits that there exists two types of decisions operating under distinct cognitive processes: intuitive (Type 1) and strategic (Type 2) decisions, which require significantly more effort than Type 1 [33]. In this chapter, we dive into multiple prominent perspectives of decision-making. We discuss how the researchers can apply frameworks and models pertaining to visualization design and evaluation in the context of decision-making. We propose that dual-process cognitive models are some of the most useful and easily applied for visualization research. This chapter will be helpful for designers and visualization researchers looking to adopt a more granular approach to decision-making and conduct holistic evaluations for better visualization recommendations.

4.1.1 Evaluation Methods for Decision-Making

Research on visualization evaluation is vast and varied [17, 43], with high tendencies toward evaluating visualization based on speed and accuracy in perceptual judgments [64]. A relatively small number of studies have focused on evaluating people's visualization-aided decisions. Researchers have investigated how visualizations impact attitudes toward risk and hypothetical decisions [22, 62]. For example, Ruiz et al. [62] conducted a study where they asked at-risk patients to decide whether they would opt for screening based on hypothetical risk information about a disease [62]. They found that people are more risk-averse when presented with icon arrays. Kay et al. [37] evaluated how well different visualizations communicate the uncertainty of transit data by asking participants to estimate the likeliness of bus arrival times on a scale of 0 to 100 [37].

In traditional visualization empirical studies, visualizations are often evaluated by their ability to prompt accurate and fast responses in behavioral tasks, that may or may not involve making a decision. While it is common to extrapolate the appropriateness of visualizations for decision-making through these performance-based measures, there are less attempts to evaluate visualization designs based on the quality of the decisions they elicit [51]. Empirical evaluations of visualization are generally challenging [9, 17, 56]. Thus, one possible reason for the lack of evaluations with decision-making is that it is generally more straightforward to gauge effectiveness via the speed and accuracy of perceptual judgments. Consider, for example, the chart shown in Fig. 4.1, which shows a given person's chance

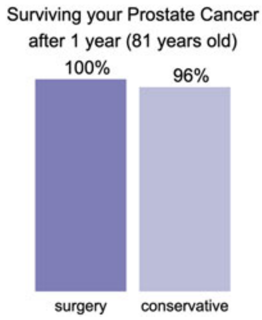

Fig. 4.1 A bar chart comparing the survival rates after one year of surgery versus conservative management for a 80-year-old prostate cancer patient [23]

of surviving prostate cancer after one year if they choose to have surgery (e.g., radical prostatectomy) compared to conservative treatments (e.g., watchful waiting). One could evaluate this chart based on how well it facilitates fast and accurate comparisons of the two quantities, or based on the responses from semi-structured interviews with prostate cancer patients [23]. Experiment protocols like these are more straightforward than those that measure decisions because it is feasible to define a ground truth or expected behavior for the analysis of study findings.

In practice, we often use performance-based findings to inform the selection of visualization designs, implying that accurate decoding likely leads to better and more informed decisions. Based on our current understanding of perceptual judgments, the bar chart in Fig. 4.1 uses position for data encoding, and therefore is ideal for comparing quantities and seeing small differences [12, 13]. However, one could reasonably assert that the difference between the survival rates for surgery (100%) and conservative treatment (96%) is statistically insignificant, but the bar chart might inadvertently emphasize a potentially minor disparity. Existing studies show that the ideal visualization depends on the task. For example, the superior representation for magnitude estimation might not be optimal for part-to-whole judgments [20, 65, 66]. Some researchers have used simulations to observe the direct impact of visualization design on decisions. In one study by Bancilhon et al. [4], participants played a lottery game and chose to either enter the lottery or receive guaranteed monetary gains based on five standard visualization designs. They analyzed the quality of the decisions based on economic optimality and found that people made significantly more risk-seeking decisions with circle and triangle charts [4] (see Sect. 4.3.1.2).

Decision-making is complex and multifactorial. In addition to the graph's appropriateness, a patient's decision to have surgery (or not) will depend on various factors including illness severity age, commodities, and personal finances. People are also prone to various cognitive biases [16], and individual differences in personality and cognitive abilities may also influence usability and choice [40, 53]. At a fundamental level, the decision-maker's perspective drives the decision, and the typical approach of defining a ground truth in an evaluation is non-trivial. Despite this challenge, other fields have demonstrated success in modeling and predicting, and reasoning about how people make decisions [33, 35, 55, 57]. We argue that for visualization to be a practical tool for supporting decision-making, we need to understand the underlying cognitive processes behind decision-making and adopt a unifying cross-discipline framework to evaluate visualization in this context.

To aid this discussion, we adapt Balleine's definition of decision [3]:

> A *decision* is a choice between competing courses of actions [3].

4.2 The Science of Making Decisions

Decisions are governed by complex systems of reasoning that scholars have studied for decades. Researchers in the visualization community have pursued two dominant approaches to study decision-making under risk. The first provides a detailed and quantifiable view of decision-making. It assumes that humans make decisions rationally by weighing the risk and expected outcome of different prospects, two factors that can be measured and modeled. The second posits that many factors can influence decision-making. It proposes that humans make both intuitive (Type 1) and strategic (Type 2) decisions and that decision-makers usually default to using intuition. These two distinct types of decisions operate under a *dual-process system*. To improve visualization research in the context of decision-making, it is crucial to understand the meaning and implications of decision-making under both umbrellas. We structure this chapter around two prevalent approaches: *The Utility-Optimal Perspective* and *The Dual-Process Perspective*.

4.3 The Utility-Optimal Perspective

Behavioral economists have long studied how people make choices under risk by investigating prospects or gambling scenarios. A prospect is a contract:

$$[(x_1, p_1), (x_2, p_2), \ldots, (x_n, p_n)], \tag{4.1}$$

which yields x_i with probability p_i, where $\sum_{i=1}^{n} p_i = 1$ [35]. Prospects provide a simple model for understanding risky decisions. The classical method for evaluating a gamble is through assessing its expected value. The expected value of a prospect is the sum of the outcomes where the probabilities weigh each value:

$$ev = \sum_{i=1}^{n} p_i x_i. \tag{4.2}$$

Consider the gambling scenario from Kahneman and Tversky's book [35]:

> Which do you prefer?
> **Option A:** 50% chance to win \$1000, 50% chance to win \$0
> **Option B:** \$450 for sure

The expected value of option A is 500 (.5 × 1000 + .5 × 0) and the expected value of option B is 450 (1 × 450). A *rational* decision-maker would then choose option A over option B. However, most people would choose the sure payment of \$450. This example highlights the perhaps obvious conjecture that humans are not always rational [35].

Expected Utility Theory (EUT) is one of the foundational theories of decision-making and has served for many years as both a model describing economic behavior [21] and a rational choice model [38]. In particular, it states that people make choices based on their *utility*—the psychological values of the outcomes. For instance, if a person prefers an apple over a banana, it stands to reason that they would prefer a 5% chance of winning an apple over a 5% chance of winning a banana. Using EUT, we can assess the overall utility of a gamble:

$$EU = \sum_{i=1}^{n} p_i u(x_i), \tag{4.3}$$

where the function u assigns utility to an outcome. We sum the utilities u of the outcomes x_i weighted by their probabilities p_i. This model has its limitations. It also assumes that humans are consistent and primarily decide on prospects based on their utility [35, 69]. Nevertheless, EUT provides a standardized tool for researchers to evaluate peoples' behavior when choosing among risky options and is the foundation for the other dominant theory in behavioral economics, *Prospect theory* [35].

Unlike EUT, prospect theory embraces the human factors present in decision-making. Kahneman and Tversky [35] are the pioneer contributors to this knowledge on bias in decision-making. For example, in their early work, they found that 72 out of 100 experiment participants favored the option of getting \$5000 with a probability of 0.001 (e.g., a small probability event) over the prospect of getting \$5 for sure [35]. Both options have the same expected value, yet most participants favored the probability associated with getting \$5000. In its simplest form, we can represent the equation for prospect theory as

$$V = \sum_{i=1}^{n} \pi(p_i) \upsilon(x_i), \tag{4.4}$$

where the function υ assigns value to an outcome and the function π is a probability weighing function that encodes the idea that people are likely to overreact to small probabilities and underreact to large probability events. In summary, prospect theory stipulates that (1) people tend to favor the option of getting a large gain with a small probability over getting a small gain with certainty and (2) people tend to prefer a small loss with certainty over a large loss with tiny probability.

4.3.1 Using Utility-Optimality to Evaluate visualizations

Visualization researchers have leveraged utility-optimal theories to investigate how visualization impacts decisions under risk. By approaching decision-making from this angle, they create an environment where choices have weights, and their evaluation considers the utility-optimal option. We highlight two empirical studies from the visualization community and examine their experimental design, methodology, and research questions. We will begin with a recent publication investigating the impact of uncertainty visualization design by simulating a fantasy football scenario.

4.3.1.1 A Fantasy Football Study

Kale et al. [36] leveraged utility-optimal theories to observe effect size judgments and decision-making with the four uncertainty visualizations. They used a fantasy football game to elicit decisions under risk. Participants were shown the number of points scored by a certain team with and without the addition of a new player. First, they asked participants to estimate a measure of effect size by asking the following question: *"How many times out of 100 do you estimate that your team would score more points with the new player than without the new player?"*. They also asked participants to make binary decisions indicating whether they would *Pay for the new player* or *Keep their team without the new player*. On each trial, the participant's goal was to win an award worth $3.17M, and they could pay $1M to add a player to their team if they thought the new player improved their chances of winning enough to be worth the cost.

They tested four uncertainty visualizations: 95% containment intervals, hypothetical outcome plots (HOPs), density plots, and quantile dot plots, each with and without means added. They found that while adding means to quantile dot plots produced significantly more utility-optimal decisions at low variance, it had no reliable effect on bias in magnitude estimation. Similarly, adding means to HOPs caused significantly more bias in magnitude estimation across both low and high variance but had no reliable effect on decisions. By evaluating uncertainty visualizations using utility-optimality, Kale et al. [36] observed a decoupling of performance across tasks, where the visualization designs that support the least biased effect size estimation do not support the best decision-making and vice versa. The authors attribute this inconsistency to the reliance on different heuristics across the two different tasks, consistent with Kahneman and Tversky's theory [35]. This finding highlights the value of leveraging utility-optimal theories when studying visualization for decision-making.

4.3.1.2 A Classic Lottery Game

Many studies that leveraged utility-optimal decision-making theories employed tasks with hypothetical gains and losses (e.g., [10, 31, 36, 49]). However, it is unclear if people make the same risk judgments when gains and losses do not tangibly affect them. To evaluate visualization decision-making with greater *ecological validity* (i.e., more closely matching real-world conditions), Bancilhon et al. [4] created a gambling game that immersed participants in an environment where their actions impacted the bonus payments they received. The experiment investigated the effect of five charts on decision-making. Replicating the experiment design of prior work in the economic decision-making domain [8], the researchers presented participants with two-outcome lotteries: take the sure gain or gamble at a risk. The experiment employed a point system for payoff quantities where 1 point equaled \$0.01. The probabilities, p_i, were drawn from the set $P = \{.05, .1, .25, .5, .75, .9, .95\}$ and the outcomes x_1 and x_2 ranged from 0 to 150 points (\$0 to \$1.50).

Figure 4.2a shows an example of the lottery sheet used in the study. At the end of the experiment, the game randomly selected one row from each of the 25 lottery sheets that they saw, and the participant's choice in that row determined their bonus. If the participant chose the sure payout in the selected row, their bonus increased by that amount. If they opted to enter the lottery, the game simulated the lottery to determine the payment, with the potential gains and the probabilities as parameters.

Overall, the findings from the study [4] validate that we can use utility-optimal theories to evaluate visualization designs, and that the latter can influence gambling behavior. They had three major findings. First, the *icon array* was most likely to elicit risk neutrality and is, therefore, the most effective design for decision-making. Second, they found that participants who saw a *bar* chart exhibited behavior that was slightly risk-averse, mirroring behavior in the control text-only group. Third, the *triangle* chart and *circle* chart elicited risk-seeking behavior with the greatest deviation from risk neutrality. It is important to note that these findings are in line with the magnitude estimation from the prior literature [13] that shows that proportion estimates with *bar* charts are more accurate than with *triangle* and *circle* charts.

4.3.2 Outlook on Using Utility-Optimal Theories for Visualization Evaluation

Although we only highlighted a few studies in this section, it is essential to note that other researchers have also examined decision-making with visualization using a similar framework (e.g., [10, 26, 31, 49, 71]). For example, Padilla et al. [49] conducted a scenario where participants made resource allocation judgments by comparing the cost of sending cold-weather aid to alpaca farmers in Peru who were at risk of losing their livestock due to cold temperatures and the expected

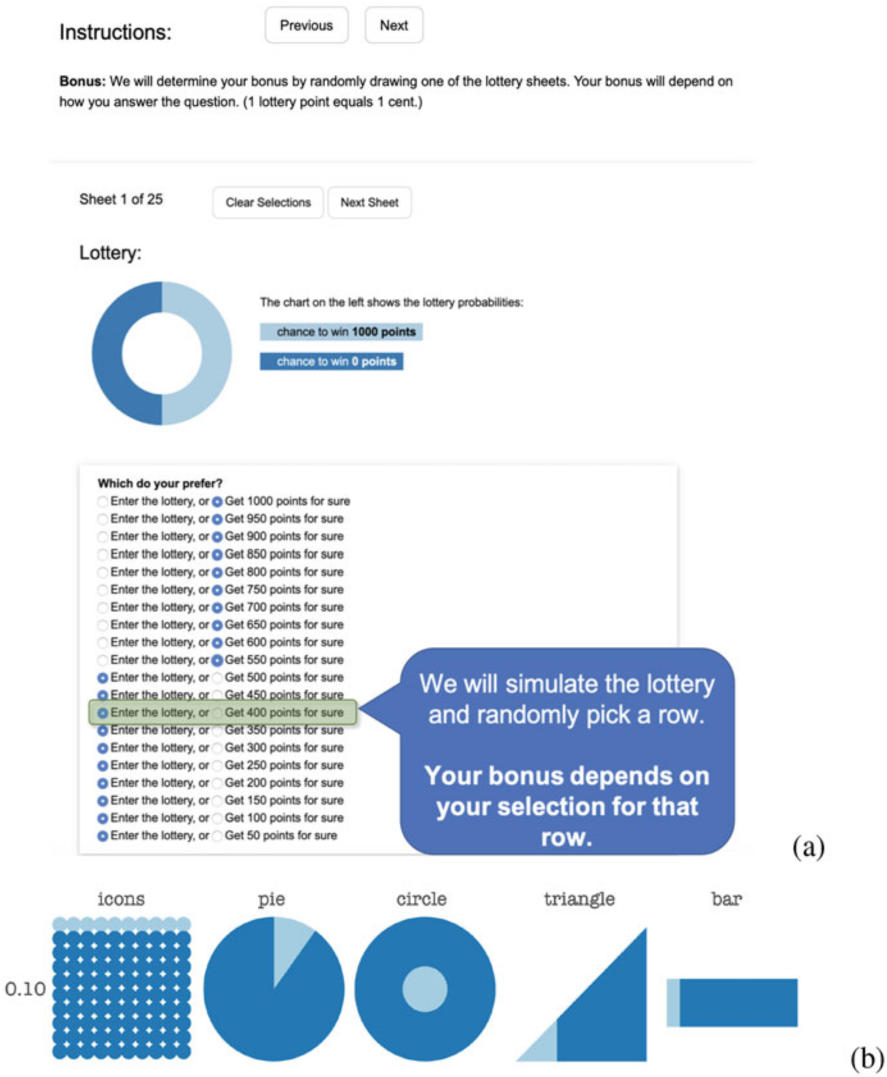

Fig. 4.2 The charts and lottery sheet used in the study by Bancilhon et al. [4]. Participants played a gambling game in which their choices determined their bonuses

value of the penalty for not sending aid, resulting in the deaths of alpacas (see also, [10, 31]). Perhaps most importantly, for visualization evaluation, the utility-optimal perspective provides a tractable approach to quantifying and modeling decision-making under risk. In both Kale et al.'s and Bancilhon et al.'s studies [4, 36], the researchers leveraged the framework to isolate the effect of visualization design. In some cases, their results suggest that using visualizations might help to reduce biases and guide people towards utility-optimality [4].

It is typical for researchers to design games or simulations to observe people's decisions in action. In many cases, it is difficult, if not impossible, to test the impact of visualizations on decisions in real life as it may give rise to safety, health, and ethical issues. For example, it might be unsafe and unethical for a gambling game to test the effect of visualizations that communicate information about a severe health condition that a participant has or a natural disaster affecting the participant at the time of the study. The utility-optimal framework using the situational scenarios in the two studies [4] and [36] provides a good test bed for evaluating visualizations for decision-making. In order to apply this framework to behavioral studies, there needs to be a cost associated with each course of action. The utility-optimal decision should be defined as the one where prospective gains are maximized and losses are minimized. By quantifying user choices and comparing them to the utility-optimal decision, we can infer the risk behavior elicited by the visualization design. It is important to take into account people's patterns of risk behavior since humans do not normally default to risk neutrality regardless of the type of representation used. By providing an incentive to decision-makers, such an experiment design can more closely mimic real-life choices over hypothetical decision scenarios.

While Bancilhon et al. [4] have shown that the visualizations that lead to better accuracy also induce more optimal decisions, Kale et al. [36] have shown that the visualization designs that lead to the least bias did not lead to the most optimal decisions and vice versa. First, this shows that task and visualization choice matter in evaluation. Second, it raises an important question: how do we define the best visualization when accuracy and utility-optimal decisions are inconsistent? In Kale et al.'s study [36], one approach to determine the best uncertainty visualization would be to pick the one with the best compromise between high accuracy and optimal decision-making. Huang et al. [27] have developed a model of visualization efficacy that includes speed, accuracy, and cognitive load, which is often overlooked. One way forward could be to refine this model to include decision-making. Another approach would be to simply not attempt to choose a single best visualization for reasoning about uncertainty. Kale et al. [36] have shown that different visualizations are best for different tasks. There needs to be a common recognition in the visualization community that a one-size fits all approach could be obsolete.

Furthermore, using utility-optimality for visualization evaluation raises another crucial question: how do we define the *best* decisions? Some would argue that rationality should be the golden standard since it maximizes the potential outcome. Bancilhon et al. [4] question whether or not that should be the case. If the goal is rationality, their findings suggest that the icon array was the most likely to elicit risk-neutral choices. However, since people make decisions according to their personal inclination to risk, there might be a cost in attempting to steer them toward utility-optimality. Perhaps an ideal visualization should support the users in making a decision based on their individual risk behaviors.

In the next section, we examine a different perspective on decision-making, positing that humans default to intuitive reasoning when making decisions. We discuss working memory as a metric for usability in visualization decision-making (Fig. 4.3).

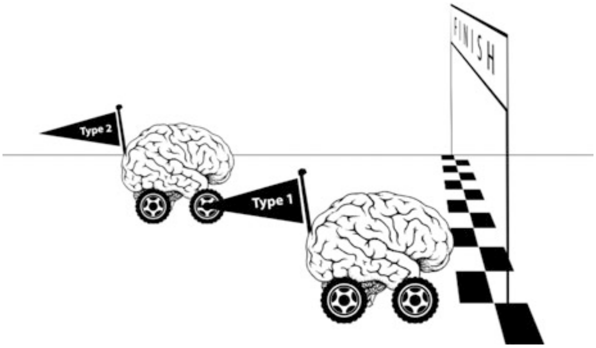

Fig. 4.3 An illustration of Type 1 and Type 2 reasoning as conceptualized by Tversky and Kahneman [33]. Type 1, our intuitive system, is at the forefront of decision processes, while Type 2, our analytic system, operates secondarily

> *Working memory* consists of various mental components that can hold a limited amount of transformable information for a finite period [14]. In visualization research, working memory is commonly associated with mental effort [47]. Note that there is an ongoing debate on the definition of working memory [14]

4.4 The Dual-Process Perspective

In addition to the biases associated with gains and losses (e.g., prospect theory), many other cognitive biases are involved when making decisions under risk. One perspective that describes a large body of biases proposes that people rely on quantitative reasoning and gist-based intuition—two systems that operate in parallel [33].

Daniel Kahneman published a book entitled *Thinking Fast and Slow*, where he summarized decades of research on a dual-system of decision-making [33]. In his earlier work, his collaborators and himself differentiated between two types of processing systems, termed *System 1* (or intuition) and *System 2* (or reasoning) [32] (later termed Type 1 and Type 2). Type 1 processing guides our intuition and recognition patterns, which occur automatically without effort. In contrast, Type 2 processing is responsible for analytical thinking and requires directed effort to use [33].

Dual-Process Theory introduces a reasoning model that formalizes the differences between Type 1 and Type 2 and their impact on decision-making [34, 67]. Proponents of Dual-Process Theory posit that most decisions stem from intuitive thinking rather than rational and calculated thinking [33]. Type 1 reasoning involves fast and intuitive thinking, while Type 2 is a slow and analytical method of thinking.

Scholars propose that Type 2 processing uses significant working memory, while Type 1 only uses negligible working memory [18]. Using this definition, the researchers can determine when people are using Type 2 processing by identifying when people show an increase in working memory demand. Visualization researchers have demonstrated how to measure an increase in working memory demand using pupillometry (e.g., dilation of pupils [47]), dual-tasking (e.g., doing two tasks a once [11, 47]), individual differences measures (e.g., working with participants with high- and low-working memory capacity [10]), the NASA-TLX (e.g., self-reported work-load [10]), and electroencephalography (e.g., neuroimaging [1]). Type 1 is at the forefront of cognitive processes, and it often requires significant effort to switch from Type 1 to Type 2 in order to avoid cognitive biases and misleading heuristics. Despite utilizing different strategies, dual-process theories propose that the processes do not necessarily occur in separate cognitive or neurological systems [19].

Other frameworks have adapted the general dual-process perspective as well. Notably, Reyna and Brainerd introduced *Fuzzy Trace Theory* (FTT) [58]. The theory posits that people form two types of mental representations from information: *Gist* and *Verbatim* representations. A verbatim representation is a detailed representation of an event that often comprises precise numbers and facts. Gist representation, on the contrary, is vague and high-level and captures the essential meaning of information. FTT asserts that people make decisions by extracting meaning from verbatim input to make a gist-based judgment. According to Reyna and Brainerd [58], the human memory contains various reasoning-relevant information, ranging from preserving the exact form of input or only retaining abstract representations. People operate somewhere between the highest level of gist and the highest level of verbatim, on a gist-to-verbatim continuum [58]. Typically, humans rely on the least precise gist representation necessary to make a decision, and this characteristic is generally referred to as “fuzzy processing preference” [58].

Although there is a long history of theories on dual-processes, the high-level ideas are similar. They assert that there are two kinds of reasoning. One is implicit, intuitive, and unconscious, and the other is explicit, conscious, and slow. For simplicity, we will refer to this general class of theories as *Dual-Process* theories.

4.4.1 Dual-Process in Decision-Making

Fuzzy Trace Theory states that people make decisions by extracting meaning from verbatim input to make a gist-based judgment. Because precision is often associated with accuracy, many believe that quantitative reasoning is superior to qualitative reasoning. However, in some cases, fuzzy representation of information does not affect reasoning accuracy [60]. Reyna and Lloyd [59] have shown that experts in the medical field tend to engage more in gist-based decision-making than novices. Tversky and Kahneman made the argument that intuition is a synonym for recognition [33]. Experts recognize familiar situations and can therefore make fast and accurate decisions even when they are complex.

Although Type 1 has been proven to be efficient [59, 60], it is also more susceptible to false first impressions and framing effects [33]. Consider the following question:

A bat and ball cost $1.10. The bat costs $1 more than the ball. How much does the ball cost?

More than 50% of students at Harvard, Princeton, and the Massachusetts Institute of Technology routinely gave the incorrect answer, insisting the ball costs 10 cents [33].[1] Type 1 is at the forefront of cognitive processes, and in order to obtain the correct answer, a switch from Type 1 to Type 2 is required to overcome cognitive biases.

Before the acknowledgement of the role of Type 1, many believed that Type 2 was solely in charge of decision-making operations. Expected Utility Theory posits that people make decisions rationally, using Type 2 to compute the utility of events. The recognition of dual modes of reasoning lead to the development of prospect theory [35] (see Sect. 4.3) and revolutionized decision-making research.

4.4.2 Dual-Processes and Visualization Evaluation

In the medical field, researchers have investigated the impact of visualization design on gist reasoning. For example, Feldman et al.'s first goal [20] was to investigate which graphical formats induced the most accurate perception of quantitative information in patients making treatment decisions. Second, they inquired about the formats that facilitate processing. The authors highlight the importance of ease of processing, especially when the patient feels overwhelmed by the diagnostic. They conducted an experiment to test the performance of variations of 6 visualization formats. Participants had to minimize how long the visualizations appeared on the screen while remaining accurate when answering questions about the charts. They were shown two quantities and were asked to make a gist judgment by choosing the one that showed the larger chance of survival or the smaller chance of side effects. They were also asked to make a verbatim judgment by determining the size of the difference.

In this study, Feldman et al. [20] used response time as a proxy for ease of information processing. Their results suggest that systematic ovals, which encode data in a natural frequency format, are likely the format that represents the best compromise for accurate processing of both gist and detailed information while also demanding relatively little effort. Similarly, Hawley et al. [24] conducted an experiment investigating gist and verbatim reasoning through similar comparison

[1] The correct answer to this problem is that the ball costs 5 cents and the bat costs –at a dollar more– $1.05 for a grand total of $1.10.

and estimation tasks. They found that viewing a pictograph was associated with adequate levels of both gist and verbatim knowledge and that superior medical treatment choices were made in both cases.

In their work, Feldman et al. [20] question the overall effectiveness of vertical bars with scales, which was the best visualization for gist reasoning. The authors state that many patients demand detail-level information, and they defined the best visualization as the one that is effective in eliciting both types of reasoning. While this prior work gives evidence that charts using natural frequency encoding perform better under both gist and verbatim reasoning in comparison tasks, further research is required to examine whether the findings are generalizable to other tasks.

4.4.3 Outlook on Using the Dual-Processing Approach for Visualization Evaluation

While the Expected Utility Framework provides a method to mathematically model decisions, the Dual-Process framework is not straightforward. Feldman et al. [20] and Hawley et al. [24] have studied how visualization affects Type 1 and Type 2 reasoning in a comparison task. Note that it is possible for both processes to be used to make a decision. In their respective work, they posit that a magnitude estimation task brings about Type 2 reasoning, whereas asking the participant to make a comparison choice triggers Type 1 reasoning. If we apply this inference to Bancilhon et al.'s lottery game study [4] in Sect. 4.3.1.2, their results are consistent with Feldman et al.'s work [20] since the icon array outperforms the other visualizations in the decision task. Considering Kale et al.'s fantasy football study [36] in Sect. 4.3.1.1, which observed a magnitude estimation task and a decision task, it is possible that the selected visualizations have different effects under Type 1 and Type 2 reasoning.

However, our conclusions are solely based on the assumption that the tasks used actually elicit two distinct types of reasoning. To further research in this area, we need to answer the following research questions, which are core to understanding the role of visualization in decision-making:

- How does the mode of reasoning influence decision-making when using visualizations?
- Can different visualizations elicit different modes of reasoning?

It is crucial to understand how people make decisions from visualizations. Understanding whether a visual encoding facilitates gist or verbatim reasoning can have enormous implications for visualization designers. By expanding our knowledge in this area, we can tailor visualizations to our audience or a specific problem area. Bridging the gap between how psychologists and visualization researchers reason about decision-making can revolutionize how we evaluate and design visualizations.

Such knowledge can have massive implications for visualization designers. For example, visualizations can be tailored and personalized to a specific problem area or level of audience expertise. Some visualizations are only seen for a short time so we need a quick way of displaying information so that people get the gist of it. Moreover, some people might be more prone to gisting and others to probabilistic reasoning. Factors such as numeracy and spatial ability likely play a role.

Further investigations are needed to understand *how* people reason under this dual mode and how it affects their decisions. In the following sections, we examine cognitive models of decision-making with visualization and advocate for their integration into visualization research to deepen our understanding of decision-making processes with different charts.

4.5 Cognitive Models of Decision-Making with Visualization

Cognitive models are an integration of approaches and can be illustrated as process diagrams that conceptualize their mechanisms processes. By applying a cognitive model to a problem, a visualization researcher can better understand, model, or even evaluate the interaction between the user and the visual design at a cognitive level of analysis, as opposed to strictly behavioral. Before diving into the integration of a dual-process approach into decision-making research with visualization, we must first understand how the mind perceives and understands visualization. Pinker [55] proposed a cognitive model depicting the distinction between two mechanisms in graph comprehension: bottom-up and top-down mechanisms [55].

> *Bottom-up* processing is when the mind is directly influenced by a visual stimulus which is utilized to construct a visual description.
> *Top-down* processing is based on the viewer's goals, experiences, and other individual differences.

Prior knowledge about the graph is then retrieved from long-term memory in the form of an established graph schema. It is essential to point out that with familiar charts, the visual schema will be retrieved from memory faster and more efficiently, facilitating Type 1 reasoning [48]. This *match process* also occurs when visual properties are altered. The viewer then retrieves the graph schema that is the most similar to the visual array. When a graph schema is retrieved, the viewer uses the information from the graph schema to interpret the visualization. Bottom-up attention is influenced by saliency in the visualization design. Features that attract bottom-up attention are color, edges, lines, and foreground information.

Graph schema is memorized graphic conventions [55].

When external factors impact knowledge retrieval, the viewer is considered to be taking a top-down approach. Top-down attention is based on the viewer's goals, experiences, and other individual differences. There are other factors that can affect visualization comprehension, such as the nature of the task. Viewers may need to transform their mental representation of the visualization based on their task or conceptual questions, and working memory plays are central role in the process (Fig. 4.4).

4.5.1 Padilla's Dual-Process Model and the Importance of Working Memory

Padilla et al. [48] devised a model that combines theories of visualization comprehension, decision-making, and working memory. The motivation for this work is the lack of formalization of research from different fields, making it difficult for scientists to integrate cross-domain findings. The authors explored a cognitive model of decision-making with visualizations and provide practical recommendations for visualization designers.

In the previous section, we defined two types of graph comprehension mechanisms: bottom-up and top-down. The understanding of these two mechanisms is crucial in the understanding of Padilla's Dual-Process Model, with the addition of working memory, which are the mental processes associated with effort [48].

Padilla et al. [48] assert that working memory plays an important role in decision-making, but it is often overlooked by visualization researchers as an evaluation factor. Before diving into how working memory is involved in the dual reasoning system, let's look at some of its properties. It is important to note that working memory capacity is limited [42, 63]. Working memory also increases with task difficulty and diminishes over time. Researchers such as Cowan et al. [15] suggest that our ability to store information begins to decay after approximately 5–10 seconds, depending on factors such as the task, type of information, and individual differences in working memory capacity. One property of working memory capacity that is relevant to dual-process theory is that it limits the amount of attention we can allocate to task-relevant information [48].

Padilla et al.'s model [48] suggests that when we deliberately employ working memory in our decision-making process, we can make slower and more strategic but cognitively demanding decisions with visualizations. In other words, working

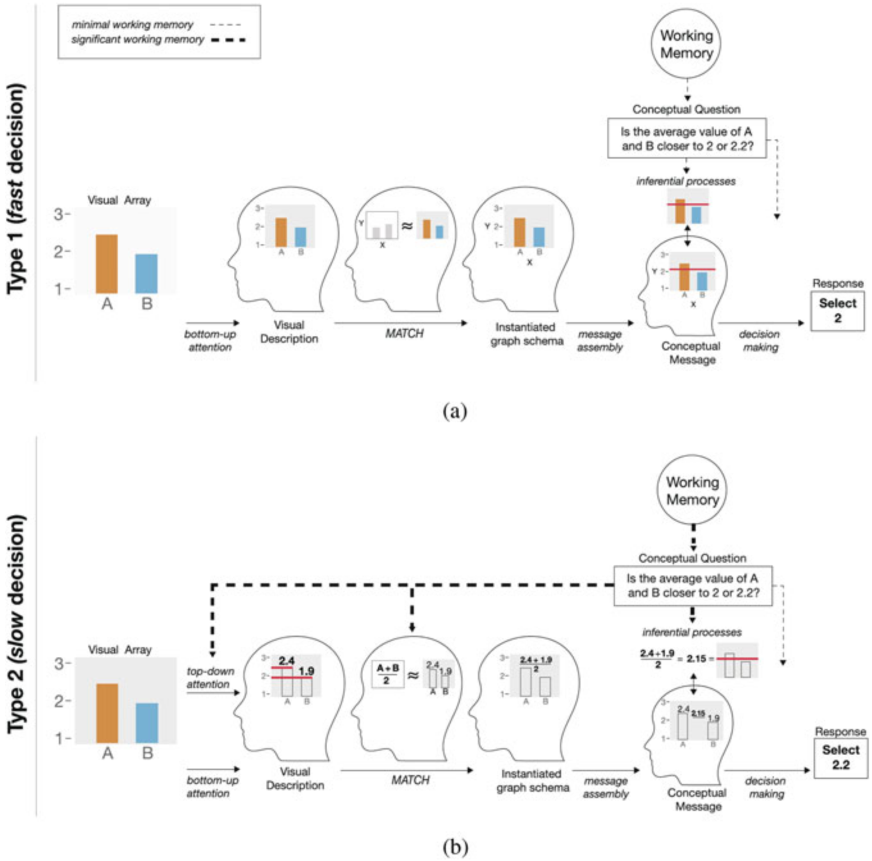

Fig. 4.4 An illustrative example of Type 1 versus Type 2 decision as characterized by Padilla et al.'s model [48]. (**a**) An example of a Type 1 decision process in which the viewer is tasked with computing the average of the two bars in the graph. A Type 1 approach might make a quick guess of the middle point between the two bars. (**b**) An illustration of a Type 2 decision process. The task is the same as subfigure (**a**) above. In this example, the viewer takes a slower approach and estimates the length of each bar. They then compute the average of the two values $\frac{2.4+1.9}{2}$. Type 2 activates working memory and can lead to a more effortful but precise estimate if done correctly

memory is what we use to switch from Type 1 reasoning (requiring nominal working memory) to Type 2 (requiring significant working memory). As described in the previous section, both Type 1 and Type 2 reasoning can be used to complete the decision step. Differences in working memory capacity can influence judgments and consequently decision-making. Lohse [41] found that when participants made judgments about budget allocation using profit charts, individuals with less working memory capacity performed equally well compared to those with more working memory capacity when they only made decisions about three regions (easier task). However, when participants made judgments about nine regions (harder task), individuals with more working memory capacity outperformed those with less working memory capacity. Other work finds that participants with low-working memory capacity make more accurate resource allocation decisions when using density plots and quantile dot plots compared to 95% confidence intervals, point estimates, or textual expressions of uncertainty [10]. Furthermore, participants with high-working-memory capacity were most accurate with quantile dot plots and reported less effort than all other tested methods. This work suggests that 95% confidence intervals, point estimates, are textual expressions of uncertainty require more working memory than densities and quantile dot plots [10]. The results of this study suggest that individual differences in working memory capacity primarily influence performance on complex decision-making tasks [10, 41].

4.5.2 Outlook on Using Cognitive Models in Visualization

Padilla et al.'s cognitive model [48] in Sect. 4.5.1 formalizes the implications of this dual mode of reasoning for visualization research. This cognitive model is an integration of multiple theories and takes a holistic approach to modeling decision-making with visualization. Applying this model can have a significant impact on design and evaluation of visualization interfaces. We provide some practical guidance for designers and visualization researchers on how to leverage visual features to generate Type 1 or Type 2 reasoning and evaluate visualization designs from a dual-process perspective.

One of the reasons why visualizations are so prominent is because they seem effortless. In other words, to design charts that bring about accurate, fast, and effortless reasoning, there needs to be a conscious effort to incorporate design principles that elicit bottom-up attention on task-relevant information. Padilla's model proposes that bottom-up attention is associated with Type 1 reasoning and top-down attention is more likely to generate Type 2 reasoning. Using this principle, Padilla et al. allow us to examine core design questions and provide guidelines to elicit either reasoning type by altering visual features.

Modeling visual attention is an important area of research in psychophysics, computational modeling, and neurophysiology (see a review of existing work by Borji and Itti [7]). When making a choice, the decision-maker must first decode the visualization via their visual system [70]. One way to elicit bottom-up attention is to align visual features to the users' existing graph schema. Figure 4.5 shows a figure from Padilla et al. where at first glance, it might appear that the introduction of the predator species caused a decline in the population of disease X [48]. If we look more closely at the graph, we notice that the y-axis is flipped and the predator species in fact contributed to the growth of species X. When decoding a visualization, we search our long-term memory for knowledge about how to interpret the chart and retrieve the graph schema that is the most similar. Altering graph conventions can cause errors because the graph schema will no longer match the chart. For example, multiple studies find that when the y-axis is inverted people

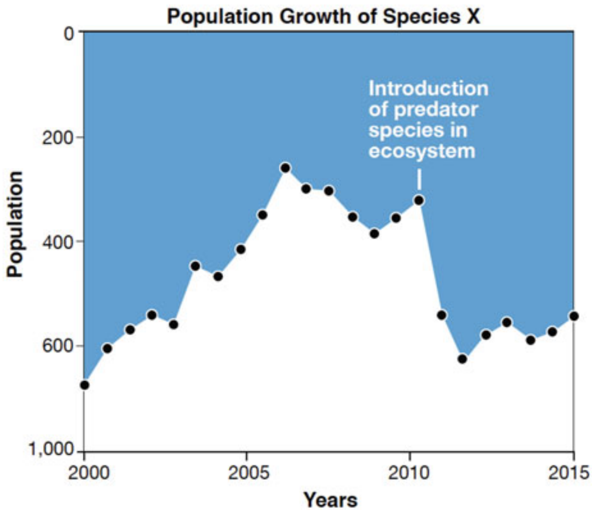

Fig. 4.5 Fictional relationship between the population growth of Species X and a predator species, where the Y-axis ordering does not match standard graphic conventions [48]

consistently come to the wrong interpretation of the chart [52, 72]. These errors are likely due to our reliance on graph schema to interpret graphs so much so that we do not notice when the schema does not match the chart.

One of the main design features that can affect decision type is saliency. Numerous studies showed that salient information in a visualization draws viewers' attention (e.g., [25, 25, 30, 45, 50, 61, 68]). First, it is important to identify the main piece of information that needs to be communicated and then we can direct the user's attention to this information using visual features. There exist behaviorally validated saliency models to determine the prominence of different visual encodings that will attract viewer's bottom-up attention, e.g., [28–30]. There is a long history of using saliency algorithms in computational imagery. For example, pioneering work by Koch and Ullmnan [39] created a *saliency map*—a two-dimensional topological map that encodes conspicuity across the entire scene. The central thesis of their work is that salient features within a stimulus "stand out," thus attracting overt attention. There have been some attempts in the visualization community to use this general principle to model visual attention in exploratory search tasks [45]. Still, future work is needed to model attention in the context of decision-making.

A critical component of Padilla et al.'s model is the principle that working memory is vital for Type 2 processing [48]. It is possible to gain insight into the type of decision-making generated by a visualization by measuring the user's working memory capacity. The amount of working memory generated by a task is commonly referred to as *cognitive load*. Remember that Type 1 reasoning does not require significant working memory contrarily to Type 2. There exists some prior work where the researchers have used measures of working memory to evaluate ease of use of visualization. Borgo et al. challenged traditional notions about chart junk and showed that embellishments do not generate higher cognitive load compared to other visualizations. By using a dual-task paradigm to evaluate different charts, they were able to evaluate differences in working memory elicited by different charts [6] by observing the dual-task cost. *Dual-task cost* is described as the decrease in performance between single and dual tasks. When the user completes two tasks simultaneously, significant memory is required, and by comparing dual-task cost across representations, differences in cognitive load can be inferred. There are a number of other ways to measure working memory. Castro et al. investigated the effect of various uncertainty visualizations on working memory using an operation span (OSPAN) task as part of a dual-task paradigm as well as self-reported measures [10]. They found that quantile dot plots and density plots are equally effective for low-working-memory individuals, while quantile dot plots elicit more accurate responses with less perceived effort for high-working-memory individuals. Moreover, Peck et al. used fNIRS to evaluate information visualization interfaces and found no difference in cognitive load in bar graphs and pie charts [54]. Other physical methods include *electroencephalogram (EEG)* [2] and *pupillometry*, which has shown high levels of correlation with working memory [47].

To summarize, two practical ways to elicit decision type are to design according to *graph schema* and *saliency*. For example, to elicit Type 1 reasoning, some elementary steps include verifying that your visualization does not violate any graphical conventions and brings forward important information using salient visual features. To examine decision type, one can observe *working memory* through self-reported measures, behavioral, and psychological methods. Padilla et al.'s model [48] is the most updated description of decision-making with visualizations, and we advocate that research incorporates this model when evaluating visualization design. Although we examined various decision-making models that appear in prior literature, they do not describe the entire visualization decision-making process using dual-process theory. For example, other models do not account for how framing effects of the visual or textual data might influence decisions [46]. Other factors such as individual differences (e.g., working memory capacity or spatial ability) can mediate the decision process [40, 44, 73] but are not encompassed in other models. Numerous researchers have voiced the importance of diversifying evaluation measures in the field of visualization [5], which is possible when using a cognitive framework. Ultimately, this chapter advocates for measures beyond the traditional usability measures, which capture *how* and *why* the brain processes visualizations.

4.6 Conclusion

Adopting decision models can have a significant impact on chart design and visualization evaluation. For instance, measuring working memory will diversify visualization research by tailoring chart design to individuals with varying levels of working memory capacity. Knowledge about dual-process reasoning and insight into cognitive load will enable tailoring visualization design to various tasks. We assert that for visualization to be reliably effective in real-world decision-making settings, research should consider leveraging existing decision theories when evaluating visual designs. We reviewed various utility-optimal theories, dual-process models, and cognitive science frameworks and discussed existing and future directions for visualization research. Much of the work discussed in this chapter raises valid concerns about evaluation paradigms that emphasize speed and accuracy measures. Overall, we advocate for evaluation techniques that go beyond traditional usability measures for better theoretical and practical advancements.

References

1. E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva. A user study of visualization effectiveness using EEG and cognitive load. *Computer Graphics Forum*, 30(3):791–800, 2011.
2. E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva. A user study of visualization effectiveness using EEG and cognitive load. *Computer Graphics Forum*, 30(3):791–800, 2011.
3. B. W. Balleine. The neural basis of choice and decision making. *Journal of Neuroscience*, 27(31):8159–8160, 2007.
4. M. Bancilhon, Z. Liu, and A. Ottley. Let's gamble: How a poor visualization can elicit risky behavior. *In proceedings of IEEE VIS Short Papers*, 2020.
5. E. Bertini, A. Perer, C. Plaisant, and G. Santucci. Beliv'08: Beyond time and errors: novel evaluation methods for information visualization. In *CHI'08 extended abstracts on Human factors in computing systems*, pages 3913–3916. 2008.
6. R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, L. Floridi, and M. Chen. An empirical study on using visual embellishments in visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2759–2768, 2012.
7. A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):185–207, 2012.
8. A. Bruhin, H. Fehr-Duda, and T. Epper. Risk and rationality: Uncovering heterogeneity in probability distortion. *Econometrica*, 78(4):1375–1412, 2010.
9. S. Carpendale. Evaluating information visualizations. In *Information visualization*, pages 19–45. Springer, 2008.
10. S. Castro, P. S. Quinan, H. Hosseinpour and L. Padilla. (2022). Examining Effort in 1D Uncertainty Communication Using Individual Differences in Working Memory and NASA-TLX. *IEEE transactions on visualization and computer graphics, 28*(1), 411–421.
11. S. C. Castro, D. L. Strayer, D. Matzke, and A. Heathcote. Cognitive workload measurement and modeling under divided attention. *Journal of experimental psychology: human perception and performance*, 45(6):826, 2019.
12. W. S. Cleveland and R. McGill. An experiment in graphical perception. *International Journal of Man-Machine Studies*, 25(5):491–500, 1986.
13. W. S. Cleveland and R. McGill. Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society: Series A (General)*, 150(3):192–210, 1987.
14. N. Cowan. The many faces of working memory and short-term storage. *Psychonomic bulletin & review*, 24(4):1158–1170, 2017.
15. N. Cowan, J. S. Saults, and L. D. Nugent. The role of absolute and relative amounts of time in forgetting within immediate memory: The case of tone-pitch comparisons. *Psychonomic Bulletin & Review*, 4(3):393–397, 1997.

16. E. Dimara, S. Franconeri, C. Plaisant, A. Bezerianos, and P. Dragicevic. A task-based taxonomy of cognitive biases for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(2):1413–1432, 2020.
17. N. Elmqvist and J. S. Yi. Patterns for visualization evaluation. *Information Visualization*, 14(3):250–269, 2015.
18. J. S. B. Evans and K. E. Stanovich. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, 8(3):223–241, 2013.
19. J. S. B. Evans and K. E. Stanovich. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3):223–241, 2013.
20. D. Feldman-Stewart, M. D. Brundage, and V. Zotov. Further insight into the perception of quantitative information: Judgments of gist in treatment decisions. *Medical Decision Making*, 27(1):34–43, 2007.
21. M. Friedman and L. J. Savage. The utility analysis of choices involving risk. *Journal of political Economy*, 56(4):279–304, 1948.
22. M. Galesic and R. Garcia-Retamero. Using icon arrays to communicate medical risks: Overcoming low numeracy. *Health Psychology*, 28(2):210–216, 2009.
23. A. Hakone, L. Harrison, A. Ottley, N. Winters, C. Gutheil, P. K. Han, and R. Chang. Proact: iterative design of a patient-centered visualization for effective prostate cancer health risk communication. *IEEE transactions on visualization and computer graphics*, 23(1):601–610, 2016.
24. S. T. Hawley, B. Zikmund-Fisher, P. Ubel, A. Jancovic, T. Lucas, and A. Fagerlin. The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Education and Counseling*, 73:448–455, 2008.
25. M. Hegarty, A. Friedman, A. P. Boone, and T. J. Barrett. Where are you? the effect of uncertainty and its visual representation on location judgments in GPS-like displays. *Journal of experimental psychology: applied*, 22(4):381, 2016.
26. J. M. Hofman, D. G. Goldstein, and J. Hullman. How visualizing inferential uncertainty can mislead readers about treatment effects in scientific results. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.
27. W. Huang, P. Eades, and S.-H. Hong. Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*, 8(3):139–152, 2009.
28. L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10–12):1489–1506, 2000.
29. L. Itti and C. Koch. Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.
30. L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.
31. S. L. Joslyn and J. E. LeClerc. Uncertainty Forecasts Improve Weather-Related Decisions and Attenuate the Effects of Forecast Error. *Journal of Experimental Psychology: Applied*, 18(1):126–140, 2012.
32. D. Kahneman. A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, 58(9):697, 2003.
33. D. Kahneman. *Thinking Fast and Slow*. Farrar, Straus and Giroux, New York, 2011.
34. D. Kahneman and S. Frederick. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49:81, 2002.
35. D. Kahneman and A. Tversky (2013) Prospect Theory: An Analysis of Decision under Risk. In: MacLean, L.C. and Ziemba, W.T., Eds., Handbook of the Fundamentals of Financial Decision Making: Part I, World Scientific Publishing, Hackensack, 99–127.
36. A. Kale, M. Kay, and J. Hullman. Visual reasoning strategies for effect size judgments and decisions. *IEEE Transactions on Visualization and Computer Graphics*, 2020.
37. M. Kay, T. Kola, J. Hullman, and S. Munson. When (ish) is my bus?: User-centered visualizations of uncertainty in everyday mobile predictive systems. *In Proceedings of the CHI Conference on Human Factors in Computing System*, page 5092–5103, 2016.

38. R. L. Keeney and H. Raiffa. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press, 1993.
39. C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. In *Matters of intelligence*, pages 115–141. Springer, 1987.
40. Z. Liu, R. J. Crouser, and A. Ottley. Survey on individual differences in visualization. *Computer Graphics Forum*, 39(3):693–712, 2020.
41. G. L. Lohse. The role of working memory on graphical information processing. *Behaviour & Information Technology*, 16(6):297–308, 1997.
42. G. A. Miller. *11. The magical number seven, plus-or-minus two or some limits on our capacity for processing information*. University of California Press, 2020.
43. T. Munzner. A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics*, 15(6):921–928, 2009.
44. A. Ottley. Adaptive and personalized visualization. *Synthesis Lectures on Visualization*, 7(1):1–117, 2020.
45. A. Ottley, R. Garnett, and R. Wan. Follow the clicks: Learning and anticipating mouse interactions during exploratory data analysis. *Computer Graphics Forum*, 38(3):41–52, 2019.
46. A. Ottley, E. M. Peck, L. T. Harrison, D. Afergan, C. Ziemkiewicz, and R. Taylor, H. A.... & Chang. Improving Bayesian reasoning: The effects of phrasing, visualization, and spatial ability. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):529–538, 2015.
47. L. M. Padilla, S. C. Castro, P. S. Quinan, I. T. Ruginski, and S. H. Creem-Regehr. Toward objective evaluation of working memory in visualizations: a case study using pupillometry and a dual-task paradigm. *IEEE transactions on visualization and computer graphics*, 26(1):332–342, 2019.
48. L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. Decision making with visualizations: A cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, 3(1):1–25, 2018.
49. L. M. Padilla, M. Powell, M. Kay, and J. Hullman. Uncertain about uncertainty: How qualitative expressions of forecaster confidence impact decision-making with uncertainty visualizations. *Frontiers in Psychology*, 11, 2020.
50. L. M. Padilla, I. T. Ruginski, and S. H. Creem-Regehr. Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive research: principles and implications*, 2(1):1–16, 2017.
51. L. M. K. Padilla, S. C. Castro, P. S. Quinan, I. T. Ruginski, and S. H. Creem-Regehr. Toward objective evaluation of working memory in visualizations: A case study using pupillometry and a dual-task paradigm. *in IEEE Transactions on Visualization and Computer Graphics*, 26(1):332–342, Jan. 2020.
52. A. V. Pandey, K. Rall, M. L. Satterthwaite, O. Nov, and E. Bertini. How deceptive are deceptive visualizations? an empirical analysis of common distortion techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1469–1478, 2015.
53. E. M. Peck, B. F. Yuksel, L. Harrison, A. Ottley, and R. Chang. Position paper: towards a 3-dimensional model of individual cognitive differences. *Beyond Time & Errors–Novel Evaluation Methods for Visualization (Proceedings of BELIV 2012)*, 2012.
54. E. M. M. Peck, B. F. Yuksel, A. Ottley, R. J. Jacob, and R. Chang. Using fNIRS brain sensing to evaluate information visualization interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 473–482. ACM, 2013.
55. S. Pinker. A theory of graph comprehension. *Artificial intelligence and the future of testing*, pages 73–126, 1990.
56. C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the working conference on Advanced visual interfaces*, pages 109–116, 2004.
57. G. Loewenstein and D. Prelec. (1991). Decision Making Over Time and Under Uncertainty: A Common Approach. Management Science. 37. 770–786. 10.1287/mnsc.37.7.770.
58. V. Reyna and C. Brainerd. Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences*, 7(1):1–75, 1995.

59. V. Reyna and F. Lloyd. Physician decision making and cardiac risk: Effects of knowledge, risk perception, risk tolerance, and fuzzy processing. *Journal of Experimental Psychology: Applied*, 12(3):179–195, 2006.
60. V. F. Reyna. A theory of medical decision making and health: fuzzy trace theory. *Medical Decision Making*, 28(6):850–865, 2008.
61. I. T. Ruginski, A. P. Boone, L. M. Padilla, L. Liu, N. Heydari, H. S. Kramer, M. Hegarty, W. B. Thompson, D. H. House, and S. H. Creem-Regehr. Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition & Computation*, 16(2):154–172, 2016.
62. J. G. Ruiz, A. D. Andrade, R. Garcia-Retamero, R. Anam, R. Rodriguez and J. Sharit. (2013). Communicating global cardiovascular risk: are icon arrays better than numerical estimates in improving understanding, recall and perception of risk?. *Patient education and counseling, 93*(3), 394–402.
63. T. L. Saaty and M. S. Ozdemir. Why the magic number seven plus or minus two. *Mathematical and computer modelling*, 38(3–4):233–244, 2003.
64. B. Saket, A. Endert, and J. Stasko. Beyond usability and performance: A review of user experience-focused evaluations in visualization. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 133–142, 2016.
65. D. Simkin and R. Hastie. An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82(398):454–465, 1987.
66. I. Spence and S. Lewandowsky. Displaying proportions and percentages. *Applied Cognitive Psychology*, 5(1):61–77, 1991.
67. K. E. Stanovich. *Who is rational?: Studies of individual differences in reasoning*. Psychology Press, 1999.
68. E. R. Stone, W. R. Sieck, B. E. Bull, J. F. Yates, S. C. Parks, and C. J. Rush. Foreground: background salience: Explaining the effects of graphical displays on risk avoidance. *Organizational behavior and human decision processes*, 90(1):19–36, 2003.
69. A. Tversky. A critique of expected utility theory: Descriptive and normative considerations. *Erkenntnis*, 9(2):163–173, 1975.
70. C. Ware. *Visual thinking for design*. Elsevier, 2010.
71. S. J. Westwood, S. Messing, and Y. Lelkes. Projecting confidence: How the probabilistic horse race confuses and demobilizes the public. *The Journal of Politics*, 82(4):1530–1544, 2020.
72. G. Woodin, B. Winter and L. Padilla, "Conceptual Metaphor and Graphical Convention Influence the Interpretation of Line Graphs," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 2, pp. 1209–1221, 1 Feb. 2022.
73. C. Ziemkiewicz, A. Ottley, R. J. Crouser, K. Chauncey, S. L. Su, and R. Chang. Understanding visualization by understanding individual users. *IEEE computer graphics and applications*, 32(6):88–94, 2012.

Chapter 5
Supporting Diverse Research Methods for Observing Huge Variable Space in Empirical Studies for Visualization

Alfie Abdul-Rahman, Min Chen, David H. Laidlaw, and Brian Fisher

Abstract In each of the last 5 years, a few dozen empirical studies appeared in visualization journals and conferences. The existing empirical studies have already featured a large number of experimental variables. There are many more variables yet to be studied. While empirical studies enable us to obtain knowledge and insight about visualization processes through observation and analysis of user experience, it seems to be a stupendous challenge for exploring such a huge variable space at the current pace. In this chapter, we discuss the implication of not being able to explore this space effectively and efficiently and propose means for addressing this challenge. In particular, we first reason the need for more empirical studies to examine hypotheses about how the "mind" works in visualization and visual analytics (VIS) processes. We then outline several progressive approaches to address such needs. We argue that an important aspect that the VIS research community can learn from psychology is to increase the diversity of publications in studying the "mind." We observe the changing definitions of empirical research papers in IEEE VIS conferences over the past two decades, suggesting an existing trend of increasing the diversity of publications in the field of VIS. We present some statistics about paper types in a number of psychology journals, showing an extensive range of empirical research in terms of paper types. Our analysis supports the arguments for studying the "mind" in the context of VIS, for providing empirical research in

A. Abdul-Rahman (✉)
King's College London, London, UK
e-mail: alfie.abdulrahman@kcl.ac.uk

M. Chen
University of Oxford, Oxford, UK
e-mail: min.chen@oerc.ox.ac.uk

D. H. Laidlaw
Brown University, Providence, RI, USA
e-mail: dhl@cs.brown.edu

B. Fisher
Simon Fraser University, Burnaby, BC, Canada
e-mail: bfisher@sfu.ca

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_5

VIS with a diverse range of paper types, and for further developing the synergy between VIS and psychology.

5.1 Introduction

Many controlled and semi-controlled empirical studies have provided empirical evidence to compare and measure effectiveness and efficiency of different visualization techniques (or approaches, algorithms, systems, workflows, and so on). Some have provided support to existing theories or models for visualization and visual analytics, while several have challenged some commonly known assumptions, wisdoms, and guidelines. Most of such studies consist of one or a few experiments, each features a few independent and dependent variables. One might wish for empirical studies to capture all possible independent variables that may be featured in commonly used visual representations and all dependent variables that could be used to measure the performance of typical visualization tasks. However, the sheer number of these variables presents a hindrance to any controlled or semi-controlled studies. On the other hand, distributing these variables to many studies, each focusing on a few variables, demands a large research community and a lot of resources.

Recently, Abdul-Rahman et al. conducted a survey of 32 empirical study papers [2] in the field of visualization and visual analytics (VIS). They identified 64 types of independent variables and categorized them into five classes. The first four classes (56 types) all focused on visual signals, while the fifth class (8 types) focused on non-visual variables (e.g., task, teaching method, etc.). They observed that "*there is no shortage of studies on independent variables in each category*," but "*there are many more research questions yet to be asked or answered, and the scope of visualization-related empirical studies is huge*." They concluded:

> "*It may thus be desirable for the visualization researchers who conduct empirical studies to be more coherently organized, instead of being distributed sparsely in InfoVis, SciVis, VAST, and other areas of visualization. This will allow these researchers to share their expertise (e.g., in the review processes) more easily and to formulate research agenda in a more ambitious and structured manner.*"
> "*By providing some opportunities to bring all these researchers together, we may soon see the emergence of a new area of **visualization psychology**.*"

This echoes an earlier observation in another survey [1]: "*There are many branches of applied psychology ... One has to ask that 'is there a room for **visualization psychology**?'* " In this chapter, we provide further discourse on how to address the huge variable space in visualization psychology.

In Sect. 5.2, we present our observation of the traditional focuses of empirical research in VIS on hypotheses about the artifacts in visualization images and the needs for more empirical studies to examine hypotheses about how the "mind" works in VIS processes. In Sect. 5.2.3, we outline several progressive approaches to address the need for more studies on the "mind." We argue that an important aspect that the VIS research community can learn from the discipline of psychology is to increase the diversity of publications in studying the "mind."

In Sect. 5.3, we first observe the changing definitions of empirical research papers in the IEEE VIS conferences over the past two decades and recognize an existing trend of increasing the diversity of publications in the field of VIS. We then present some statistics on paper types in a number of psychology journals, showing an extensive range of empirical research in terms of paper types. This indicates that studying the "mind" is facilitated by a diverse range of research activities, which need to be reported and disseminated in publications of a diverse range of paper types.

In Sect. 5.4, we summarize the arguments for studying the "mind" in the context of VIS, for providing empirical research in VIS with a diverse range of paper types, and for further developing the synergy between VIS and psychology. In addition, we propose a set of criteria for evaluating empirical research papers, including artifact- and mind-focused empirical study papers.

5.2 Observations

The main obstacles to the scalability of empirical studies in visualization include (i) the relatively small number of visualization researchers who design and conduct empirical studies, (ii) the complex variations in visualization in a combinatoric manner, and (iii) the narrow hypothesis-based experimental design suitable for publication requirements. A new area of visualization psychology may adopt the following strategies to help overcome these obstacles.

5.2.1 More Experimental Scientists

Building on the references collected by Lam et al. [26], Kijmongkolchai et al. [24], Fuchs et al. [15], and Roth et al. [28], Abdul-Rahman et al. surveyed 129 papers on visualization-focused empirical studies [1] until 2018. Their statistics show that on average the Journal of Psychological Review published about 38 papers per year between 1978 and 2018, while the average number of visualization-focused empirical studies is about 12 per year between 2010 and 2018. Considering that a Wikipedia page [33] lists 144 psychology journals, empirical studies that focus on visualization and visual analytics are drops in the ocean.

The situation is unlikely to improve substantially within the field of visualization and visual analytics (VIS) as the overall number of scientists, researchers, and practitioners is small, while a large portion of them are busy with other sub-areas, such as applications, systems, algorithms, designs, theories, and so on.

Having Visualization Psychology as an interdisciplinary field and a branch of applied psychology can potentially attract many researchers in psychology to design and conduct experiments focused on or closely related to visualization. One important step to develop the synergy between VIS and psychology is to give an adequate emphasis on cognition in VIS research. In other words, there are needs for more studies on the "mind."

5.2.2 *More Studies on the "Mind"*

Most visualization-focused empirical studies examine hypotheses about the *artifacts* in visualization images. For example, Laidlaw et al. compared four techniques for visualizing 2D vector fields [25], Chen et al. compared four visual representations for depicting motion signatures in videos [8], and Kanjanabose et al. compared data tables, scatter plots, and parallel coordinates plots [23]. Sometimes, such studies of artifacts (e.g., techniques, plots, visual representations, systems, etc.) have led to findings about the *mind*. In their artifact-based study, Chen et al. [8] by chance discovered that participants unconsciously remembered the video visualization skills acquired in the first study and performed better 3 months later in the second study than those who did not take part in the first study. This is a finding about memory and learning—aspects of cognition. Similarly, when studying data tables, scatter plots, and parallel coordinates plots, Kanjanabose et al. [23] found that participants could retrieve data values more quickly and accurately with data tables than with scatter plots and parallel coordinate plots. Since visualization was commonly considered as a means for viewing data values and many empirical studies compare artifacts with data retrieval tasks, this raises a question: what would have happened if data tables had been involved in the comparison, or more fundamentally, in what condition visualization is better for data retrieval tasks than data tables?

In recent years, more studies were designed explicitly to study the mind, and artifacts were moved to a secondary role as stimuli for observing the mind. There have been studies on memory [5], attention [17], visual grouping [16], knowledge [24], and so on. Although artifacts were used as stimuli, the experimenters were aiming for discoveries about the mind, which can be applied to other artifacts that were not examined in the studies. For example, when Szafir found that the perception of colors was size-dependent [30], this naturally led to many hypotheses that the perception of **A** might be **Y**-dependent. Here, **A** is a placeholder for a set of artifacts being studied, while **Y** is one or a set of cognitive factors. It could also lead to a more fundamental hypothesis: must visual encoding be always isomorphic and can it be polymorphic [9] since human perception could not hold up the isomorphic requirement anyway [30]? If the latter is true, what cognitive factor (or factors) can condition polymorphic perception?

The needs for more studies on the "mind" also reflect the increasing research activities in **Visual Analytics**—a subfield of VIS—developed during the last two decades. The research in visual analytics usually focuses on integrated uses of machine-centric processes (e.g., statistics and algorithms) and human-centric processes (e.g., visualization and interaction) in workflows for data-informed decision-making and knowledge acquisition. The goals of decision-making and knowledge acquisition naturally place more emphasis on cognition than perception. From the perspective of visual analytics, there is an urgent requirement for scaling up the empirical research on the "mind" in VIS.

Meanwhile, the needs for more studies on the "mind" are warranted and entailed by the recent developments in the **Theoretical Foundation of Visualization**. One strand of these developments is an information-theoretic measure of the cost-benefit of human- and machine-centric processes in visual analytics workflows [10]. It provides many sound practices in VIS with a mathematical explanation, such as visual abstraction [32] and overview-first and detail-on-demand [9, 11]. In particular, it does not demonize information loss in VIS processes but recognizes that it is a phenomenon common in statistics, algorithms, visualization, and interaction. It also postulates that the soft knowledge of humans in the loop can alleviate the negative impact of information loss, especially in VIS processes.

To validate the postulation in [10], Kijmongkolchai et al. designed and conducted an empirical study to detect and quantify the human knowledge used in a set of trials featuring VIS processes [24]. They successfully detect the significant role of human knowledge and found a way to convert the measures of accuracy and response time to those of benefit and cost as outlined in [10]. Kijmongkolchai et al. also reported some difficulties in dealing with the unbounded nature of the Kullback–Leibler divergence, which is part of the original formula in [10]. This prompted further effort to improve the original formula by evaluating several bounded divergence measures using both conceptual analysis [12] and empirical data [7]. The empirical studies for collecting the data involved two types of VIS techniques, volume visualization and London underground maps, both of which are known to be useful but feature a significant amount of information loss. The collected empirical data were also used to exemplify two knowledge measures proposed in [7]. The example shows the importance of co-development in theoretical and empirical research in the field of VIS.

Focusing on the mind potentially allows VIS empirical research to make a big stride in making fundamental advances in the field of VIS. It is likely that studying the mind is harder than studying artifacts. However, any discovery about the mind can be translated into inferences about many artifacts. Of course, this is not to say that we should not study artifacts. Indeed, as mentioned earlier, findings about artifacts can lead to hypotheses and potentially major discoveries about the mind. Building on the past studies of artifacts and empirical researchers in VIS, hopefully, together with more and more colleagues in psychology, we will be able to conduct more studies on the mind.

5.2.3 Progressive Approaches

Studying a hypothesis about the mind is entrenched in almost every empirical study in psychology. It is also a tradition in psychology that a hypothesis is typically investigated in many empirical studies by several teams. It has been rare that a hypothesis is confirmed or disproved after the first empirical study on the hypothesis. A switching of emphasis from artifacts to the mind may instigate more progressive approaches to studying a challenging hypothesis.

Firstly, empirical researchers in visualization should embrace the tradition of psychology in scholarly contention and disagreement and should welcome any serious challenge to an existing theory or finding as long as there is an adequate empirical evidence or analytical rationale suggesting that the existing theory or finding might not be 100% correct as many thought. While it is not easy for reviewers to read papers that challenge their past theories or findings, reviewers in such situations should exercise a high level of integrity and professionalism, e.g., in making an objective assessment, declaring a conflict of interest if appropriate, and overcoming the prepossession for suppressing the debate through nitpicking.

Secondly, empirical researchers in visualization may explore other forms of empirical studies that do not involve controlled or semi-controlled experiments. The BELIV Workshop (https://beliv-workshop.github.io/) is a biennial event. Since it was established in 2006, it has been encouraging empirical researchers to develop "*new and innovative evaluation methods for visualization tools and techniques.*" While BELIV has a strong focus on artifacts, findings obtained from the evaluation of some visual representations, interaction techniques, and visualization tools can also inform the development of new hypotheses, conceptual models, and qualitative theories about the mind in the context of visualization.

Thirdly, visualization scientists are data scientists and are used to process a variety of data using data mining and data visualization. Meanwhile, empirical studies, controlled as well as uncontrolled, collect data about various variables in visualization processes, including the variables about artifacts as well as those about the mind. Often such data may not be adequate for confirming a binary hypothesis in a statistically significant manner. It may feature too many variables, or some variables may have too many values that cannot be clustered into a few groups.

Nevertheless, if the collected data feature some strong variations in the relation between the independent and dependent variables, we can discover such relations using visual analytics workflows where statistics, algorithms, visualization, and interaction are integrated. We can also use the data to develop data-driven models and data-driven metrics. Such a model or metric defines a complex causal relation in a probabilistic or functional manner, which is sometimes perceived to be less definite than a hypothesis confirmed by an empirical study. In fact, a data-driven model or a data-driven metric is just an intermediate step stone toward a grand theory.

Fourthly, there are two main types of models: *data-driven* and *theory-driven* models. A data-driven model is typically built directly on the data collected in one or a set of empirical studies. In psychology, researchers often propose theories based on the findings of empirical studies, usually including their own studies as well as those in the literature. Such theories are usually in the form of one or a series of causal relations and are treated as hypotheses to be evaluated by further empirical studies. A theory may become widely accepted after a sufficient number of experiments have evidenced the theory and no credible counter-example has been found to falsify the theory. A theory-driven model usually encodes the theorized causal relations mathematically or procedurally. We should regard developing and evaluating "hypotheses theories" as a progressive approach in studying visualization.

Considering both the third and fourth points, empirical researchers in visualization should welcome and embrace such data-driven and theory-driven models, simply because studying a hypothesis about the mind is usually much more complex than studying artifacts. Evaluating whether artifact **A** is better than artifact **B** may need one or a few empirical studies. Determining whether a function of the mind, ***X***(), causes **A** to be better than artifact **B** will likely require many intermediate steps.

5.3 The Diversity of Publications in Studying the "Mind"

The types of publications in a discipline reflect the research scopes, methods, historical predilections, and trend movements in the discipline. In this section, we first observe the journey of empirical studies in the field of visualization through the changing lens of paper types defined for IEEE VIS Conferences. The journey very much encapsulates the changes from early artifact-focused empirical research to the gradual increase of mind-focused empirical research in the field of visualization and visual analytics (VIS).

In order to anticipate the potential changes that may benefit mind-focused VIS empirical research in the future, we provide a short survey of paper types in psychology. The diversity of paper types exhibited in psychology publications reflects the publication mechanisms needed for supporting mind-focused empirical research. This suggests that empirical studies in visualization will enjoy and benefit from a broader spectrum of empirical and theoretical research in VIS.

5.3.1 The Types of Empirical Research Papers in Visualization

In the field of visualization, the study of human perception and cognition has always been encouraged. In 2003, the InfoVis program co-chairs first a categorization visualization paper types. The studies of human perception and cognition fall into the paper type "Evaluation" [21], which was defined as:

▷ [1] **Evaluation papers** are an empirical comparative study of InfoVis techniques or systems. The authors are not necessarily expected to implement the systems used in these studies themselves; the research contribution will be judged on the validity and importance of the experimental results, as opposed to the novelty of the systems or techniques under study. The conference committee appreciates the difficulty and importance of designing and performing rigorous experiments, including the definition of appropriate hypotheses, tasks, datasets, selection of subjects, measurement, validation, and conclusions. The goal of such efforts should be to move from mere description of experiments toward prediction and explanation. We suggest that the potential authors who have not had formal training in the design of experiments involving human subjects may wish to partner with a colleague from an area such as psychology or human–computer interaction who has experience with designing rigorous experimental protocols and statistical analysis of the resulting data.

Munzner detailed the motivation and rationale of the introduction of the five types of papers: *technique*, *system*, *design study*, *evaluation*, and *model* [27]. IEEE Vis in 2009 (sometimes referred to as SciVis) adopted the definitions of the five types with some minor modifications, such as changing *technique* to *algorithm/technique*, *design study* to *application/design study*, and *model* to *theory/model*. In particular, in the definition of *evaluation* papers, phrases "human users" and "empirical study" were explicitly mentioned [19]. Later, the definition was also adopted by IEEE VAST 2011, after it was transformed from a symposium to a conference in 2010.

In order to reinforce the purposes of empirical studies being not only for evaluating visual designs and visualization systems but also for studying human perception and cognition, IEEE VAST 2015 renamed the category of "Evaluation" papers as "Empirical Study" papers and offered the following definition [18]:

▷ **Empirical Study Papers.** In VAST, the goal of empirical studies is typically to gain knowledge and insight about aspects of visual analytics (VA) through direct and indirect observation and analysis of user experience. They may provide empirical evidence to support VA theories or models, compare and measure effectiveness and efficiency of a set of VA techniques (or approaches, algorithms, systems, workflows, etc.), and collect data for data-driven metrics. An accepted empirical study paper may feature one of the following qualities:

1. Novelty. An empirical study reports new discoveries and findings that have not been previously obtained. The study may examine a new phenomenon in VA or provide evidence to support or contradict an unconfirmed theoretical hypothesis or practical wisdom.
2. Innovation. An empirical study features new study methodologies that are previously unknown to or uncommon in VAST research and are technically

[1] We use a symbol ▷ to indicate that a piece of text is quoted directly from the reference cited just before the text in order to avoid overloading the meaning of quotation marks and italic fonts.

sound and beneficial in the direct and indirect observation of user experience and the collection of empirical data. Such a methodology may become a new template for empirical studies in VA.

3. Significance. An empirical study presents an experiment that may be substantially more comprehensive or lead to more meaningful statistical inference than previous studies on the topic.
4. Impact. An empirical study that may lead to a significant change of our fundamental understanding of major VA aspects or result in new guidelines and practices in VA. Such impact may have been evidentially confirmed, or an initial assessment may have convincingly suggested the potential.

Following the work of a committee called reVISe for some 2 years, IEEE VIS 2021 introduced a new area model for categorizing visualization papers [20]. The studies of human perception and cognition fall into "Area 1: Theoretical & Empirical," where the section of "Empirical" is defined as:

▷ **Empirical Research** aims to contribute research methodologies or concrete results of assessments of a visualization/visual analytics contribution or its context of use. Topic of interest include:

- *Research Methodology*: general methodologies for conducting VIS research, e.g., typology, grounded theory, empirical studies, design studies, task analysis, user engagement, qualitative and quantitative research, etc.
- *Empirical Studies*: controlled (e.g., typical laboratory experiments), semi-controlled (e.g., typical crowdsourcing studies), and uncontrolled studies (e.g., small group discussions, think aloud exercises, field observation, ethnographic studies, etc.), which may be in the forms of qualitative or quantitative research and which may be further categorized according to their objectives as follows:
- *Empirical Studies for Evaluation*: studies for assessing the effectiveness and usability of specific VIS techniques, tools, systems, and workflows, for collecting lessons learned from failures, and for establishing the best practice.
- *Empirical Studies for Observation, Data Acquisition, and Hypothesis Formulation*: studies for observing phenomena in visualization processes, stimulating hypothesis formulation, and collecting data to inform computational models and quality metrics.
- *Empirical Studies for Understanding and Theory Validation*: studies for understanding the human factors in visualization processes, including perceptual factors (e.g., visual and non-visual sensory processes, perception, attention, etc.) and cognitive factors (e.g., memory, learning, reasoning, decision-making, problem-solving, knowledge, emotion, etc.)

While the methodologies for empirical research are now formally included in the area, the conceptualization and theorization based on empirical research is also part of the other section, i.e., "Theoretical" in the same area. In particular, one particular subcategory of **Theoretical Work** is *Model Development*, which is defined as "conceptual models and simulation models for describing aspects of visualization processes (e.g., color perception, knowledge acquisition, collaborative decision making, etc.).” This is a serious encouragement for formulating qualitative and quantitative models based on empirical research. In addition, a number of research topics strongly associated with empirical research, such as quality metrics, taxonomies and ontologies, also fall into “Area 1: Theoretical & Empirical.”

The changes during the last decade from **Evaluation** papers to **Empirical Studies** papers and then to **Empirical Research** papers in “Area 1: Theoretical & Empirical” have enabled the increase of the diversity of publications on human perception and cognition in the field of visualization. However, as suggested in the short survey in the next section, the studies of the “mind” in the context of visualization have yet to be as diverse as those in psychology.

5.3.2 A Survey of Paper Types in Psychology Journals

There are many types of papers featured in different psychology journals. According to the list at Wikipedia [33], there are nearly 150 psychology journals in 2021. In this section, we present a survey of the paper types in eleven psychology journals, all of which are well-established journals and have had, or potentially will have, strong inference on visualization psychology. We use this survey to inform us about the diversity in ways of conducting research, discussing research ideas, and disseminating research results. These fourteen journals are as follows (in alphabetical order of their abbreviations):

- *The American Journal of Psychology* (AJP), University of Illinois Press
- *Annual Review of Psychology* (ARP), Annual Reviews
- *Annual Review of Vision Science* (ARVS), Annual Reviews
- *Behavioral and Brain Science* (BBS), Cambridge University Press
- *British Journal of Psychology* (BJP), The British Psychological Society
- *Cognitive Research: Principles and Implications* (CRPI), Springer
- *European Journal of Psychology of Education* (EJPE), Springer
- *Frontiers in Psychology* (FiP), Frontiers
- *International Journal of Psychology* (IJP), Wiley
- *Journal of Vision* (JoV), Association for Research in Vision and Ophthalmology
- *Perception* (Pec), Sage
- *Personality and Social Psychology Review* (PSPR), Sage
- *Psychological Reviews* (PRe), American Psychological Association
- *Theory & Psychology* (TaP), Sage

Our main selection criteria are:

1. *Ease of identifying paper types*—When a journal defines paper types explicitly in its guidance to authors and places papers in each issue under specific categories of paper types, we can precisely count the number of papers of a specific paper type. All selected journals on the above list met this criterion. When

such information is not available, one would have to categorize each paper subjectively, which is both time-consuming and error-prone.

2. *Strong focus on visual perception*—Research in the field of visualization and visual analytics (VIS) has relied extensively on perception research in psychology. These include ARVS, JoV, and Pec in the above list of selected journals. For example, Franconeri et al. [13] recently reviewed empirical research findings that support guidelines for creating effective and intuitive visualizations for disseminative visualization. The review covers many aspects of human vision and perception, e.g., ratio perception, color perception, shape perception, visual illusions, color blindness, attention, and so on.
3. *Representative of cognition research*—The review by Franconeri et al. [13] also covers several aspects of cognition, such as working memory, cognitive biases, and uncertainty and risk reasoning, suggesting the role of many other cognitive activities (in addition to sensory processing, perception, and attention) in visualization processes. As the survey by Abdul-Rahman et al. [1] summarized, visualization may feature perceptual and cognitive activities for sensing, storing, learning, thinking, motivating, feeling, externalizing, and deviating, and it will be beneficial for VIS research to draw knowledge and practices from a broader scope of psychology. We therefore intentionally selected a good number of highly reputable journals with a strong focus on cognition, which include AJP, ARP, BBS, BJP, CRPI, IJP, PRe, and TaP. While many visual representations and visualization systems are designed for data analysts and domain experts, some are designed for information and knowledge dissemination to the general public. We therefore included EJPE and PSPR, which feature empirical research topics particularly relevant to the latter.
4. *Paper types relevant but less familiar to the VIS community*—Some well-establish journals, such as BBS, FiP, and PRe in the above list, feature paper types that are highly relevant to VIS empirical research but are yet available in VIS publication venues. For example, the BBS format of "target article → open peer commentary → authors' response" enables scholarly, transparent, and democratic discourse in research disciplines, where empirical evidence can sometimes be conflicting with each other or can often lead to different interpretations and conclusions. The mission of PRe, which includes "systematic evaluation of alternative theories," encourages the scholarly review and analysis of competing theories and the findings resulting from empirical research. The diverse range of paper types in FiP complements the common paper type "research article" and encourages exploratory theoretical, conceptual, and methodological research, as well as practical applications, system development, and technical innovation.

Collectively, a wide range of paper types were featured in these journals. Table 5.1 summarizes the occurrences of different category names, in alphabetic order, in these journals. Each value indicates the number of occurrences of a category name (row) in a journal (column) in the period between January 2010 and December 2020.

Table 5.1 Category names of research papers in eleven psychology journals. Each number indicates the number of papers or small writings that were labeled with a category name (row) in a journal (column) during the period between January 2010 and December 2020. The only exception is ARVS, which published its first volume in 2015

Category names	AJP	ARP	ARVS	BBS	BJP	CRPI	EJPE	FiP	IJP	JoV	Pec	PSPR	PRe	TaP
Anniversary article	24	0	0	0	0	0	0	0	0	0	0	0	0	0
Article	306	0	0	0	3	0	0	0	36	2445	834	147	387	435
Authors' response	0	0	0	95	10	0	0	0	0	9	0	0	0	0
Book review	205	0	0	0	141	0	0	93	0	0	76	0	0	19
Brief research report	0	0	0	0	0	14	0	453	89	0	0	0	0	0
Clinical case study	0	0	0	0	0	0	0	53	0	0	0	0	0	0
Clinical study protocol	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Clinical trial	0	0	0	0	0	0	0	50	0	0	0	0	0	0
Commentary	0	0	0	0	36	0	0	0	2	0	0	0	36	45
Community case study	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Conceptual analysis	0	0	0	0	0	0	0	216	0	0	0	0	0	0
Debate	0	0	0	0	0	0	0	0	3	0	0	0	0	0
Data report	0	0	0	0	0	0	0	29	0	0	0	0	0	0
Editor's choice	0	0	0	0	4	0	0	0	0	0	0	0	0	0
Empirical study	0	0	0	0	0	0	0	6	0	0	0	0	0	0
Essay review	0	0	0	0	0	0	0	0	0	0	0	0	0	6
Evaluation	0	0	0	0	0	0	0	4	0	0	0	0	0	0
Focused review	0	0	0	0	0	0	0	37	0	0	0	0	0	0
Forum	2	0	0	0	0	0	0	0	0	0	0	0	0	0
Frontiers commentary	0	0	0	0	0	0	0	20	0	0	0	0	0	0
General commentary	0	0	0	0	0	0	0	484	0	0	0	0	0	0
Invited article	0	0	0	0	1	0	0	0	0	0	0	0	0	0
History of psychology	19	0	0	0	0	0	0	0	0	0	0	0	0	0

Hypothesis and theory	0	0	0	0	0	0	0	875	0	0	0	0	0	0
Letter	0	0	0	0	0	0	0	0	0	17	6	0	0	0
Methods	0	0	0	0	0	0	0	265	0	62	0	0	0	0
Mini review	0	0	0	0	0	0	0	350	0	0	0	0	0	0
Open peer commentary	0	0	0	2594	0	0	0	0	0	0	0	0	0	29
Opinion	0	0	0	0	0	0	0	809	0	0	0	0	0	0
Original research	0	0	0	0	313	202	436	13,512	376	0	0	0	0	0
Perspective	0	0	0	0	0	0	0	485	0	3	0	0	0	0
Policy and practice review	0	0	0	0	0	0	0	3	0	0	0	0	0	0
Protocol	0	0	0	0	0	0	0	33	0	0	0	0	0	0
Regular article	0	0	0	0	89	0	0	0	0	0	0	0	0	0
Reports	0	0	0	0	0	0	1	0	0	0	98	0	0	0
Research articles	0	0	0	0	0	0	0	0	228	370	0	44	0	0
Response	0	0	0	0	1	0	0	0	0	9	0	0	0	19
Reviews	0	292	130	0	0	18	5	1031	13	28	56	0	0	59
Technology report	0	0	0	0	0	0	0	28	0	0	0	0	0	0
Special issue paper	0	0	0	0	29	0	0	0	7	0	42	0	0	0
Special section paper	0	0	0	0	0	0	0	0	67	0	0	0	0	0
Specialty grand challenge	0	0	0	0	0	0	0	27	0	0	0	0	0	0
Study protocol	0	0	0	0	0	0	0	7	0	0	0	0	0	0
Systematic review	0	0	0	0	0	0	0	229	0	0	0	0	0	0
Target article	0	0	0	92	2	0	0	0	0	0	0	0	0	1

(continued)

Table 5.1 (continued)

Category names	AJP	ARP	ARVS	BBS	BJP	CRPI	EJPE	FiP	IJP	JoV	Pec	PSPR	PRe	TaP
Technology and code	0	0	0	0	0	0	0	5	0	0	0	0	0	0
Theoretical notes	0	0	0	0	0	0	0	0	0	0	0	0	27	0
Tutorial review	0	0	0	0	0	3	0	0	0	0	0	0	0	0
Total	556	292	130	2781	629	237	442	19,108	821	2943	112	191	450	613

In addition, there are various small writings for communicating additional, supplementary, professional, or organizational information, such as the Addendum, Award, Call for papers, Correction, Corrigendum, Editorial, Editorial Acknowledgment, Erratum, Introduction, In memoriam, Obituary, Precis, Preface, and Retraction.

5.3.3 A High-level Categorization

Figure 5.1 shows two word clouds of the words in the category names listed in the leftmost column of Table 5.1. All plural words are converted to their singular forms except "notes." All words are case-insensitive, while all function words are excluded. Under these processing conditions, there are a total of 58 words in each word cloud. The color and font size of each word encode the frequency of the word. When a word appears in multiple category names, the frequency of the word is the sum of the frequencies of these category names.

In Fig. 5.1a, we compute the frequency of a category name by counting the number of journals that feature such a category name. For example, according to Table 5.1, five journals have the category "book review," and its frequency is thus 5. In Fig. 5.1b, we compute the frequency of a category name by counting the number of papers and small writings under such a category in all journals that feature the

(a) based on journal occurrences (b) based on paper occurrences

Fig. 5.1 Two word clouds of the words in the category names column of Table 5.1. In (**a**), the frequency of each category name is computed based on counting the number of journals that have this category name. In (**b**), the frequency of each category name is computed by counting the number of papers or small writings labeled with this category name. When a word appears in multiple category names, the frequency of the word is the sum of the frequencies of these category names. The word clouds were created using WordItOut (https://worditout.com/). (**a**) Based on journal occurrences. (**b**) Based on paper occurrences category. Thus, the frequency of “book review” is 534 (205+141+93+76+19). The word “book” appears only once among all category names, and thus its frequency is 5 in (a) and 534 in (b). On the other hand, the word “review” appears in seven category names, and its frequency is the sum of the frequencies of these seven category names. It is 20 in (a) and 2784 in (b).

From Fig. 5.1a, we can observe that words “review” (20 occurrences) and “article” (17) occur most frequently. Other words that may draw our attention include “research” (11), “commentary” (8), “report” (7), “response” (6), “book” (5), “original” (5), and “study” (5). These more frequently occurred words represent a diverse range of papers. In addition, two words have 4 occurrences, six words have 3 occurrences, seven words have two occurrences, and thirty-five words have 1 occurrence.

From Fig. 5.1b, we can observe that the words “research” (16,037 occurrences) and “original” (14,839) occur most frequently. The next group of words includes “article” (5444), “commentary” (3246), “review” (2794), “open” (2623), and “peer” (2623). In addition, eighteen words have occurrences in the 3-digit range of [100, 999], and twenty-one words in the 2-digit range of [10, 99], and thirteen words in the single-digit range of [1, 9].

From Table 5.1, We can also observe that the word “book” only occurs in the category of “Book review,” while the word “review” is featured in eight category names. The words “open” and “peer” only occur in the category of “Open peer commentary,” while the word “commentary” is featured in four category names. The word “original” only occurs in the category of “Original research,” while the word “research” is featured in three different category names.

Therefore, we can consider that the category features “book,” “open,” “peer,” or “original” in its name may likely be part of a super-category. After excluding these words for further consideration, the words with high-frequent occurrences in Fig. 5.1 are:

article, **commentary**, **report**, **research**, **response**, **review**, **study**

It is not so difficult to see a high-level categorization emerging. In this chapter, we broadly divide papers and small writings in these eleven journals into the following five super-categories.

- **Articles**—These are papers considered as the *most typical* papers in the journal concerned. They typically feature original research and are presented in a regular, full-length format. We intentionally do not use any word to modify “articles” because some commonly used noun adjuncts or adjectives may stimulate narrow-minded interpretations. For example,
  - Using the word “research” as a noun adjunct might imply other types of papers are not research papers.
  - Using the adjective “original” might suggest that papers in other super-categories do not contain much original research and would not rather unfair to many papers in the super-category of **Reports**.

  – Using the adjective “regular” might exclude papers that are regular in content and format, but less regular in submission, review, and editorial processes, e.g., special issue and special section papers.

- **Commentaries and Responses—Commentaries** are normally small writings that offer comments and opinions on a particular topic or a published paper, while **Responses** are the authors’ feedback to the commentaries on specific papers. In terms of the categories listed in Table 5.1, this super-category includes all category names featuring words such as “commentary,” “response,” “opinion,” “debate,” and “forum.”
- **Reviews**—From Table 5.1, we can observe several types of reviews. This super-category simply includes all category names featuring the word “review.”
- **Reports**—These are relatively short papers that typically offer brief communication about a research project or a technical aspect. Journals that cater for papers in this super-category normally label them with category names featuring words such as “report” and “study,” while avoiding words “article” and “research” in order to differentiate them from full-length papers in the super-category of **Articles**. However, it is more fair-minded to consider most, if not all, such short papers as *research* work. We decided not to include the word “studies” in naming this super-category because many full-length papers in psychology report empirical studies. It would not be appropriate to associate the word “study” with only short papers.
- **Others**—This super-category includes all other types of small writings that do not fall into the super-categories of **Commentaries and Responses**, **Reviews**, **Reports**. They are typically for communicating additional, supplementary, professional, or organizational information, such as Addendum, Award, Call for papers, Correction, Corrigendum, Editorial, Editorial Acknowledgement, Erratum, Introduction, In memoriam, Obituary, Precis, Preface, and Retraction.

5.3.4 Further Categorization of “Articles”

According to Table 5.1, in six journals (out of eleven), most of their papers fall into the category “Article,” i.e., AJP 55%, JoV 83%, Pec 75%, PSPR 77%, PRe 86%, and TaP 71%. This is indeed another reason for the super-category **Articles** to adopt the term “article.” Meanwhile, in five other journals, the most common category name is “Original Research,” i.e., BJP 50%, CRPI 85%, EJPE 99%, FiP 71%, and IJP 46%.

Publisher Springer, which hosts the journal EJPE, defines the “Original Research” papers as follows [29]:

▷ This is the most common type of journal manuscript used to publish full reports of data from research. It may be called an *Original Article*, *Research Article*, *Research*, or just *Article*, depending on the journal. The Original Research format is suitable for many different fields and different types of studies. It includes full Introduction, Methods, Results, and Discussion sections.

Most journals in psychology place a strong emphasis on empirical research. For example, the American Psychological Association (ASA), which hosts PRe, defines "Research articles" as [3]:

▷ Behavior analysis deals with relations between environmental inputs and behavioral outputs using a behavior analytic conceptual framework. Research articles present original empirical findings depicting these relations.
Such articles must provide a compelling rationale for the experimental question, employ methods that are appropriate for answering that question, include sufficient detail about those methods to allow for replication, present meaningful data, analyze those data appropriately, and interpret them meaningfully.

Some other journals may focus on more theoretical and methodological research. For example, TaP defines its scope as [31]:

▷ Theory & Psychology publishes scholarly and expository papers which explore significant theoretical developments within and across such specific sub-areas as cognitive, social, personality, developmental, clinical, perceptual, or biological psychology.

Many journals are platforms for supporting a diverse range of research work, including empirical, theoretical, and methodological research. One such journal is FiP, which makes the diversity particularly explicit through its paper types [14]. FiP offers nine types of full-length papers, each of which may have up to 12,000 words. We have included six categories, namely "Original research," "Hypothesis and theory," "Clinical trial," "Method," "Study protocol," and "Technology and code," in the super-category **Articles**, and three other categories of "Review," "Policy and practice review," and "Systematic review" in the super-category **Reviews** (see also Sect. 5.3.4).

In addition, there are categories indicating whether a paper is submitted and reviewed in a *regular* process of the journal or a special process, such as in relation to a special call for papers. These special processes include "Anniversary article," "Editor's choice," "Invited article," "Special issue paper," and "Special section paper."

Among the eleven journals in Table 5.1, only BBS has fewer full-length articles (3.3%) than its extensive collection of papers in the "Open peer commentary" category (93.3%). This apparent anomaly is because of the unique format of BBS, which refers a full-length article as a "Target article" [4]:

▷ A BBS target article can be (i) the report and discussion of empirical research that the author judges to have broader scope and implications than might be more appropriately reported in a specialty journal, (ii) an unusually significant theoretical article that formally models or systematizes a body of research, or (iii) a novel interpretation, synthesis, or critique of existing experimental or theoretical work. Occasionally, articles dealing with social or philosophical aspects of the behavioral and brain sciences will be considered.

In BBS, each target article is a significant and controversial piece of work and is published in the same issue together with 20–40 commentaries from specialists within and across the discipline concerned and the author's response to the commentaries. The level of openness and rigor in scholarly discourse is inspirational to visualization researchers. We have included the category of "Target article" in the super-category **Articles**, while placing "Open peer commentary" and "Author's response" in the super-category **Commentaries and Responses**.

The super-category **Articles** includes the following categories listed in Table 5.1:

- Anniversary article (AJP)
- Article (AJP, BJP, IJP, JoV, Pec, PSPR, PRe, TaP)
- Editor's choice (BJP)
- Clinical trial (FiP)
- Invited article (BJP)
- Hypothesis and theory (FiP)
- Method (FiP, JoV)
- Original research (BJP, CRPI, EJPE, FiP, IJP)
- Regular article (BJP)
- Research article (IJP, JoV, PSPR)
- Special issue paper (BJP, IJP, Pec)
- Special section paper (IJP)
- Study protocol (FiP)
- Target article (BBS, BJP, TaP)
- Technology and code (FiP)

5.3.5 Further Categorization of "Commentaries and Responses"

Papers in this super-category are common in the field psychology, encapsulating a research culture that embracing openness in discussion, discourse, and debate. Difference or disagreement in theoretical understanding and postulation is not a barrier in accepting a paper that may contain important postulation, which some reviewers do not agree with. Instead, such difference or disagreement is "welcomed" and facilitates more papers commonly referred to as **Commentaries** and **Responses**.

FiP defines its categories "General Commentary" and "Opinion" as [14]:

▷ General Commentary articles provide critical comments on a previous publication at Frontiers. The authors wishing to submit commentaries on articles published outside of Frontiers are encouraged to reformat and submit them as an Opinion type.

▷ Opinion articles allow the authors to contribute viewpoints on the interpretation of recent findings in any research area, value of the methods used, as well as weaknesses and strengths of scientific hypotheses.

To complement such commentary and opinion papers, the “Response” category provides the authors of articles that receive open critical commentaries with an opportunity to respond openly and formally through a short paper.

Behavioral and Brain Science (BBS) provides a unique and ample platform where the authors and commentators engage in open, extensive, and constructive interaction on a topic judged to be of broad significance.

These papers are usually much shorter than the full-length papers. For example, FiP limits the length of a paper in the “Opinion” category to 2000 words and that in the category of “General commentary” to 1000 words. JoV offers a paper type “Perspectives,” which present authors’ personal viewpoints on topics and limits the length of each paper in this category to 4 pages. JoV also offers a paper type “Point/CounterPoint” that presents two invited articles with opposing views and limits the length of each paper to 2–3 pages [22].

Some journals use the category name “Letter” for a broad range of small writings including commentaries. The American Psychological Association (ASA), which hosts PRe, defines letters as [3]:

▷ Letters, which comprise no more than 850 words, provide a means for behavior analysts to share potentially important information that would not be appropriate for publication in another format. Examples include commentaries on books or articles, descriptions of interesting research findings or other observations that merit further investigation, and reports of political or legal events likely to affect the field.

The *Journal of Vision* defines letters as [22]

▷ The journal welcomes submission of Letters to the Editor to be considered for publication. Letters may concern material published in the journal or issues of general interest to vision scientists. Letters about material published in the journal may correct errors or offer different points of view, clarification, or additional information or analyses in a civil manner. Letters will be evaluated for their scientific merit, technical quality, and significance. Letters whose arguments or conclusions require support from experimental evidence or theoretical analyses are more appropriate as regular submissions and may be declined without review. The authors whose article is discussed in a Letter will be given an opportunity to reply.

Our survey of the eleven journals has found the following categories under the super-category **Commentaries and Responses**:

- Authors’ response (BBS, BJP, JoV)
- Commentary (BJP, IJP, PRe, TaP)

- Debate (IJP)
- Forum (AJP)
- Frontiers commentary (FiP)
- General commentary (FiP)
- Letter (JoV, PRe)
- Open peer commentary (BBS)
- Opinion (FiP)
- Perspective (FiP, JoV)
- *Point/CounterPoint* (JoV)
- Response (BJP, JoV, TaP)
- Theoretical notes (PRe)

where the category "Point/CounterPoint" was shown in italics because it is a paper type offered by JoV, but our survey has not found any paper with this category name in JoV between January 2010 and December 2020.

5.3.6 Further Categorization of "Reviews"

Papers in the super-category of **Reviews** feature extensive discussions on previously published research. Among the eleven journals in Table 5.1, nine journals (ARP, ARVS, EJPE, FiP, IJP, JoV, Pec, PRe, and TaP) offer a generic category "Review."

Publisher Springer, which hosts the journal EJPE, defines the most common type of review papers as follows [29]:

▷ Review articles provide a comprehensive summary of research on a certain topic and a perspective on the state of the field and where it is heading. They are often written by leaders in a particular discipline after invitation from the editors of a journal. Reviews are often widely read (for example, by researchers looking for a full introduction to a field) and highly cited. Reviews commonly cite approximately 100 primary research articles.

Publisher Annual Reviews hosts 52 journals that focus almost entirely on review papers. While the collection covers a broad spectrum of academic disciplines, it includes four psychology journals (cf. one for computer science). In addition to ARP and ARVS included in our survey, there are also *clinical psychology* and *developmental psychology*.

As mentioned earlier in Sect. 5.3.4 in conjunction with the category of "Target articles" in BBS, literature review and theoretical discourse and development often go hand in hand in the psychology literature. American Psychological Association (ASA), which hosts PRe, defines papers in the "Review category" as *review and conceptual articles* and offers the following definition [3]:

▷ Articles in this category summarize previously published research or address theoretical or conceptual issues of interest to behavior analysts. Such articles support conclusions of potential theoretical, clinical, or practical importance to behavior analysts and are written in a clear and comprehensible style.

The other four journals (AJP, BBS, BJP, and PSPR) also publish review articles without an explicit category named as "Review" or alike. Some journals offer specific categories of review papers, and these category names are rather self-explanatory. Together with the generic category of "Review," the eleven journals in Table 5.1 offer the following categories of review papers (in alphabetic order):

- Book review (AJP, BJP, FiP, Pec, TaP)
- *Emerging trends in vision science* (JoV)
- Essay review (TaP)
- Focused review (FiP)
- History of psychology (AJP)
- Mini review (FiP)
- Policy and practice review (FiP)
- Review (ARP, ARVS, CRPI, EJPE, FiP, IJP, JoV, Pec, PRe, TaP)
- Systematic review (FiP)
- Tutorial review (CRPI)

where "Emerging trends in vision science" is a paper type offered by JoV, though our survey has not found any paper of this type during the period between January 2010 and December 2020.

5.3.7 Further Categorization of "Reports"

This super-category consists of relatively shorter research papers in comparison with the full-length research papers. The definition of short- vs full-length is sensitive to the context of individual journals. These papers are often referred to as *research reports* or *case studies*.

The American Psychological Association (ASA), which hosts PRe, defines research reports as [3]:

▷ Research reports are similar to research articles, but no more than 2500 words in length, with no more than two tables or figures.
Research reports are a convenient venue for reporting findings that are suggestive but not compelling, technological devices or applications, follow-up data not adequate to support a research article, or any study that can be accurately described in few words.

Publisher Springer, which hosts the journal EJPE, defines short reports and letters collectively as [29]:

▷ These papers communicate brief reports of data from original research that editors believe will be interesting to many researchers and that will likely stimulate further research in the field. As they are relatively short, the format is useful for scientists with results that are time sensitive (for example, those in highly competitive or quickly changing disciplines). This format often has strict length limits, so some experimental details may not be published until the authors write a full Original Research manuscript. These papers are also sometimes called Brief communications.

Meanwhile, Springer defines case studies as [29]:

▷ These articles report specific instances of interesting phenomena. A goal of Case Studies is to make other researchers aware of the possibility that a specific phenomenon might occur. This type of study is often used in medicine to report the occurrence of previously unknown or emerging pathologies.

ASA uses the term "Case conference" and defines it as [3]:

▷ The goal of this type of submission is to provide professionals and students with an opportunity to acquire skills in behavioral case conceptualization, behavioral assessment methods, and intervention in the context of service delivery (e.g., Behavior Therapy, Clinical Behavior Analysis, Behavioral Medicine, and Applied Behavior Analysis).
A secondary goal is the dissemination of behavior analytic assessment and intervention methods to the broader community of readers. To achieve these goals, manuscripts describing the use of a controlled case study (A-B design) or single-subject research design allowing demonstration of functional relationships are appropriate.

FiP is a publication venue accepting a variety of short papers in addition to full papers (maximum 12,000 words). These short papers include [14]:

- Maximum 8000 words: "Conceptual analysis"
- Maximum 5000 words: "Community case study," "Curriculum, instruction, and pedagogy"
- Maximum 4000 words: "Brief research report"
- Maximum 3000 words: "Case report," "Data report," "Mini review," "Perspective," "Policy brief," "Registered report"
- Maximum 2000 words: "Specialty grand challenge"

In addition, FiP has published papers under category names "Clinical case study," "Empirical study," "Evaluation," "Protocol," and "Technology report." Meanwhile, although categories "Case report," "Curriculum, instruction, and pedagogy," "Policy brief," and "Registered report" are listed as FiP paper types [14], and no FiP paper in our survey period appears to be labeled with any of these category names.

In particular, "Registered report" is referred to as a *Stage 1* paper, *outlining a proposed methodology and analysis which is pre-registered before data collection* [14]. *Following the In-Principle Acceptance authors have 1 year to collect data and* *submit a complete manuscript for Stage 2 of peer review* [14]. "Registered report" is also a special paper type offered by BJP, which defines it as [6]:

▷ Registered Reports are a form of empirical article in which the methods and proposed analyses are pre-registered and reviewed prior to research being conducted. This format is designed to minimize bias in deductive science, while also allowing complete flexibility to conduct exploratory (unregistered) analyses and report serendipitous findings.

The following list shows the category names under the super-category **Reports**, including those listed in Table 5.1 as well as those paper types defined by the eleven journals. Those defined by the journals but not in the table are listed in italics.

- Brief research report (CRPI, FiP, IJP)
- *Case conference* (PRe)
- *Case report* (FiP)
- Clinical case Study (FiP)
- Clinical study Protocol (FiP)
- Community case Study (FiP)
- Conceptual analysis (FiP)
- *Curriculum, instruction, and pedagogy* (FiP)
- Data report (FiP)
- Empirical study (FiP)
- Evaluation (FiP)
- *Policy brief* (FiP)
- Protocol (FiP)
- Report (EJPE, Pec)
- *Registered report* (BJP, FiP)
- Specialty grand challenge (FiP)
- *Technology* (PRe)
- Technology report (FiP)

5.3.8 Further Categorization of "Others"

In addition to the category names listed in Table 5.1, we can find other categories in some journals. The names of these categories are mostly self-explanatory. Almost all of them are associated with small writings for communicating additional, supplementary, professional, or organizational information. They are listed below in alphabetic order:

- Addendum
- Awards
- Call for papers
- Correction (FiP)
- Corrigendum

- Editorial (FiP)
- Editorial Acknowledgement
- Erratum (FiP)
- Introduction
- In Memoriam—a Latin term, meaning “in memory of”
- Obituary
- Precis
- Preface
- Retraction

5.3.9 Observations and Discussions

From the above survey, we can make a number of observations:

1. **Scale and diversity**—In terms of the number of papers and the diversity of paper types, the current effort of studying the mind in the field visualization is a drop in the ocean. While it is necessary to increase the effort within the field of visualization, it will be more effective to encourage researchers in psychology to investigate research questions about visualization and utilize the well-established research expertise and publication platforms in the discipline of psychology.
2. **Theory development and evaluation**—Research in psychology is not limited to empirical studies. The transformation from studying artifacts to studying the mind will require visualization researchers to be more interested in theory development and evaluation. While we can be built on a large and diverse collection of theories in psychology literature, many of us may have to become accustomed to the notion that a theory is a hypothesis. It is necessary to propose theories for abstracting and explaining findings of empirical research, while empirical research has an important role in testing and questioning existing theories as well as suggesting new theories.
3. **Scholarly discussion, discourse, and debate**—It is highly desirable to introduce new papers in the field of visualization to encourage open and rigorous discussion, discourse, and debate as exemplified by the target articles and open peer commentaries in BBS.
4. **Short papers and reports**—The field of visualization has conference-based platforms for publishing short papers, but no journal-based platform. It is desirable to develop journal-based platforms to encourage and sustain a diverse range of practical, analytical, and technical research effort that may not result in full-length research papers.

5.4 Conclusions

In data science, interactive visualization and visual analytics brings together machine-centric processes and human-centric processes. It can provide psychologists with one of the best platforms for studying the human mind. Therefore, creating a new interdisciplinary area of visualization psychology will not only benefit the research and development in the field of visualization but also benefit the scientific agenda in psychology. In particular, the aforementioned fundamental questions in visualization are also fundamental questions in perception and cognition. Many currently imperfect guidelines in visualization reflect some limited understanding in terms of perception and cognition. Failures or shortcomings in the human mind often inspire some best research topics in psychology. Similarly, failures or shortcomings of visualization guidelines could inspire some best research topics in visualization psychology.

Meanwhile, many visualization scientists and researchers are highly skilled in data analysis and have access to many practical applications. Visualization psychology can benefit from such skills and applications in developing new research methodologies and delivering high impact applications.

Having more studies on the mind and having more progressive approaches naturally lead to an update of the existing evaluation criteria for artifact-focused empirical study papers. An accepted empirical study paper in visualization psychology may feature one of the following qualities:

- **Novelty.** An empirical study reports new discoveries and findings that have not been previously obtained. The study may examine a new phenomenon in visualization or provide evidence to support or contradict an unsupported theoretical hypothesis or practical wisdom.
- **Innovation.** An empirical study features new study methodologies that are previously unknown to or uncommon in visualization research and are technically sound and beneficial in the direct and indirect observation of user experience and the collection of empirical data. Such a methodology may become a new template for empirical studies in visualization.
- **Significance.** An empirical study presents an experiment that is substantially more comprehensive, or leads to more meaningful statistical inference, than previous studies on the topic.
- **Impact.** An empirical study that may lead to a significant change of our fundamental understanding about visualization or result in new guidelines and practices in visualization. Such an impact may have been evidently confirmed, or an initial assessment may have convincingly suggested the potential.
- **Data, Evidence, Measurement, and Analysis.** An empirical study reports important data samples, evidence, measurement, and analysis that have not been previously obtained. The study may contribute toward the discoveries and findings of a major, fundamental, and complex hypothesis that is difficult to confirm or disapprove through one or a few empirical studies.

In addition, we need to develop new threads of research work and scholarly publications beyond empirical studies, which may include but not limited to "Hypothesis and theory," "Modelling and simulation," "Method," "Study protocol," "Technology and code," "Open peer commentaries," "Author's response," "Brief Research Report," "Case study," "Data report," and so on.

Acknowledgments The authors would like to thank Professor Sine McDougall, Bournemouth University, for providing us with helpful suggestions. The authors also appreciate very much the comments and suggestions made by the anonymous reviewers.

References

1. A. Abdul-Rahman, R. Borgo, M. Chen, D. J. Edwards, and B. Fisher. Juxtaposing controlled empirical studies in visualization with topic developments in psychology. arXiv 1909.03786, 2019.
2. A. Abdul-Rahman, M. Chen, and D. H. Laidlaw. A survey of variables used in empirical studies for visualization. In *Foundations of Data Visualization*. Springer, 2020.
3. American Psychological Association. Types of articles accepted. https://www.apa.org/pubs/journals/bar/article-types, accessed in April 2021.
4. Behavioral and Brain Sciences. Instructions for authors. https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/information/instructions-contributors, accessed in July 2021.
5. I. Boyandin, E. Bertini, and D. Lalanne. A qualitative study on the exploration of temporal changes in flow maps with animation and small-multiples. *Computer Graphics Forum*, 31(3pt2):1005–1014, 2012.
6. British Journal of Psychology. Guidelines for authors. https://bpspsychub.onlinelibrary.wiley.com/hub/journal/20448295/homepage/registeredreportsguidelines.htm, accessed in July 2021.
7. M. Chen, A. Abdul-Rahman, D. Silver, and M. Sbert. A bounded measure for estimating the benefit of visualization (part ii): case studies and empirical evaluation. *Entropy*, 24(2):282, 2022.
8. M. Chen, R. P. Botchen, R. R. Hashim, D. Weiskopf, T. Ertl, and I. M. Thornton. Visual signatures in video visualization. *IEEE Trans. Visualization & Computer Graphics*, 12(5):1093–1100, 2006.
9. M. Chen and D. J. Edwards. "Isms" in visualization. In *Foundations of Data Visualization*. Springer, 2020.
10. M. Chen and A. Golan. What may visualization processes optimize? *IEEE Transactions on Visualization and Computer Graphics*, 22(12):2619–2632, 2016.
11. M. Chen and H. Jänicke. An information-theoretic framework for visualization. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1206–1215, 2010.
12. M. Chen and M. Sbert. A bounded measure for estimating the benefit of visualization (part i): theoretical discourse and conceptual evaluation. *Entropy*, 24(2):228, 2022.
13. S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman. The science of visual data communication: What works. *Psychological Science in the Public Interest*, 22(3):110–161, 2021.
14. Frontiers in Psychology. Article types. https://www.frontiersin.org/journals/psychology#article-types, accessed in July 2021.
15. J. Fuchs, P. Isenberg, A. Bezerianos, and D. Keim. A systematic review of experimental studies on data glyphs. *IEEE Trans. Visualization & Computer Graphics*, 23(7):1863–1879, July 2017.

16. C. C. Gramazio, K. B. Schloss, and D. H. Laidlaw. The relation between visualization size, grouping, and user performance. *IEEE Trans. Visualization & Computer Graphics*, 20(12):1953–1962, 2014.
17. S. Haroz and D. Whitney. How capacity limits of attention influence information visualization effectiveness. *IEEE Trans. Visualization & Computer Graphics*, 18(12):2402–2410, 2012.
18. IEEE VAST 2015. Vast paper types. http://ieeevis.org/year/2015/info/call-participation/vast-paper-types, 2015.
19. IEEE Vis 2009. Call for participation: Vis papers. http://vis.computer.org/VisWeek2009/vis/cfp_papers.html, 2009.
20. IEEE VIS 2021. The vis area model for 2021+. http://ieeevis.org/governance/area-model#area-1-theoretical--empirical, 2021.
21. InfoVis 2003. Call for papers. https://infovis.org/infovis/2003/CFP/#papers, 2003.
22. Journal of Vision. Submitting manuscripts for peer review. https://jov.arvojournals.org/SS/PeerReview.aspx, accessed in July 2021.
23. R. Kanjanabose, A. Abdul-Rahman, and M. Chen. A multi-task comparative study on scatter plots and parallel coordinates plots. *Computer Graphics Forum*, 34(3):261–270, 2015.
24. N. Kijmongkolchai, A. Abdul-Rahman, and M. Chen. Empirically measuring soft knowledge in visualization. *Computer Graphics Forum*, 36(3):73–85, 2017.
25. D. H. Laidlaw, J. S. Davidson, T. S. Miller, M. da Silva, R. M. Kirby, W. H. Warren, and M. Tarr. Quantitative comparative evaluation of 2D vector field visualization methods. In *Proc. IEEE Visualization*, pages 143–150, 2001.
26. H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Trans. Visualization & Computer Graphics*, 18(9):1520–1536, Sept 2012.
27. T. Munzner. Process and pitfalls in writing information visualization research papers. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization*, volume LNCS 4950. Springer, 2008.
28. R. E. Roth, A. Çöltekin, L. Delazari, H. F. Filho, A. Griffin, A. Hall, J. Korpi, I. Lokka, A. Mendonça, K. Ooms, and C. P. J. van Elzakker. User studies in cartography: Opportunities for empirical research on interactive maps and visualizations. *International Journal of Cartography*, 3(sup1):61–89, 2017.
29. Springer. Types of journal articles. https://www.springer.com/gp/authors-editors/authorandreviewertutorials/writing-a-journal-manuscript/types-of-journal-articles/10285504, accessed in April 2021.
30. D. A. Szafir. Modeling color difference for visualization design. *IEEE Trans. Visualization & Computer Graphics*, 24(1):392–401, Jan 2018.
31. Theory & Psychology. Aims and scope. https://journals.sagepub.com/aims-scope/TAP, accessed in July 2021.
32. I. Viola, M. Chen, and T. Isenberg. Visual abstraction. In *Foundations of Data Visualization*. Springer, 2020.
33. Wikipedia. List of psychology journals. https://en.wikipedia.org/wiki/List_of_psychology_journals, accessed in April 2021.

Part II
Visualization Psychology from a Visualization Perspective

In developing a book on a new subject discipline called visualization psychology (VisPsych), different approaches and perspectives should be drawn from both psychology and visualization (a subject residing mainly in computer science) in order to present a balanced perspective of this new discipline. Visualization research has developed a conscious awareness of the two domains naturally and the almost-inevitable intersection between them.

This part explores studies that are drawn largely from the computer science domain as opposed to psychology, acting as a complement to the previous part that draws largely from the psychology domain. The chapters in this part attempt at making foundational links between visualization and psychology, outlining and proving, from both theoretical and empirical points of view, the implicit synergy of the two disciplines. The chapters encourage readers to develop more formal and better structured collaborations across their disciplinary boundaries, fostering a virtuous cycle of mutual benefit, which one needs for the progress and further development of both.

In the first chapter of this part, Chap. 6 “Visualization Onboarding Grounded in Education Theories,” Christina Stoiber and colleagues present a survey of approaches from the academic community as well as from commercial products relating to how to support users in learning how to use new digital technologies. They emphasize the approach of onboarding, define the concept, and then systematically lay out the design space of onboarding in the context of visualization and as a conceptual framework using learning theories.

In the second chapter of this part, Chap. 7 “Adaptive Visualization of Health Information Based on Cognitive Psychology: Scenerios, Concepts and Research Opportunities,” Tobias Schreck and colleagues discuss how evidence-based medical knowledge, cognitive mechanisms, and novel interactive data visualizations can potentially be combined to form adaptive and interactive consumer health information systems that take into account individual health information needs such as health literacy.

In the third chapter of this part, Chap. 8 “Design Cognition in Data Visualization,” Paul Parsons introduces the field of design cognition and its relevance to visualization (VisPsych). He highlights two relevant paradigms—the rational solving problem and the reflective practice paradigm. Paul then outlines the strengths and weaknesses of these in order to reconcile their differences and then examines these implications in relation to four data visualization topics (defining, automating, modeling, and teaching data visualization design).

In the fourth chapter of this part, Chap. 9 “Visualization Psychology: Foundations for an Interdisciplinary Research program,” Amy Rae Fox and James D. Hollan introduce the first ever Visualization Psychology framework. Unique to the chapter is the interpretation of the framework as a set of theoretical premises that should guide any inquiry concerned with psychological intrinsic dimensions of visualization. The chapter offers a unique view on how visualization is a fertile laboratory where theories of perception and cognition can thrive and advance. The authors provide a strong argument as to why the intersection between visualization and psychology is not a new one but rather traces its roots back into the origins of human–computer interaction. They suggest that visualization should be situated in the much broader context of external representation, semiotic activity, information processing, and distributed cognitive systems, rather than being relegated within the computer science realm.

In the fifth chapter of this part, Chap. 10 “Visualization Psychology for Eye Tracking Evaluation,” Maurice Koch and colleagues provide further empirical evidence of the implicit synergy between Visualization and Psychology already highlighted in the previous chapters. The authors’ focus point is this temporal hardware, i.e., eye-tracking devices in particular. The chapter provides empirical evidences showing the advantages of employing cognitive models when evaluation of visualizations is performed through the means of eye-tracking devices. Eye-tracking technology enables visualization research to deepen its understanding of the perceptual and cognitive processes at play when interpreting a visualization. Meanwhile, theories and methodologies from psychology and cognitive science can benefit the design and evaluation of eye-tracking experiments for visualization.

The chapters in this part draw from visualization theory and practice to provide strong arguments in favor of the intimate and intertwined relation between visualization and psychology. Moreover, they push the boundaries of the discussion toward an emergent theory that sees a co-dependency of the two disciplines, which are capable of influencing each other’s research advancements.

Part II Editors: Darren J. Edwards[1] and Rita Borgo[2]
[1]Swansea University, UK and [2]King’s College London, UK

Chapter 6
Visualization Onboarding Grounded in Educational Theories

Christina Stoiber, Markus Wagner, Florian Grassinger, Margit Pohl, Holger Stitz, Marc Streit, Benjamin Potzmann, and Wolfgang Aigner

Abstract The aim of visualization is to support people in dealing with large and complex information structures, to make these structures more comprehensible, facilitate exploration, and enable knowledge discovery. However, users often have problems reading and interpreting data from visualizations, in particular when they experience them for the first time. A lack of visualization literacy, i.e., knowledge in terms of domain, data, visual encoding, interaction, and also analytical methods can be observed. To support users in learning how to use new digital technologies, the concept of onboarding has been successfully applied in other domains. However, it has not received much attention from the visualization community so far. This chapter aims to fill this gap by defining the concept and systematically laying out the design space of onboarding in the context of visualization as a descriptive design space. On this basis, we present a survey of approaches from the academic community as well as from commercial products, especially surveying educational theories that inform the onboarding strategies. Additionally, we derived design considerations based on previous publications and present some guidelines for the design of visualization onboarding concepts.

C. Stoiber (✉) · M. Wagner · F. Grassinger · W. Aigner
St. Poelten University of Applied Sciences, St. Poelten, Austria
e-mail: christina.stoiber@fhstp.ac.at; markus.wagner@fhstp.ac.at; florian.grassinger@fhstp.ac.at; wolfgang.aigner@fhstp.ac.at

M. Pohl · B. Potzmann
Vienna University of Technology, Vienna, Austria
e-mail: margit.pohl@tuwien.ac.at; benjamin.potzmann@tuwien.ac.at

H. Stitz
datavisyn GmbH, Linz, Austria
e-mail: holger.stitz@datavisyn.io

M. Streit
Johannes Kepler University Linz, Linz, Austria
e-mail: marc.streit@jku.at

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_6

6.1 Introduction

The term onboarding was originally coined in the context of HR processes to support new employees in learning about their tasks that are part of their job within a particular company [40]. The aim of this ongoing process is to communicate not only formal knowledge about their tasks but also informal knowledge about organizational culture and its unwritten rules, to the new employees. This concept has been transferred to other domains such as human–computer interaction (HCI) [6, 8, 10, 18, 22, 38, 50]. More recently, the focus of onboarding has shifted toward mobile applications. Hulik[1] introduced the concept of supporting users in learning smartphone applications and software tools. Kumar defined *user onboarding* as "the process of increasing the likelihood that new users become successful when adopting your product" [42].

We think that it is also useful to conceptualize the process of learning about complex visualizations that cannot be understood at a first glance by having visualization onboarding concepts. We define visualization onboarding as follows: ***"Visualization onboarding is the process of supporting users in reading, interpreting, and extracting information from visual representations of data."*** [67]. This learning process often takes place immediately before or while users work with the visualization and is highly task-oriented. In this context, theories about learning play an important role. In the visualization community, a considerable amount of research has addressed the question of how to increase visualization literacy (see, e.g., [31, 57]). This research is generally based on educational theories from psychology, especially on constructivist research. The basic assumption is that knowledge about visualizations can best be acquired by creating one's own visualization and actively generating one's own view about this topic. Similarly, educational theories can also be adopted to explain the usefulness of onboarding approaches.

In the literature, several different possibilities of how to realize onboarding have been suggested (see Table 6.1). Some of them are primarily based on cognitivist approaches (e.g., tutorials) [43, 49] and Gestalt psychology (using analogy as a learning principle [58]). The educational theory on which these solutions are based is sometimes reflected explicitly and sometimes not. The discussion of this topic could help to clarify which approaches in the design of onboarding systems are more helpful than others. Informal evidence indicates that tutorials are often not read, and users just proceed and start working and exploring features of the system themselves. Nevertheless, commercial systems often rely on tutorials as well as help websites as onboarding systems, e.g., [1, 32, 47, 69].

We present a descriptive design space, presented in Fig. 6.2, covering aspects of visualization onboarding especially with the focus on educational theories. We conduct a systematic literature review to identify the state of the art in visualiza-

[1] https://useronboard.com, accessed: 2021-04-30.

Table 6.1 Overview of available visualization onboarding approaches (rows), systematically characterized along the aspects of our conceptual framework (columns). The table is divided into academic research and concepts (upper half) and commercial tools (lower half) which make use of various onboarding concepts. With a main focus on the questions, we took four of them and mapped them to the categorization of available approaches. The colors refer to the equivalent questions explained in the subsections of 3 ■– —applicable □– —not applicable, and *n.a.*—not available/unknown

Name	Who?				How?						Where?	When?
	Domain Knowledge	Data Knowledge	Visual Encoding & Interaction Knowledge	Analytical Knowledge	Type	Medium	Tool-specific	Context-sensitivity	Interactivity	Educational Theory	external / internal	before / while
Alper et al., 2017 [4]	■	■	■	■	teaching tool	text, pictograms	—	context-free	active	concreteness fading	external	before, while
Ola & Sedig, 2017 [50]	n.a.	n.a.	n.a.	n.a.	video-tutorial	video	✓	context-free	passive	n.a.	external	before
Kwon & Lee, 2016 [44]	□	□	■	■	interactive tutorial walkthrough, video, static	screenshots, video, text, visual elements	—	context-free	active, passive, reactive	experimental learning model	external	before, while
Yalçın, 2016 [80]	■	■	■	□	topic listing, point & learn, guided tour, notifications, topic answers	text	✓	context-sensitive, embedded	reactive	n.a.	internal	while
Tanahashi et al., 2016 [71]	□	■	■	□	InfoVis guide	text-plus-questions	—	context-free	active	top-down & bottom-up	external	before, while
Ruchikachorn & Mueller, 2015 [59]	□	□	■	□	video tutorial	animated visualization sequences	—	context-free	active, passive	learning-by-analogy	external	before, while
Kang et al., 2003 [35]	□	■	■	□	step-by-step overlays	text	✓	context-sensitive, embedded	reactive	n.a.	internal	while
Bishop et al., 2020 [10]	□	□	■	□	free-form constructive visualization tool	visual elements	—	context-free	reactive	scaffolding via visual feedback, learning from shared experiences	external	before, while
Firat et al., 2020 [21]	□	■	■	□	instructional software tool	visual elements, text	—	context-free	active	active learning	external	before, while
SAS *JMP* [62]	□	■	■	■	website, video, step-by-step overlays	text, videos, images, visual elements	✓	context-sensitive, -free	active, passive, reactive	n.a.	internal, external	before, while
Advizor Solutions *Advizor* [1]	□	■	■	■	website, overlay	text, videos, images	✓	context-sensitive, -free	passive	n.a.	external	before, while
SAP *Lumira* [61]	□	■	■	■	website, video	text, videos, images	✓	context-free	passive	n.a.	external	before, while
IBM *Cognos Analytics* [33]	□	□	■	□	website, video, interactive guided tour	text, videos, images, visual elements	✓	context-sensitive, -free	active, passive, reactive	n.a.	internal, external	before, while
TIBCO *Jaspersoft* [73]	□	■	■	□	website, overlay	text, videos, images	✓	context-sensitive, -free	passive	n.a.	internal, external	before, while
Microsoft *Power BI* [48]	□	■	■	□	website, video, examples, ask questions in app	text, videos, images	✓	context-sensitive, -free	passive	n.a.	external	before, while
SAS *Visual Analytics* [63]	□	■	■	■	website, video, interactive guided tour, courses, books	text, videos, images, visual elements	✓	context-sensitive, -free	active, passive	n.a.	internal, external	before, while
Tableau Software *Tableau* [70]	□	■	■	■	website, video, courses, books	text, videos, images	✓	context-free	passive	n.a.	external	before, while
TIBCO *Spotfire* [74]	□	□	■	■	website, video, courses, books	text, videos, images	✓	context-free	passive	n.a.	external	before, while
QlikTech *QlikView* [54]	□	■	■	□	website, videos, courses, overlays	text, videos, images	✓	context-sensitive, -free	passive	n.a.	external	before, while

tion onboarding and to categorize the work by summarizing existing onboarding concepts in scientific publications and commercial visualization tools using the *Five W's and How* [24, 25]. **WHY** is visualization onboarding needed? **WHAT** is visualization onboarding? **WHO** is the user? Which knowledge gap does the user have? **HOW** is visualization onboarding provided? **WHERE** is visualization onboarding provided? **WHEN** is visualization onboarding used? Additionally, we derived design considerations based on the collected publications and provide some existing guidelines for the application of educational theories for visualization onboarding in Sect. 6.4.4. Overall, we can report that whether other approaches are better for onboarding or it is still not an open question. Empirical research based on educational theories could help to gain more systematic information about this area.

6.2 Related Work

As visualization onboarding aims at filling the knowledge gaps of users by supporting the learning of new concepts, it makes sense to build upon knowledge from the fields of learning theories and cognitive science (see Fig. 6.1). Therefore, we present the related work for visualization onboarding, educational theories in visualization and cognitive science, as well as how explicit knowledge relates to onboarding in the following subsections.

6.2.1 Visualization Onboarding

So far, there has been little discussion about onboarding concepts for visualization techniques and visual analytics (VA) tools. Tanahashi et al. [70] investigated top-down and bottom-up teaching methods as well as active or passive learning types. The bottom-up teaching method is a method focusing on small, detailed pieces of information on which students then incorporate together for comprehensive understanding. A top-down teaching method is given when a broad overview helps to understand the abstract, high-level parts of an idea which then provide context for understanding its components in detail [70]. Passive learning means that students only receive the information without participatory dialog. In contrast, active learning describes an active participation [70]. Their analysis indicates that top-down exercises were more effective than bottom-up and active learning types with top-down tasks were the most effective ones. In their comparative study, Kwon and Lee [43] explored the effectiveness of active learning strategies. Three tutorial

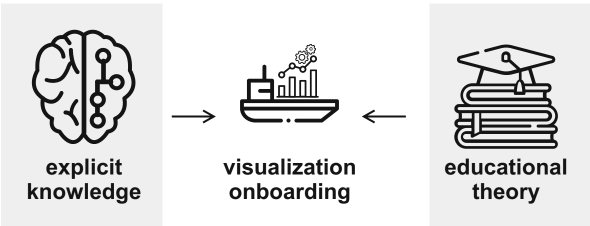

Fig. 6.1 Visualization onboarding aims to support end users in comprehending data visualizations and take full advantage of the tools at hand. With effectively designed onboarding methods, the knowledge gap of users could be filled. Thus, it makes sense to tap in the field of educational theories as well as identify how onboarding can benefit from explicit knowledge

types—static, video-based, and interactive—were used to support the learning of scatterplot visualizations. They observed that participants who used interactive and video tutorials outperformed participants who used static or no tutorials at all. In a study which set out to determine the power of teaching unfamiliar visualization by linking it to a more familiar one, Ruchikachorn and Mueller [58] found that the learning-by-analogy concept can be useful when the visualization method to be learned is inherently more powerful than its counterpart. They assessed four combinations and compared their difference in visual literacy: scatter plot matrix against hyperbox, linear chart against spiral chart, hierarchical pie chart against treemap, and data table against parallel coordinates. The spiral charts seemed to be the most difficult one to understand for the participants. The authors describe also another advantage of learning-by-analogy over other forms of demonstrations such as textual or oral descriptions is the power of visuals, as they bridge any language barriers. The educational community has also studied how students interpret and generate data visualizations [5] and how to teach bar charts in early grades [3] using a tablet app, called "C'est la vis," supporting elementary school pupils to learn how to interpret bar charts based on the concreteness fading approach. Concreteness fading is a pedagogical method where new concepts are presented with concrete examples at first, before progressively abstracting them. Recently, Bishop et al. [9] developed a tablet-based tool called Construct-A-Vis, which supports elementary school children in creating visualization based on free-form activities. They used scaffolding as a pedagogical method. In detail, they integrated feedback mechanisms by showing if the visual mapping was correct. Additionally, Bishop et al. [19] developed an interactive pedagogical method for training and cognition of a treemap design, as well as a treemap literacy test. The user study showed that students who interacted with the teaching tool outperformed those students who learned through slides.

Besides, there are platforms and websites available. For example, *The graphic continuum* [68], which provides an overview and helps choose the appropriate design or visualization type. Additionally, the *Data Viz Catalog* [55], a library of different visualization types, seeks to help users understand the encoding and building blocks of different visualization types. Furthermore, a decision tree provided by *From Data to Viz* [28] helps to find an appropriate visualization type based on the input data. The catalog offers definitions, variations, and the use of each visualization type in addition to potential issues that may arise during use and interpretation. These systems are not related to a particular visualization tool, neither integrate any educational theories nor provide validations. Recently, Wang et al. [77] presented a set of cheat sheets to support visualization literacy around visualization techniques inspired by infographics, data comics, and cheat sheets that are established onboarding methods in domains such as machine learning.

Besides scientific literature, onboarding concepts are integrated in commercial visualization tools as well. Nowadays most of these commercial visualization tools already integrate onboarding concepts focusing on the explanation of features, see Table 6.1. Yalçin [79] presented HelpIn, a design of a contextual in situ help system to explain features of Keshif [39]. Furthermore, IBM Cognos Analytics [32], for example, uses step-by-step tours with tooltips and overlays for onboarding new users. A more traditional approach is used by the commercial visualization tool Advizor [1], which makes use of textual descriptions to explain the visual mapping for visualization techniques.

6.2.2 Educational Theories in Visualization and Cognitive Science

Visualization onboarding supporting users in learning new concepts [67]; therefore it makes sense to build upon the knowledge from the field of learning theories and cognitive science. We distinguish between three main educational theories: behaviorism, cognitivism, and constructivism [16]. Behaviorism is an educational theory that only focuses on objectively observable behaviors and discounts any independent activities of the mind [78]. It is based on positive and negative reinforcement techniques. Besides, cognitivism is a philosophy of learning, founded on the premise that learning can be modeled as a kind of information processing [16]. Each of us generates our own "rules" and "mental models," which we use to make sense of our experiences. Learning, therefore, is the process of adjusting our mental models to accommodate new experiences. E-learning systems often integrate elements from different educational theories. This also applies to most onboarding systems.

Constructivist theories seem to be the one most appropriate for explaining learning processes with onboarding systems because they reflect on the application of learning in a practical context. The concept of cognitive apprenticeship plays an important role in constructivism [15, 63]. *Cognitive apprenticeship* is a kind of guided participation by learners in real processes of knowledge generation. This is related to the concept of *scaffolding* [27] where teachers gradually reduce the level of support for the student until the student is able to work autonomously. Cognitive apprenticeship and scaffolding can explain processes related to onboarding because the goal of the learners is to solve a real task, while the guidance is gradually reduced.

Another theoretical framework relevant for onboarding is *graph comprehension*, a theory that aims to explain how users make sense of graphs. Most of the investigations in this context deal with simple, small graphs [52]. Nevertheless, the findings from graph comprehension yield interesting results that can inform the design of visualizations. This is especially valuable for onboarding systems because investigations in this area often address the issue of how to design graphs that are appropriate for use in educational contexts. One of the most influential models in the context of the theory of graph comprehension describes this activity as consisting of three stages [21]. These three stages are (1) reading the data (i.e., finding individual data values), (2) reading between the data (i.e., finding relationships between the data), and (3) going beyond the data (i.e., interpreting the data, developing hypotheses about the data). Educational graphs are supposed to support all three stages, but the ultimate goal is to induce learners to "go beyond the data," that is, to reflect on the data and draw conclusions. Shah et al. [66] argue that inexperienced users typically concentrate on single data points or single lines in line graphs, whereas experts are able to actually interpret patterns in the data. Peeck [51] investigated whether it is possible to motivate learners to process graphs more comprehensively. In this context, the author successfully tested whether specific instructions for the processing of graphs support learning. The author also postulates that other measures such as cues to draw the learner's attention or motivating the learner to solve simple tasks by using the graphs are beneficial. Based on this approach, it can be recommended that onboarding should especially support "going beyond the data" and that instructions and visual cues can help users to better understand visualizations.

A further learning theory relevant for onboarding is *Microlearning*. Microlearning as an approach is a reaction to several technical developments. First, mobile technologies enable learners to learn flexibly, e.g., on the way to work, while traveling on public transport or while waiting for a physician. In addition, microlearning is also relevant for workplace learning and continuing education [64]. Employees in companies or other organizations do not need lengthy explanations but focused information that is necessary to continue their work. Microlearning has been defined as "special moments or episodes of learning while dealing with specific tasks or content, and engaging in small but conscious steps." [30]. Microlearning usually encompasses small units of learning that never take longer than 15 min. The situation described for Microlearning in the context of workplace learning is similar to the situation of users of complex information visualization systems.

Finally, *Gestalt psychology* is a theory that might be relevant for the design of onboarding systems. It is well known that Gestalt psychology has made important contributions in the area of perceptual psychology. It is less well-known that Gestalt psychologists also conducted relevant research in the area of reasoning and problem solving (see, e.g., [29, 46]). This is especially interesting for the design of visualizations as Gestalt psychologists conceptualized problem solving as the (sudden) perception of structure in a problem domain. The so-called Aha-moment is the moment when pieces fall into place and coherent structure is identified. In this context, the usage of analogies plays an important role because the transfer of structural knowledge from a well-known domain to an unknown domain is one of the learning methods that was suggested by Gestalt psychologists. Analogies can also be used to support onboarding in improving the understanding of complex visualizations as shown in the concept by Ruchikachorn and Mueller [58].

6.2.3 Knowledge Integration for Onboarding

In this section, we describe how user onboarding can benefit from *explicit knowledge* sources and contribute to generate new knowledge and insights. Based on the previously introduced terminology, we further characterize knowledge in Sect. 6.3.2.1, listing all the possible knowledge types which are needed, supporting meaningful onboarding.

In this work, we mainly consider *explicit knowledge*, i.e., knowledge as the source for providing onboarding. Usually, two types of *prior knowledge* are needed by a user to analyze data: *operational knowledge* (how to interact with the information visualization system) and *domain knowledge* (how to interpret the content) [11]. While a focus on usability and a perception- and cognition-aware design can alleviate the need for operational knowledge, domain knowledge cannot be easily replaced [11]. Stoiber et al. [67] further enhanced the levels of the users' *prior knowledge* for visualization onboarding based on the nested model [48] as (1) *domain knowledge* (e.g., vocabulary and concepts), (2) *data knowledge* (understanding the particular datatype), (3) *visual encoding knowledge* (understanding the visual mapping), (4) *interaction knowledge* (for performing tasks and understand relations in the data), and (5) *analytical knowledge* (knowledge of different automated data analysis methods)—see Sect. 6.3.2.1 for more details. However, Chen et al. [11] as well as Stoiber et al. [67] described the term *prior knowledge* at different granularities, whereby *operational knowledge* [11] can be seen as similar to the combination of *visual encoding, interaction, and analytical knowledge* [67].

Thomas and Cook [71] describe that the proper representation of final as well as intermediate *generated knowledge* can be useful to support the analytical discourse. By retaining quality and provenance information, it supports the interoperation between human and machine components, the collaboration between different users, as well as tracing the relations between data and derived knowledge products. In 2005, Thomas and Cook [71, p. 35] incorporated *prior domain knowledge* and *building knowledge structures* as one of the open challenges of the VA agenda. This is supported by the central role of knowledge in the VA process model by Keim et al. [36, 37] and further process models such as the knowledge generation model by Sacha et al. [59] and the visualization model by van Wijk [75]. However, these process models do not differentiate between knowledge in the human and the machine space. Based on Wang et al. [76] as well as Federico and Wagner et al. [17], *tacit knowledge* is exclusively available to/by human reasoning and can be extracted as *explicit knowledge* to become machine usable in the VA environment. Additionally, the integration of *explicit knowledge* into the VA process is formalized in several recent models by Wang et al. [76], Ribarsky et al. [54], as well as Federico and Wagner et al. [17]. Beyond the role of knowledge in the VA process, only few works discuss the content and structure of *explicit knowledge* on a general level [2, 4, 44, 56, 65, 74]. As visualization onboarding aims to fill different knowledge gaps of the user, the former described knowledge generation and transformation concepts can be used.

6.3 Descriptive Design Space

In our previous work [67], we introduced a descriptive design space for visualization onboarding. This work enhances the design space and discusses the role of educational theories in the context of onboarding.

6.3.1 Construction of Design Space

We structured the design space based on *Five W's* and the appended *How* [24, 25]. These questions are frequently used to describe a matter from its most relevant angles in technical documentation and communication. Furthermore, the same questions have already been employed for structuring the use of visualization for healthcare informatics [80] and in a survey on the role of visual analytics in deep learning research [26]. We describe the space of visualization onboarding along the following questions: **WHO** is the user? Which knowledge gaps does the user have? **HOW** is visualization onboarding provided? **WHERE** is visualization onboarding provided? **WHEN** is visualization onboarding used? Inside of each dimension (question), we defined several categories which are described in detail in the section below. We followed an open coding approach for the survey of onboarding concepts where we unified top-down approaches as well as bottom-up categorizations. Where available, we used the existing taxonomies or frameworks, which we adapt to the specifics of visualization onboarding.

6.3.2 Design Space Dimensions

The aim of visualization onboarding is to support human in dealing with large and complex information structures, to make them more comprehensible, facilitate exploration, and enable knowledge discovery. Nevertheless, the users often have problems in reading and interpreting data from visualizations, in particular when they experience them for the first time. In this section, we present the design space dimensions of visualization onboarding and show its various aspects.

6.3.2.1 WHO Is the User?

Users need to understand the process and reasoning that lead to the visual appearance, interactive behavior, and findings. Making this process transparent to the users is a central aspect in the design of visual analytics solutions. For conceptualizing this aspect, we adapt the nested model by Munzner and colleagues [48] as the guiding framework for presenting different levels of knowledge. The nested model

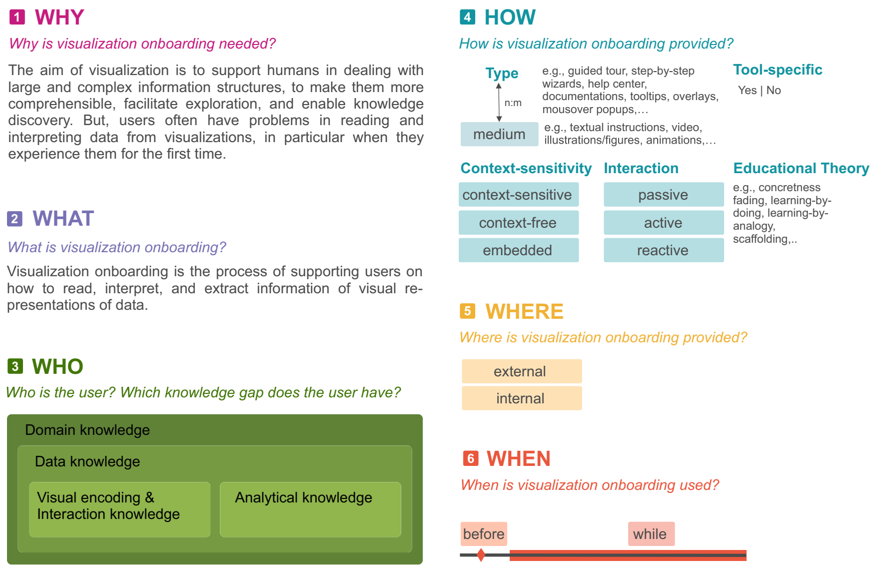

Fig. 6.2 A visual overview of the onboarding design space and of how all six questions "Why, What, Who, How, Where, and When" relate to one another. Each question corresponds to one paper section as indicated by the numbered tag near each question title

is a unified approach that splits the design into four levels and combines these with appropriate evaluation methods to mitigate threats to validity at each level. In order to be able to cover visual analytic approaches and include automated data analysis components, we expand the original model by adding analytical methods alongside visual encoding/interaction idioms. Analytical knowledge—such as different automated data analysis approaches, machine learning methods, or statistical methods applied to the data—is necessary to understand complex visualization interfaces and data. Figure 6.2 (3) keeps the nesting but shows an altered representation of the different levels. The model components represent the different levels of knowledge that (a) visualization users need in order to correctly interpret (interactive) visualization artifacts and (b) visualization designers have to consider when developing onboarding concepts.

The levels consider the users' prior knowledge such as domain knowledge, data knowledge, knowledge of visual encoding and interaction concepts, and analytical knowledge.

Domain knowledge: A specific domain is a particular field of interest by target users of a visualization tool (e.g., medicine, data journalism, bioinformatics). Each domain has its own vocabulary for describing the data and problems, workflows, and how data can be used to solve a problem. Domain knowledge is also an ensemble of concepts, intellectual tools, and informational resources that a user can draw upon to put the visualized data into context.

Data knowledge: Many visualization tools are specific to a particular *type of data*, such as multivariate data, hierarchical data, network data, or time-oriented data. Data knowledge refers to the necessary knowledge for understanding the data types and structures or statistical properties of the data. In many cases, users need to know how to get their data into a specific visualization tool as a first step. This relates to a more technical level of knowledge about a particular file format (e.g., CSV, JSON) or structure of the data—*data format*—(e.g., order and data types of individual variables).

Visual encoding knowledge: This type of knowledge is the most obvious one in the context of visualization, as it concerns the visual appearance of the data. Data elements are mapped to visual marks and channels to form visualizations. Understanding this mapping is the basis for being able to correctly interpret the visualization.

Interaction knowledge: Interactivity is crucial for visualization tools. An interactive visualization tool can support the investigation at multiple levels of detail, such as either a high-level overview or fully detailed views that show a small data subset only [48]. Understanding the interaction concepts used in a visualization tool is important for users for an active discourse with the data, i.e., to perform tasks and understand connections and relationships in the data.

Analytical knowledge is defined as the knowledge of different automated data analysis methods, for example, clustering (e.g., k-means) or data aggregation (e.g., dimensionality reduction). In certain cases, users need to have at least a basic understanding of their characteristics in order to choose or parameterize them correctly.

6.3.2.2 HOW Is Visualization Onboarding Provided?

Onboarding type, medium, context sensitivity, interaction, tool-specific, and educational theory are relevant aspects of the question of how visualization onboarding is provided. The *onboarding type* captures the used *medium*. The form of *contextual aid* is extremely important for applications [23]. The help system should be designed to guide users by demonstration in the *context* of their own interface. Chilana et al. [12] developed an approach to provide a new framework for integrating crowd-sourced contextual help into web applications. In their work, they also discussed the importance of contextual help and *adaptive help* systems. Based on these results, we also integrate the aspect of *context sensitivity* into our framework for visualization onboarding. Fernquist et al. [18] introduced a set of the most relevant aspects for interactive tutorials for a sketching software. Based on their design space for sketching software, we adopted the aspect of *interactivity*. Additionally, we integrated the category *tool-specific* indicating if the onboarding concept is connected to a visualization tool or not. Visualization onboarding supports users in learning new concepts. Hence, we integrated the category of educational theory.

Onboarding type and medium: Onboarding can be provided in different types, such as guided tours, step-by-step wizards, video-based tutorials, and help centers.

We derived this terminology from our literature review and Pronovic's blog article about context-sensitive and embedded help formats [13]. A particular type of onboarding consists of a *medium* which can be, e.g., textual instructions, video, illustrations/figures, animations, etc.

Context Sensitivity: Context-sensitive help provides assistance at a specific point in the current state of the tool. It is the smallest possible chunk of information the user needs to understand at that point. Examples are in-application help centers, guided tours, or mouseover popups including instructional material. A type of context-sensitive help is **embedded help** which goes beyond basic information and explanations by either detecting the user's need for help or offering a guided tour right on the interface. Examples are tooltips, instructions on the interface, or walkthroughs. **Context-free help** can be called at any state of usage and does not relate to the current state of help-seeking. Examples are online documentations and help videos.

Interaction: Interaction is applied within the onboarding process itself. We refer to Fernquist [18] for defining the degree of interactivity in onboarding concepts. Help systems can be **passive** if the user only consumes the learning material, such as reading an article or viewing a video. If users can try out the concepts, the onboarding concept is defined as **active**. Active tutorials that are aware of the users' interactions and can respond to these are referred to as **reactive**.

Educational theory: The aspect of learning and educational theories is crucial when it comes to onboarding approaches. A systematic categorization of the educational theories was not possible to conduct as there is no taxonomy available. Therefore, we collect educational theories, which authors described in their scientific publications (e.g., concreteness fading [3], learning by analog [58], etc.)

Tool-specific: The category describes if the onboarding concept is designed for a specific visualization tool (tool-specific) or it is decoupled from it (non-tool-specific).

6.3.3 WHERE is Visualization Onboarding Provided?

Based on Fernquist et al. [18] who introduced a set of the most relevant aspects, we also adopted the aspect of the integration of onboarding concepts by asking *Where is visualization onboarding provided?*—externally, internally, or as a learning environment. An onboarding system that is integrated **internally** into the visualization can be more helpful for users because they do not have to jump back and forth between two different systems. **External** sources for onboarding concepts can be defined as sources which can be reached independently of the current state of the tool. At the tightest level of integration, help systems can be provided **internally**. It should be pointed out, however, that integrating onboarding systems into the visualization or visual analytics tools is challenging and requires a considerable effort.

6.3.4 *WHEN Is Visualization Onboarding Used?*

The aspect of *WHEN* describes the temporal aspect of intended onboarding use (see Fig. 6.2 (6)). Onboarding concepts can be integrated **before** using the actual visualization tool (one time or repeated) or called up **while** the use of a certain tool, e.g., when support regarding a particular feature is needed.

6.4 Survey on Visualization Onboarding

In this section, we describe the method used for our systematic literature review in detail. Furthermore, we present the results of the survey based on our descriptive design space.

6.4.1 *Method*

To get a comprehensive overview of existing onboarding concepts, we systematically surveyed the literature published in the main venues in the fields of information visualization, visual analytics, and HCI. In addition to scientific publications, we reviewed commercial visual analytic tools based on a recent study about commercial systems by Behrisch et al. [7] (see Table 6.1). We focused on the following major conferences and journals: *IEEE InfoVis*, *IEEE VAST*, *EuroVis*, *Eurographics*, *EuroVA*, *IEEE TVCG*, *Information Visualization* (IV), *ACM CHI*, and *ACM UIST*. Due to the fact that the term *onboarding* is rarely used in the visualization community, we used the following keywords: *data visualization literacy*, *visualization literacy*, *instructional material*, and *learning*. We scanned the title and abstract for the specific keywords.

We additionally examined papers published as part of various relevant workshops on the topic of *visualization literacy*, especially the *IEEE VIS DECISIVE Workshop*. We took into account both full and short papers. Moreover, we identified the authors of the most relevant papers and included further publications by these researchers. We scanned through the related work sections of the relevant papers to find more literature related to our topic. We were able to identify a total of nine papers that focus on onboarding concepts and learning environments for visualization or visualization tools [3, 9, 20, 34, 43, 45, 49, 58, 70, 79] as well as ten commercial tools that use a variety of onboarding methods and concepts [1, 32, 47, 53, 60–62, 69, 72, 73].

Every selected publication was categorized by two coders who are co-authors of this chapter. After the coding of the nine papers, we discussed the coding criteria and matched our coding strategy. In case of conflicting codes, coders discussed the reasons for decisions in order to resolve inconsistencies.

6.4.2 Results

We reviewed nine scientific publications and ten commercial tools with a special focus on onboarding concepts summarized in Table 6.1. In the following sections, we discuss and highlight the most relevant factors of onboarding methods we discovered.

6.4.2.1 WHO: Who Is the User? Which Knowledge Gap Does the User Have?

For both the scientific publications and the commercial tools, we recognized strong emphasis on *visual encoding and interaction knowledge* as well as *data knowledge* [1, 3, 9, 34, 43, 45, 47, 49, 53, 58, 60–62, 70, 72, 79]. Interestingly, Kwon and Lee [43], Ruchikachorn and Mueller [58], Bishop et al. [9], and the two visualization tools *IBM Cognos Analytics* [32] and *TIBCO Spotfire* [73] do not target *data knowledge explicitly*, which appears to be surprising as basic data knowledge is crucial in order to understand the visual encoding of a visualization. Only two publications [3, 43] cover *analytical knowledge*, while six of ten commercial tools provide support in this respect, e.g., classification and regression models [1]. We were able to identify a lack of *domain knowledge* in all tools and the majority of scientific publications. Only two publications focus on *domain knowledge* in their onboarding concepts [3, 79]. The publication of Ola and Sedig [49] was an exception insofar as we could not identify any of the knowledge gaps.

6.4.2.2 HOW: How Is Visualization Onboarding Provided?

In this dimension, we distinguish between five different aspects: *onboarding type and medium, context sensitivity, interactivity, tool-specific, and educational theory* (see Sect. 6.3.2.2). In terms of the onboarding type and medium, we found some similarities within the collection of publications. However, these have been the most difficult to gather, as the publications vary the most in their onboarding approaches. In the educational setting [3, 9, 20], the teaching tools use *text*, *visual elements*, as well as pictograms as *medium* to educate students. In terms of documented *onboarding type*, Alper et al. [3] introduced a "tool for teaching bar charts." More recently, Firat et al. [20] developed an instructional software tool for treemap visualizations, and Bishop et al. [9] introduced a "free-form constructive visualization tool." Besides, Kang et al. [35] as well as Yalçın [79] only integrated *text* in their onboarding approaches on *overlays*. Kang et al. [35] focused their concept on *step-by-step overlays*, in contrast, Yalcin [79] used for his approach overlays including a combination of *topic listing, point and learn, guided tour, notification, and topic answers*. A further similarity is the usage of video and/or animation to onboard users. For example, Ola and Sedig [49] as well as Ruchikachorn and Mueller [58] developed *video tutorials* using *animated visualization sequences* [58] (see Fig. 6.3 (3)) and a *video* [49] to support users in learning. In addition, we identified other types such as interactive walkthrough tutorials [43] and InfoVis Guides using text plus questions [70]. In general, most of the collected onboarding approaches use a combination of different medium and onboarding approaches.

All commercial tools could be systematically categorized only in terms of type and media using documentation/explanation websites with *screenshots* and *textual descriptions* (medium). The majority of tools also use *videos* as a medium to onboard users. *SAS JMP* [61], *IBM Cognos Analytics* [32], and *SAS Visual Analytics* [62] integrate step-by-step tutorials or interactive guided tours and therefore also rely on visual elements (chart parts to interact with, applicable filters, etc.). *TIBCO Jaspersoft* [72] and *Advizor* [1] make use of an in-application help overlay using text and videos. Additionally, *Microsoft Power BI* [47], *SAS Visual Analytics* [62], *Tableau* [69], *TIBCO Spotfire* [73], and *QlikTech QlikView* [53] provide a combination of books and courses. One special method to highlight is the in-application *ask questions* of *Microsoft Power BI* [47], which allows the users to ask a question related to the dataset they are currently working on.

Tool-specific: For the scientific publications, we identified three onboarding approaches which can be categorized as tool-specific [33, 49, 79]. The remaining six are non-tool-specific [3, 9, 20, 43, 58, 70]. We call these onboarding concepts *learning environments*, which are independent of a specific visualization tool and can be used in general.

Context sensitivity refers to the three categories: context-free, context-sensitive, and embedded concepts. Seven out of nine papers designed context-free onboarding concepts, while only Yalcin [79] and Kang et al. [34] use context-sensitive and embedded onboarding methods. On the other hand, three out of ten commercial tools integrate context-free onboarding concepts. The other commercial tools integrate context-free and context-sensitive methods as they are using documentation websites and also in-application overlays or guided tours. One example is *Advizor* [1], which makes use of context-free and context-sensitive onboarding methods (see Fig. 6.3 (2) for the design of the context-sensitive approach).

A more detailed investigation of the *interactivity* of the onboarding concepts described in publications we found revealed a good balance between the three types of interaction. The category *interactivity* is also connected with the used educational theory. Four of the nine onboarding concepts provide *reactive* onboarding [9, 35, 43, 79]. For the commercial tools, we observed a strong trend toward passive interactivity. Only two tools—*SAS JMP and IBM Cognos Analytics* [32, 61]—cover all three interactivity types. *IBM Cognos Analytics*, for example, provides videos and a website (passive) as well as an interactive guided tour (reactive) to onboard users.

In terms of the integrated *educational theories*, we could not find any unique use of educational theories among the onboarding approaches presented in publications. Thus, we identified the following aspects: (1) *onboarding approach designed without the integration of educational theories* [35, 49, 79], and (2) *onboard-*

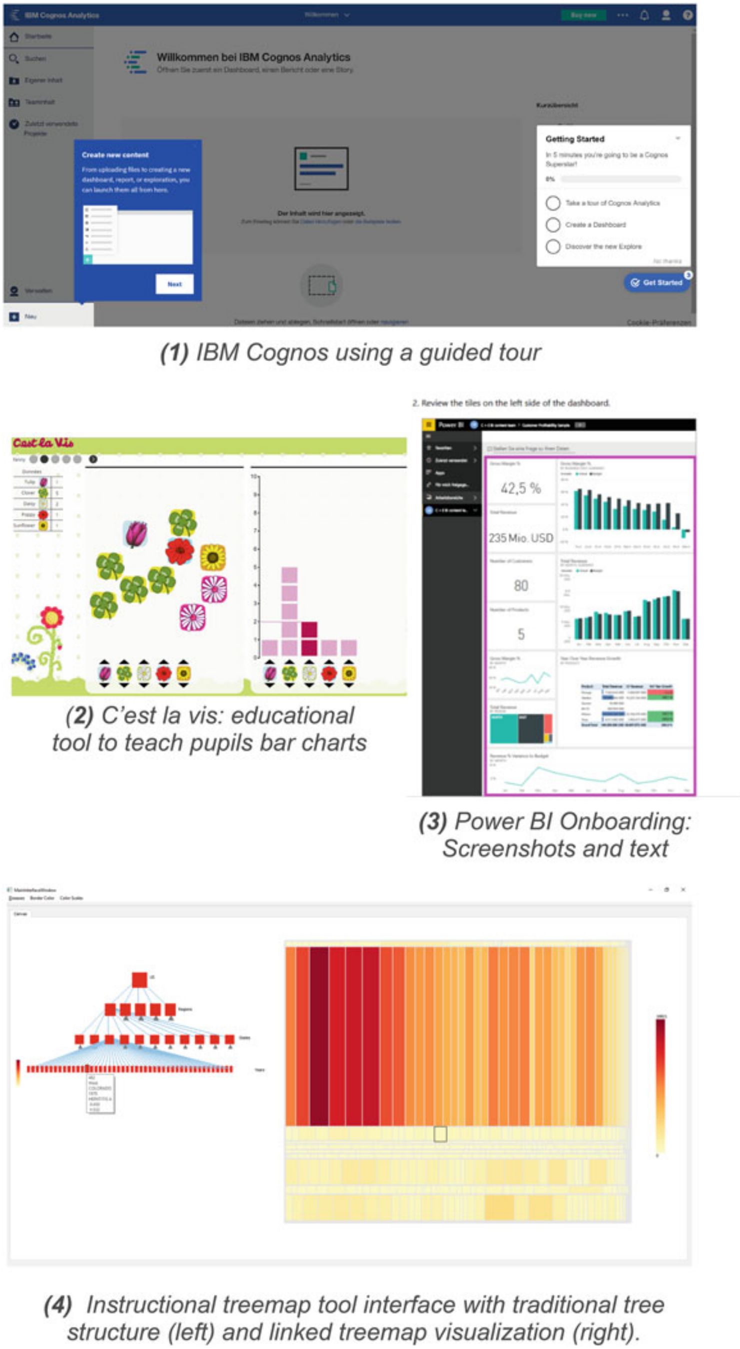

Fig. 6.3 Onboarding approaches: (1) IBM Cognos [32], (2) Alper et al. [3] onboarding method based on the concreteness fading educational theory, (3) PowerBI external webpage with instructional material (screenshots and text) [47], and (4) educational instructional material for treemap visualization [20]

ing approaches grounded in educational theories: (2a) concreteness fading [3], (2b) Experiential learning model [43], (2c) top-down and bottom-up [70], (2d) learning-by-analogy [58], (2e) scaffolding via visual feedback, learning from shared experience [9], and (2f) active learning [20]. In the following, we describe two examples in detail showing how visualization onboarding has been applied.

Example on Experiential Learning Model (2b): one example for a reactive onboarding is from Kwon and Lee [43], who developed an online learning approach for parallel coordinates following the experiential learning model (see Fig. 6.3 (6)). The model defines learning as the process in which knowledge is constructed via concrete experience and reflection on the experience [41]. Therefore, the interactive tutorial page integrates the experiential learning model's four stages (Concrete Experience, Reflective Observation, Abstract Conceptualization, and Active Experimentation). The authors implemented the model as follows. For the first stage, the *Concrete Experience*, the people are asked to complete a mission. For the *Reflection Observation* stage, the onboarding approach provides hints to the user interactions. Additionally, "the system shows the conceptual goal of the activity at a successful completion" [43] (*Abstract Conceptualization*). For the fourth level—*Active Experimentation*—the learning approach suggests to repeat the activity to strengthen the learning. They conducted a comparative evaluation with three tutorial types (static, video-based, and interactive tutorial walkthrough). They observed that participants using the interactive and video tutorials outperformed participants with static or no tutorials.

Example on Learning-by-Analogy (2d): In addition to onboarding using the experiential learning model [43], Ruchikachorn and Mueller [58] proposed a concept for the teaching of unfamiliar visualizations by using the educational theory of *learning-by-analogy*. This is an example of a combination of *passive* and *active* onboarding system. Based on animated visualization sequences (passive), the users were taught a more advanced visualization technique based on an easier one with transitions as presented in Fig. 6.3 (3). The user was able to watch the sequences which can be categorized as a *passive* interaction. Additionally, the user was able to start and stop animating the morphing (*active*).

6.4.2.3 WHERE: Where Is Visualization Onboarding Provided?

Our survey of existing work and commercial tools showed that the majority of onboarding solutions can either be classified as external or internal or a combination of both sources. Yalçın [79] and Kang et al. [34] designed an internal onboarding concept. All other solutions can be categorized as external onboarding approaches. For commercial tools, there is a fairly equal distribution between only *external* ones and those who are *external and internal*. The majority of commercial tools provide external material such as documentation sites with text, images, and videos.

6.4.2.4 WHEN: When Is Visualization Onboarding Used?

Onboarding concepts can be integrated at different states of use—before or during. Ola and Sedig [49] relied on a *before* approach, in contrast, Yalçın [79] and Kang et al. [34] provide their onboarding *while* the usage. Other onboarding approaches [3, 9, 20, 43, 58, 70] can be either used *before* or *while*. We detected a clear tendency for commercial tools as all of the onboarding concepts can be used before and during usage of the particular visualization tool.

6.4.3 Summary

Considering the **WHO** question, we observed a strong tendency toward *visual encoding and interaction knowledge* [3, 9, 20, 34, 43, 58, 70, 79]. *Data knowledge* is also prominent in the literature [3, 20, 34, 70, 79]. However, *domain knowledge* [3, 79] and *analytical knowledge* [3, 43] are covered only by two out of nine investigated papers. Only Alper et al. [3] are targeting all knowledge gaps. Regarding the question of **HOW** is onboarding provided? we found a variety of different onboarding types. This ranges from simple texts instructions [79] or videos [49, 58] to interactive visual elements [3, 9, 20, 43] or step-by-step guides [34]. Regarding context sensitivity, most of them are using a context-free approach [3, 9, 20, 43, 49, 58, 70], with two exceptions that are context-sensitive and embedded in the visualization tool [34, 79]. Those two exceptions are also *internal* looking at the **WHERE** aspect. All others are designed as non-tool-specific onboarding approaches, i.e., not directly integrated into a visualization tool which are then *external*.

In the case of *educational theory*, however, no general statement can be made based on the categorization of the papers, since each paper follows a different educational theory. However, we observed similarities regarding the educational theories, which are presented in Sect. 6.4.4. In general, most of the collected onboarding approaches of the commercial tools are designed to be used *before* and *while* interacting with a particular visualization tool (**WHEN**).

None of the commercial tools address or attempt to explain the *domain knowledge* of the users. The tools mainly cover only the *data knowledge* [1, 47, 53, 60–62, 69, 72] as well as the *visual encoding and interaction knowledge* [1, 32, 47, 53, 60–62, 69, 72, 73]. In general, the tendency to convey analytical knowledge is much higher with commercial tools [1, 60–62, 69, 73] than with the scientific papers. In relation to tools, the type of **onboarding** mainly relies on help websites, video tutorials, or courses. There are a few exceptions [32, 61, 62] that also use visual elements offering more interaction. For *context sensitivity*, it is about evenly distributed among the tools, but there is no single embedded one. Also, the *interactivity* in the tools is mostly *passive* since the help often is only provided on demand. Exceptions to this are the three approaches [32, 61, 62] that offer guides or tutorials directly or react to user interaction. Unfortunately, it was not possible to identify an *educational theory* for any of the commercial tools, but this was to be expected, since they are established visualization software. The commercial tools have a balanced ratio in the question of **WHERE**. In terms of the **WHEN** question, all the onboarding approaches can be used while or before using the actual visualization tool.

6.4.4 Existing Design Considerations for Visualization Onboarding

In this section, we present existing guidelines derived from the collected papers. We focused on the given medium, type of onboarding, as well as the education theory used to onboard users.

- Kwon and Lee [43] developed an interactive guide for parallel coordinates plots based on a **learning-by-doing** approach. They followed the "Experiential learning model," which can be defined as the process in which knowledge is constructed via concrete experience and reflection on the experience [41]. The presented interactive tutorial walkthrough integrates textual descriptions as well as interactive visual elements (see Fig. 6.4 (3)), where, for example, the user can click on points in integrated parallel coordinates, whereupon lines are drawn that then connect them.
- In their paper, Ruchikachorn and Mueller [58] developed a teaching concept to learn and teach unfamiliar visualizations by linking it to a more familiar one. They followed the **learning-by-analogy** approach. The authors commented that their system can be useful when the visualization method to be learned is inherently more powerful than its counterpart. Their approach overcomes languages barriers as it uses visuals.
- The results of the conducted study by Tanahashi et al. [70] showed that tutorials where users can directly interact with the visualization will influence the comprehension positively. They suggest to use **active learning** type (participating actively in a corresponding dialog) with **top-down** exercises. In detail, this means to ask participants to draw more advances, less direct inferences from the data. Their study revealed that their approach of text-plus-question introductory tutorials is a useful and practical way to onboarding users to information visualizations.
- A recent study shows that there is a successful knowledge transfer to another concrete domain when concrete examples were given as opposed to abstract ones [14]. Based on these results, Alper et al. [3] developed a tablet app teaching elementary school pupils bar charts using the pedagogical method of **concreteness fading**. The tool provides a space with a reference line (x- and y-axis) as well as free-form pictograph that represents data in the form of illustrative icons which are scattered around. Children can stack the icons on

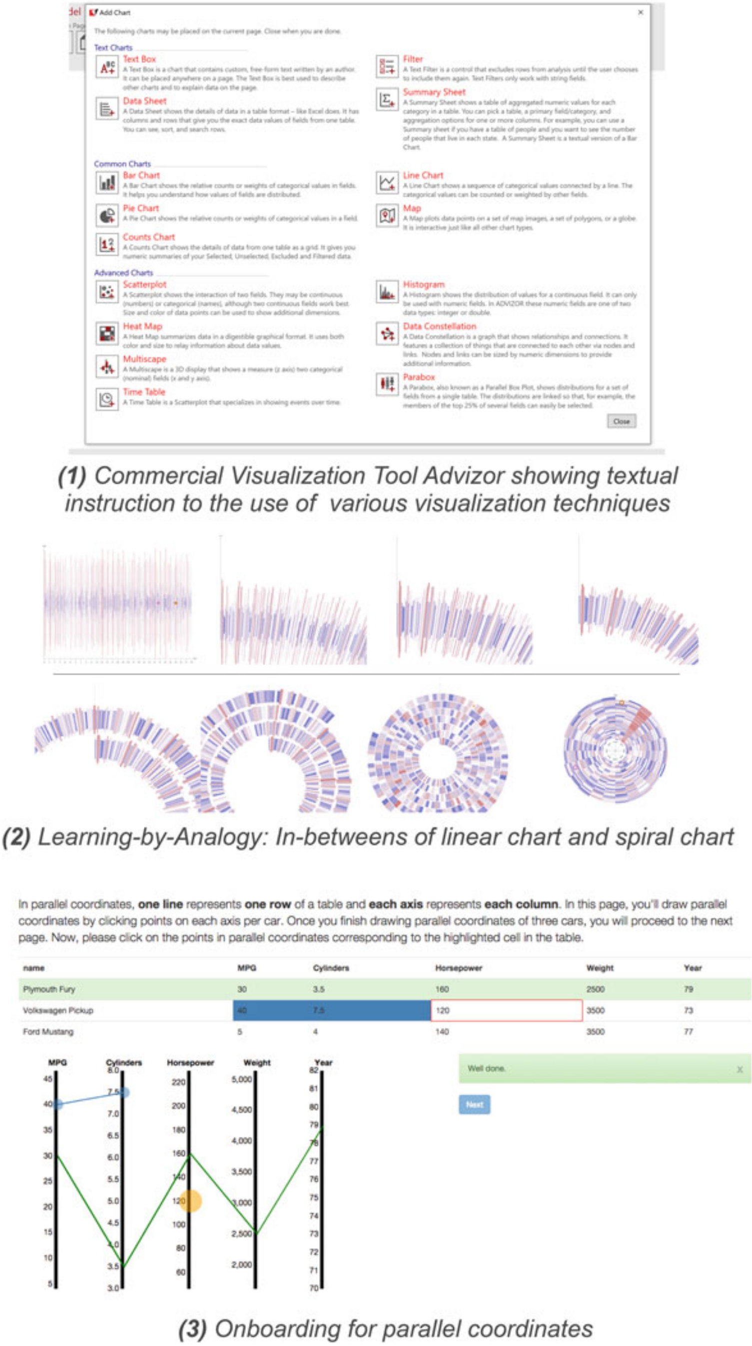

Fig. 6.4 Onboarding approaches: (1) Advizor [1], (2) learning-by-analogy developed by Ruchikachorn and Mueller [58], and (3) interactive tutorial based on Experiential Learning Model [43]

top of each other and then watch an animated transition morphing the icons into a more abstract representation of a bar chart (see Fig. 6.3 (2)).

- Bishop et al. [9] developed a free-form construction tool for tablets to engage pupils with the creation of visualization, as well as to make the visual mapping of data more explicit. **Scaffolding** was integrated as educational theory. The results of their study highlight the advantage of scaffolding within the creation process of visualizations through visual feedback, configurability, and shared interaction.

When we sum up and generalize the results of the empirical studies of the papers, as well as the results of the analysis of the design space, we propose the following guidelines when it comes to design onboarding methods: (1) explain the visual encoding and interaction concepts [3, 9, 43], (2) use interactive onboarding approaches, where users can interact with the visualization as well as with the instructional material [43, 70], (3) concrete experience and reflection can lead to higher understanding [43, 70], and (4) use animations or videos [3, 43, 58] to show the data-to-visual mapping.

6.5 Discussion and Conclusion

We presented a descriptive design space for visualization onboarding and presented design considerations based on the existing empirical studies. The design space contains the six aspects: **WHY** is visualization onboarding needed? **WHAT** is visualization onboarding? **WHO** is the user? Which knowledge gap does the user have? **HOW** is visualization onboarding provided? **WHERE** is visualization onboarding provided? **WHEN** is visualization onboarding used? We conducted a systematic literature review to develop the presented design space. Additionally, we also reviewed commercial visualization tools listed in Table 6.1. We especially focused on educational theories as the aspect of learning is important when it comes to the design of visualization onboarding (see Table 6.1 and Sect. 6.4.4). Ways to effectively support the learning process of users with different knowledge gaps can be considered by using educational theories. However, the literature lacks educational theories with a special focus on onboarding concepts. We tried to identify guidelines based on the existing literature, which we presented in Sect. 6.4.4. Nevertheless, existing theories and results of educational research can be used to inform the design of onboarding systems.

Onboarding systems can either be designed like help systems, which implies a cognitivist approach, or they might use a scaffolding approach [9], applying features such as prompts, tools to structure information, or higher order questions. Constructivist theory supports the assumption that especially higher order reasoning processes and the ability to make inferences and draw conclusions from the data are supported by cognitive apprenticeship or scaffolding in particular. Higherorder reasoning is not only the last stage in the model suggested by graph comprehension but also the ultimate goal of most visualization systems. Based on the papers, educational theories that support active learning and concrete experience are appropriate for onboarding. Further research is needed to empirically test these observations.

Acknowledgments This work was funded by the Austrian Science Fund as part of the VisOnFire project (FWF P27975-NBL), the Austrian Ministry for Transport, Innovation, and Technology (BMVIT) under the ICT of the Future program via the SEVA project (no. 874018), as well as the FFG, Contract No. 854184: "Pro2Future," which is funded within the Austrian COMET Program Competence Centers for Excellent Technologies under the auspices of the Austrian Federal Ministry for Transport, Innovation, and Technology, the Austrian Federal Ministry for Digital and Economic Affairs and of the Provinces of Upper Austria and Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

References

1. Advizor Solutions. https://www.advizorsolutions.com/software/interactive-data-visualization/, 2019. Accessed: 2019-03-14.
2. W. Aigner, A. Rind, and M. Wagner. KAVA-Time: Knowledge-Assisted Visual Analytics Methods for Time-Oriented Data. In *Tagungsband des 12. Forschungsforum der österreichischen Fachhochschulen (FFH) 2018*, 2018.
3. B. Alper, N. H. Riche, F. Chevalier, J. Boy, and M. Sezgin. Visualization literacy at elementary school. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 5485–5497, New York, NY, USA, 2017. ACM.
4. N. Andrienko, T. Lammarsch, G. Andrienko, G. Fuchs, D. Keim, S. Miksch, and A. Rind. Viewing visual analytics as model building. *Computer Graphics Forum*, published online before print, 2018.
5. R. S. Baker, A. T. Corebett, and K. R. Koedinger. Toward a model of learning data representations—eScholarship. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 23, 2001.
6. N. Banovic, T. Grossman, J. Matejka, and G. Fitzmaurice. Waken: Reverse engineering usage information and interface structure from software videos. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, UIST '12, pages 83–92. ACM, 2012.
7. M. Behrisch, D. Streeb, F. Stoffel, D. Seebacher, B. Matejek, S. H. Weber, S. Mittelstaedt, H. Pfister, and D. Keim. Commercial visual analytics systems-advances in the big data analytics field. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2018.
8. L. Bergman, V. Castelli, T. Lau, and D. Oblinger. DocWizards: A system for authoring follow-me documentation wizards. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 191–200. ACM, 2005.
9. F. Bishop, J. Zagermann, U. Pfeil, G. Sanderson, H. Reiterer, and U. Hinrichs. Construct-a-vis: Exploring the free-form visualization processes of children. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):451–460, 2020.
10. J. M. Carroll, editor. *Interfacing thought: Cognitive aspects of human-computer interaction.* Interfacing thought: Cognitive aspects of human-computer interaction. The MIT Press, 1987.
11. C. Chen. Top 10 unsolved information visualization problems. *IEEE Computer Graphics and Applications*, 25(4):12–16, July 2005.

12. P. K. Chilana, A. J. Ko, and J. O. Wobbrock. LemonAid: selection-based crowdsourced contextual help for web applications. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems—CHI '12*, page 1549. ACM Press, 2012.
13. An overview of context-sensitive and embedded help formats. https://pronovix.com/blog/overview-context-sensitive-and-embedded-help-formats, 2019. Accessed: 2019-03-14.
14. D. De Bock, J. Deprez, W. Van Dooren, M. Roelens, and L. Verschaffel. Abstract or Concrete Examples in Learning Mathematics? A Replication and Elaboration of Kaminski, Sloutsky, and Heckler's Study. *Journal for Research in Mathematics Education*, 42(2):109–126, 2011.
15. T. M. Duffy, J. Lowyck, and D. H. Jonassen, editors. *Designing Environments for Constructive Learning*. Nato ASI Subseries F:. Springer-Verlag, 1993.
16. P. A. Ertmer and T. J. Newby. Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, 26(2):43–71, 2013.
17. P. Federico, M. Wagner, A. Rind, A. Amor-Amorós, S. Miksch, and W. Aigner. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In *Proc. IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 2017. In press.
18. J. Fernquist, T. Grossman, and G. Fitzmaurice. Sketch-sketch revolution: An engaging tutorial system for guided sketching and application learning. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 373–382. ACM, 2011.
19. E. E. Firat, A. Denisova, and R. S. Laramee. Treemap Literacy: A Classroom-Based Investigation. In M. Romero and B. Sousa Santos, editors, *Eurographics 2020—Education Papers*. The Eurographics Association, 2020.
20. E. E. Firat, A. Denisova, and R. S. Laramee. Treemap Literacy: A Classroom-Based Investigation. In M. Romero and B. Sousa Santos, editors, *Eurographics 2020—Education Papers*. The Eurographics Association, 2020.
21. S. N. Friel, F. R. Curcio, and G. W. Bright. Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2):124–158, 2001.
22. T. Grossman and G. Fitzmaurice. ToolClips: An investigation of contextual video assistance for functionality understanding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1515–1524. ACM, 2010.
23. T. Grossman, G. Fitzmaurice, and R. Attar. A survey of software learnability: Metrics, methodologies and guidelines. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 649–658. ACM, 2009.
24. G. Hart. The five w's of online help systems. http://www.geoff-hart.com/articles/2002/fivew.htm.
25. G. Hart. The five W's: An old tool for the new task of task analysis. *Technical communication*, 43(2):139–145, 1996.
26. F. Hohman, M. Kahng, R. Pienta, and D. H. Chau. Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE Transactions on Visualization and Computer Graphics*, 25(8):2674–2693, 2019.
27. D. Holton and D. Clarke. Scaffolding and metacognition. *International Journal of Mathematical Education in Science and Technology*, 37(2):127–143, 2006.
28. Y. Holtz and C. Healy. From data to viz | find the graphic you need, 2021. Library Catalog: www.data-to-viz.com.
29. K. J. Holyoak and P. Thagard. *Mental leaps: Analogy in creative thought*. A Bradford book. MIT press, Cambridge, MA, US; London, England, 1995.
30. T. Hug and N. Friesen. Outline of a microlearning agenda. *eLearning Papers*, 16, 2009.
31. S. Huron, S. Carpendale, A. Thudt, A. Tang, and M. Mauerer. Constructive visualization. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, page 433–442, New York, NY, USA, 2014. Association for Computing Machinery.
32. IBM Cognos Analytics. https://www.ibm.com/at-en/marketplace/business-intelligence, 2019. Accessed: 2019-03-14.

33. M. Kandlhofer, G. Steinbauer, J. Lassnig, W. Baumann, S. Plomer, A. Ballagi, and I. Alfoldi. Enabling the Creation of Intelligent Things: Bringing Artificial Intelligence and Robotics to Schools. In *IEEE Frontiers in Education Conference (FIE)*, 2019.
34. H. Kang, C. Plaisant, and B. Shneiderman. New approaches to help users get started with visual interfaces: Multi-layered interfaces and integrated initial guidance. In *Proceedings of the 2003 Annual National Conference on Digital Government Research*, dg.o '03, pages 1–6. Digital Government Society of North America, 2003.
35. H. Kang, C. Plaisant, and B. Shneiderman. New approaches to help users get started with visual interfaces: Multi-layered interfaces and integrated initial guidance. In *Proceedings of the 2003 Annual National Conference on Digital Government Research*, dg.o '03, pages 1–6. Digital Government Society of North America, 2003.
36. D. A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann. Mastering the information age—solving problems with visual analytics. *Eurographics*, 2010.
37. D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. Visual Analytics: Scope and challenges. In S. J. Simoff, M. H. Böhlen, and A. Mazeika, editors, *Visual Data Mining*, LNCS 4404, pages 76–90. Springer, Berlin, 2008.
38. C. Kelleher and R. Pausch. Stencils-based tutorials: Design and evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 541–550. ACM, 2005.
39. Keshif | Data Made Explorable. https://www.keshif.me/. Accessed: 2020-04-15.
40. H. J. Klein, B. Polin, and K. L. Sutton. Specific onboarding practices for the socialization of new employees. *International Journal of Selection and Assessment*, 23(3):263–283, 2015.
41. D. Kolb. *Experiential Learning: Experience As The Source Of Learning And Development*, volume 1. Prentice-Hall, 1984. Journal Abbreviation: Journal of Business Ethics Publication Title: Journal of Business Ethics.
42. A. Kumar. Why user onboarding is essential for your application?, 2017.
43. B. C. Kwon and B. Lee. A comparative evaluation on online learning approaches using parallel coordinate visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '16, pages 993–997. ACM Press, 2016.
44. T. Lammarsch, W. Aigner, A. Bertone, S. Miksch, and A. Rind. Towards a concept how the structure of time can support the visual analytics process. In S. Miksch and G. Santucci, editors, *EUROVA*, pages 9–12. Eurographics, Eurographics, 2011.
45. G. E. Marai. Visual scaffolding in integrated spatial and nonspatial analysis. In *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association, 2015.
46. R. E. Mayer. The search for insight: Grappling with Gestalt psychology's unanswered questions. In R. J. Sternberg and J. E. Davidson, editors, *The nature of insight*, pages 3–12. MIT Press, Cambridge, MA, US, 1995.
47. Microsoft Power BI. https://powerbi.microsoft.com/de-de/, 2019. Accessed: 2019-03-14.
48. T. Munzner. A nested process model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics (InfoVis '09)*, 15(6):921–928, 2009.
49. O. Ola and K. Sedig. Health literacy for the general public: Making a case for non-trivial visualizations. *Informatics*, 4(4):33, 2017.
50. S. Palmiter and J. Elkerton. An evaluation of animated demonstrations of learning computer-based tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, pages 257–263. ACM, 1991.
51. J. Peeck. Enhancing graphic-effects in instructional texts: Influencing learning activities. In W. Schnotz and R. W. Kulhavy, editors, *Advances in Psychology*, volume 108 of *Comprehension of Graphics*, pages 291–301. North-Holland, 1994.
52. M. Pohl, M. Smuc, and E. Mayr. The user puzzle—explaining the interaction with visual analytics systems. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2908–2916, 2012.
53. QlikTech QlikView. https://www.qlik.com/us/products/qlikview, 2019. Accessed: 2019-03-14.
54. W. Ribarsky and B. Fisher. The human-computer system: Towards an operational model for problem solving. In *Proc. Hawaii Int. Conf. System Sciences (HICSS)*, pages 1446–1455, 2016.

55. S. Ribecca. The data visualisation catalogue. https://datavizcatalogue.com/, 2020. Accessed: 2020-03-12.
56. A. Rind, M. Wagner, and W. Aigner. Towards a structural framework for explicit domain knowledge in visual analytics. In *IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, pages 33–40, 2019.
57. J. C. Roberts, P. D. Ritsos, J. R. Jackson, and C. Headleand. The explanatory visualization framework: An active learning framework for teaching creative computing using explanatory visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):791–801, 2018.
58. P. Ruchikachorn and K. Mueller. Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE Transactions on Visualization and Computer Graphics*, 21(9):1028–1044, 2015.
59. D. Sacha, A. Stoffel, F. Stoffel, K. Kwon, Bum Chul, G. Ellis, and D. A. Keim. Knowledge Generation Model for Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1604–1613, Dec. 2014. 00000.
60. SAP Lumira. https://saplumira.com/, 2019. Accessed: 2019-03-14.
61. SAS JMP. https://www.jmp.com/en_us/software/data-analysis-software.html, 2019. Accessed: 2019-03-14.
62. SAS Visual Analytics. https://www.sas.com/en_us/software/visual-analytics.html, 2019. Accessed: 2019-03-14.
63. R. K. Sawyer and J. G. Greeno. Situativity and learning. In M. Aydede and P. Robbins, editors, *The Cambridge Handbook of Situated Cognition*, pages 347–367. Cambridge: Cambridge University Press, 2009.
64. A. Schmidt. Microlearning and the Knowledge Maturing Process: Towards Conceptual Foundations for Work-Integrated Microlearning Support. In T. Hug, M. Lindner, and P. A. Bruck, editors, *Proceedings of the 3rd International Microlearning 2007 Conference*, Innsbruck, Austria, June 2007. Innsbruck University Press.
65. H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A systematic view on data descriptors for the visual analysis of tabular data. *Information Visualization*, 16(3):232–256, 2017.
66. P. Shah, E. G. Freedman, and I. Vekiri. The comprehension of quantitative information in graphical displays. In *The Cambridge Handbook of Visuospatial Thinking*, Cambridge Handbooks in Psychology, pages 426–476. Cambridge University Press, 2005.
67. C. Stoiber, F. Grassinger, M. Pohl, H. Stitz, M. Streit, and W. Aigner. Visualization onboarding: Learning how to read and use visualizations. In *IEEE Workshop on Visualization for Communication*. IEEE, 2019.
68. J. Swabish and S. Ribecca. The graphic continuum: A poster project for your office. https://policyviz.com/2014/09/09/graphic-continuum/, 2014. Accessed: 2020-03-25.
69. Tableau Software Tableau. https://www.tableau.com/products/desktop, 2019. Accessed: 2019-03-14.
70. Y. Tanahashi, N. Leaf, and K. Ma. A study on designing effective introductory materials for information visualization. *Computer Graphics Forum*, 35(7):117–126, 2016.
71. J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE, 2005.
72. TIBCO Jaspersoft. https://www.jaspersoft.com/, 2019. Accessed: 2019-03-14.
73. TIBCO Spotfire. https://www.tibco.com/products/tibco-spotfire, 2019. Accessed: 2019-03-14.
74. C. Tominski. Event-based concepts for user-driven visualization. *Information Visualization*, 10(1):65–81, Jan. 2011.
75. J. J. van Wijk. The value of visualization. In *Proc. IEEE Visualization (VIS 05)*, pages 79–86, 2005.
76. X. Wang, D. H. Jeong, W. Dou, S.-W. Lee, W. Ribarsky, and R. Chang. Defining and applying knowledge conversion processes to a visual analytics system. *Computers & Graphics*, 33(5):616–623, Oct. 2009.

77. Z. Wang, L. Sundin, D. Murray-Rust, and B. Bach. Cheat sheets for data visualization techniques. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.
78. J. B. Watson. *Behaviorism*. Taylor & Francis, 1st edition, 1997.
79. M. A. Yalçın. *A Systematic and Minimalist Approach to Lower Barriers in Visual Data Exploration*. Dissertation, University of Maryland, College Park, 2016.
80. Z. Zhang, B. Wang, F. Ahmed, I. V. Ramakrishnan, R. Zhao, A. Viccellio, and K. Mueller. The five Ws for information visualization with application to healthcare informatics. *IEEE Transactions on Visualization and Computer Graphics*, 19(11):1895–1910, 2013.

Chapter 7
Adaptive Visualization of Health Information Based on Cognitive Psychology: Scenarios, Concepts, and Research Opportunities

Tobias Schreck, Dietrich Albert, Michael A. Bedek, Karl Horvath, Klaus Jeitler, Bettina Kubicek, Thomas Semlitsch, Lin Shao, and Andrea Siebenhofer-Kroitzsch

Abstract Consumer Health Information Systems (CHISs) are indispensable in healthcare. User-centered evidence-based medical information for patients positively influences therapy success, behavior, and cause–effect comprehension. Also, improved health literacy allows patients to accept medical advice and share decision-making and improves doctor–patient communication. Today, CHISs exist in many different forms. Yet, information is generally provided statically, i.e., the same medical content is presented to everyone. However, patients vary regarding previous knowledge and information needs and preference of perception of the

T. Schreck (✉)
Graz University of Technology, Graz, Austria
e-mail: tobias.schreck@cgv.tugraz.at

D. Albert · M. A. Bedek · B. Kubicek
University of Graz, Graz, Austria
e-mail: dietrich.albert@uni-graz.at; michael.bedek@uni-graz.at; bettina.kubicek@uni-graz.at

K. Horvath · K. Jeitler · T. Semlitsch
Medical University of Graz, Graz, Austria
e-mail: karl.horvath@medunigraz.at; klaus.jeitler@medunigraz.at; thomas.semlitsch@medunigraz.at

L. Shao
Graz University of Technology, Graz, Austria

Fraunhofer Austria Center for Data Driven Design, Graz, Austria
e-mail: l.shao@cgv.tugraz.at

A. Siebenhoferr-Kroitzsch
Medical University of Graz, Graz, Austria

Goethe University Frankfurt am Main, Frankfurt, Germany
e-mail: andrea.siebenhofer@medunigraz.at

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_7

information, e.g., in textual or visual form. This variation can depend, e.g., on gender, age, personality, perception, and cognitive aspects.
In this conceptual chapter, we envision how research and knowledge from evidence-based medical knowledge, cognitive-psychological mechanisms, and interactive data visualizations can be combined, to form adaptive and interactive consumer health information systems (CHISs) that take account of individual health information needs and increase health literacy by providing a reliable source of medical knowledge. To this end, we detail the scope and contributions of these disciplines to novel visual health information systems which can adapt them to the information needs and preferences of their consumers. We depict a concept for an advanced interactive, adaptive, personalized visual CHIS (named A^{+}CHIS). The concept is based on introducing multidimensional adaptivity in the content, visual presentation, level of detail, for example, to the provision of evidence-based medical health information, aiming at the consumers' full understanding of the meaning of the provided medical content. We argue that adaptive visual health information may provide efficiency increase for the general medical system and improved health literacy. While we do not present concrete results, we lay out the research opportunities and a possible system architecture to inform and implement A^{+}CHIS in the future.

7.1 From Static to Adaptive Visual Health Information Systems

Evidence-based medical research shows that health literacy and treatment success require that quality-assured medical information is available to the population and patients [21]. This is the aim of CHIS and providers, ranging from information folders, brochures, to media reports, web-based discussion forums, and information portals. Currently, information is presented *statically* and does not take into account that the knowledge, information needs, and health treatment situations (or contexts) of consumers differ significantly. Usually, a one-size-fits-all approach is taken to provide information, and existing CHISs provide health information unidirectionally from the system to users. However, advances in information technology enable consumers and patients to view health information on the Internet. In addition, consumers can record data related to their own health, e.g., by using consumer health trackers. By automatically collecting such data, as well as other forms of user preference feedback, and based on approaches of data analysis, we can automatically predict specific health interests of users and *dynamically adapt* health information to a particular consumer's context and requirements. Adaption can be made regarding the content and its level of detail, as well as its visual representation, e.g., as text, diagrams, interactive data visualizations, etc. Suitable adaption of information detail and form of presentation to customers' requirements, medical and visual literacy, and cognitive setting and greatly improve the reception and understanding of information [26, 47]. However, existing static CHISs do not fulfill these functions. Also, trustworthy health information targeted to the information needs of consumers and patients and effective interactive visual representations of health information are expected to drastically increase the use and acceptance of health information. This, in turn, may substantially improve patient–doctor communication and therapy compliance, foster general health literacy, and raise the effectiveness of the health system as a whole.

Hence, we motivate the need for multidimensional adaptive mechanisms for CHIS and propose to research techniques for advanced adaptive, human-centered, interactive and visual CHIS, or A^+CHIS. Advanced approaches should aim at delivering the required information to consumers at a level of detail and using the visual representations that best fit the consumer's specific, individual information needs.

We argue that results from three fields of research can be combined to research effective new approaches for adaptive, personalized health information: (1) interactive, personalized visualization of (2) evidence-based health information, and (3) fostering cognitive processing. In combination, they allow to research how novel A^+CHIS can be designed, implemented, and evaluated. It can support to define mechanisms to improve medical information processing in different patient and user groups. Novel visual interactive techniques that help adapt displayed information to suit the user context can be developed. The specific roles of these fields can be described as follows. A more detailed discussion of these areas and their contribution toward adaptive visual health information will be given in Sect. 7.3 and following.

7.1.1 Interactive Data Visualization for Health Data Visualization

In visualization, the goal is to find cognitively useful visual representations of data that enable us to understand complex data and make insightful decisions [70, 112]. Information visualization techniques are interactive, allowing to select, filter, and navigate data to support task-oriented data analysis such as exploration and hypothesis generation, comparisons, pattern searching, and the verification of dependencies [116]. Different visualization techniques have been proposed, depending on the type of data (e.g., time series, geospatial data, textual, or high-dimensional data), and application domains, including medical information.

The field of Information Visualization can provide concepts and implementations of adaptive visualizations of health data, building on evidence-based health information and cognitive psychology principles in health information use. Specifically, approaches may be developed to determine appropriate interactive visual representations of health information automatically. This requires, relying on Cognitive Psychology, finding appropriate knowledge structures for the representation of information on health, the consumers, and visualization processes. Building on this, research can be conducted into mechanisms that automatically decide what information, by which level of detail, and visual representation should be displayed based on the personal needs of the consumer. This can build on known guidelines and user studies on the effectiveness of information visualization displays and methods of machine learning and recommender systems to model and learn user interest. The latter will allow not only for consumer information according to the pull principle (consumer explicitly requests information) but also the push principle (further information is recommended).

7.1.2 Evidence-Based Consumer Health Information as Information Basis

The field of Evidence-based Consumer Health Information covers the study of existing CHIS, deriving principles, and goals used in traditional (static, non-interactive) CHIS, on which we can build on. These principles can inform new approaches, by incorporating known working methods, while extending and improving them for multidimensional adaptivity.

Based on existing quality criteria for CHIS, one can define standards for the methodological quality of the A^+CHIS to ensure their trustworthiness. A strictly evidence-based approach is required to ensure the presented information is of high quality [31, 67]. This is clearly necessary in view of the heterogeneity and differing levels of evidence presented in the medical literature, let alone the questionable quality of information available in the wilds of the Internet.

7.1.3 Cognitive Psychology Principles for Adaptive Health Information

Cognitive Psychology of adaptive health information systems is a new research direction, which can cover cognitive aspects involved when consumers seek and process health information. A key role in this research is the study of pre-knowledge, motivation, interests, cognitive biases, and expectations that influence the most suitable quantity, detail, context, and presentation of information for specific consumers. Adaptation mechanisms needed to suit individual consumer profiles can be defined. Research into Cognitive Psychology, including such aspects as the identification of mechanisms of knowledge, motivation, and learning capacity, which takes into account that human cognition is vulnerable to many known cognitive biases and misconceptions, can inform the mechanisms for novel A^+CHIS. As far as health information is concerned, this can result in problems such as over-information or over-diagnosis, which should be taken into consideration by these systems.

The remainder of this chapter is structured as follows: Sect. 7.2 emphasizes the importance of a novel A^{+}CHIS on diabetes type II and describes two scenarios on how such an A^{+}CHIS could provide suitable health information to consumers. Sections 7.3 to 7.5 provide a review of the previous work and current research challenges in the three abovementioned fields of research, interactive visualization, evidence-based health information, and cognitive psychology. Section 7.6 outlines a foreseen system architecture, and Sect. 7.7 concludes.

7.2 Scenario: Adapting Health Information for Diabetes Type II

In the following, we exemplify the need for adaptive visual health information by an important disease according to two scenarios. The examples show the different needs and phases in the course of a disease and serve as a reference for building adaptive health information systems. We select type 2 diabetes mellitus for our use case because it is highly relevant to public health, affects a broad section of the population, and is hence well suited for the evaluation of the effects of A^{+}CHIS. The type of information required varies greatly and includes general information on the disease, information on the use of devices for blood glucose monitoring or insulin applications, information on the potential advantages and disadvantages of various interventions and expected effect sizes, information on adequate footwear for persons at increased risk of amputation, information supporting behavioral changes, and specific advice.

Depending on the specific situation and interest, a lower or higher level of information detail is advantageous. Type 2 diabetes mellitus is a chronic disease, i.e., affected persons must deal with it over a long period of time. With increasing age and disease progression, the patient may require or be interested in different information. For example, information on prevention and lifestyle measures may be most important at the onset of the disease, while information dealing with secondary diseases may gain in importance later on. Furthermore, the disease affects people of different social and educational backgrounds and interests, which may affect the information they are interested in.

To be effective, health information must be adapted to suit the current needs of individual consumers. But currently, the detail and comprehensibility of health information on type 2 diabetes mellitus, whether in paper form or taken from the Internet, etc., cannot generally be influenced. Furthermore, it is often difficult for consumers to obtain the information they need, when they need it, and in a form they can understand. For much information, it is also impossible for patients to determine to what extent it is reliable. New A^{+}CHIS should offer comprehensive evidence-based information at varying levels of detail (from general information, specific questions, descriptions of the expected clinical effects of interventions, explanations of pathophysiological mechanisms to original scientific papers) and take into account individual social and educational differences. Thereby, they can make major contributions to improving the care of people with type 2 diabetes mellitus and improving health literacy.

The specification of scenarios is a useful domain context to inform the adaptation mechanisms of an adaptive CHIS system. In Sect. 7.6, we discuss an architecture and possible Machine Learning methods to incorporate scenario specifications.

7.2.1 Consumer Health Information System Scenario 1

A person with known type 2 diabetes mellitus goes to the doctor for a routine check-up. The doctor determines that the patient's blood glucose has risen since it was last checked and is now too high. The doctor recommends intensifying the blood glucose-lowering therapy. As a key information need, the questions arise *to what extent blood glucose should be lowered* and *what the potential benefits and harms of intensified blood glucose reduction are*. Based on this scenario, information content can be adapted and visually presented like exemplified in the following.

Detail Level I: Basic Information on Recommended Intensity of Blood Glucose Reduction A^+CHIS will provide information based on recommendations published in Austrian Diabetes Association and other international guidelines, i.e., the intensity of the blood glucose reduction should, in principle, be determined individually for each person, based on defined criteria. Furthermore, A^+CHIS will provide a list of the corresponding personal criteria that includes examples for better comprehensibility. For example, when considering the criterion of 'significant comorbidities', intensive glucose lowering is recommended if these comorbidities are absent or of low severity. However, if there are numerous or severe comorbidities, moderate glucose lowering is recommended. A^+CHIS will show examples of concomitant diseases (cancer, heart attack, stroke, etc.) which are considered relevant to the given consumer. The A^+CHIS relies on consumer profile information and explicit/implicit feedback from the consumer to determine what information is necessary. Furthermore, the way the information is presented gets adapted, e.g., to overview or summary texts, or symbolic representations based on cognitive profiles of consumer's preferences and perceptions, which are also continuously maintained by the A^+CHIS.

Detail Level II: Basic Information on the Benefits and Harms That May Result from Intensifying Blood Glucose-Lowering A^+CHIS will explain that data from scientific studies show that, e.g., compared to a moderate reduction in blood glucose, intensified blood glucose-lowering does not reduce mortality, prevents a heart attack in 2 out of 1000 people, prevents a microvascular event (e.g., new onset or progression of retinopathy) in 14 out of 1000 people, etc. At the same time, the studies show that for every person in whom a microvascular complication was prevented by intensification of therapy measures, there are three persons in whom severe hypoglycemic events can be expected. The A^+CHIS selects a visual representation appropriate to the interest and visual and health literacy of the consumer, e.g., using aggregated or disaggregated medical statistics, which may involve a diagrammatic representation of probabilities such as Venn diagrams, line charts, or symbolic representations.

Detail Level III: Presentation of the Content of Scientific Studies on the Extent of Potential Benefits or Harms of Intensifying Blood Glucose Reduction A^{+}CHIS will provide information on key publications, guidelines, etc. in terms of their research questions, content, and methodological quality. The exact research question, the inclusion criteria, and the number and characteristics of included studies are presented. In addition, a detailed presentation of the results and the authors' conclusions/recommendations is provided. The A^{+}CHIS chooses, based on consumer profiles, appropriate visual representations of documents. For example, the text of a document can be reduced or expanded, applying natural language processing methods to the recommended level of detail (see Sect. 7.6). It may be possible to represent medical content in different textual or visual form, e.g., to represent the dependencies of health on medication and behavior as either a dependency graph (network), or in textual form, or both. In the systems, consumers should be able to interactively mark information that interests them. This feedback is used to update the consumer profile database and search for related documents that may then be recommended. Based on demand for further information, the system could respond by giving additional references in key publications or guidelines that are relevant to the topic and corresponding links to the publications.

In terms of form and visualization and, where reasonably possible, depending on specific aspects and characteristics (such as gender and age), the content is presented differently for each level of detail. The presentation will also take into account findings on the avoidance of cognitive bias. It should ensure that content is evidence-based. In addition, based on the described information, A^{+}CHIS should present additional information on the significance of hypoglycemic events at differing degrees of detail and using different forms of presentation. Likewise, information on the possible benefits and harms of therapeutic measures (lifestyle, medication, etc.), i.e., information on the question "How should intensification in blood glucose-lowering take place?", can be visualized by means of a network of topics and the relationships between them. These can then be interactively explored by the consumer.

The A^{+}CHIS chooses from a spectrum of visualization techniques for the information, including standard diagrams and maps, symbolic representations, text visualizations like tag clouds, keyword timelines, document landscapes, citation graphs, etc. The visualizations are interactive and enable consumers to request additional information and indicate interest or disinterest in certain aspects, from which the system learns and updates the consumer profiles.

From a cognitive psychological point of view, each consumer's information processing and integration must be analyzed and optimized, taking into account (a) relevant previous knowledge, health literacy, and the user's profile, including such details as age, gender, consumer group (e.g., patient, relative, interested laypersons), and information needs, (b) her/his aims and goals associated with information needs, and (c) all information on the consumer gained from his/her dynamic interactions and provided feedback that has contributed to sustainable non-biased knowledge. Optimization is achieved by using the process-oriented formative and summative evaluation of multidimensional adaptivity and interactivity.

7.2.2 Consumer Health Information System Scenario 2

A person without known diabetes mellitus participates in a routine health exam and an elevated blood glucose level is detected. Many questions on diabetes mellitus arise for the person, e.g., *Do I really have diabetes mellitus?* And if so, *What type of the disease do I have? What consequences will this have for my health and professional future?* And *What can I do to reduce the risk of unwanted health consequences?* To answer such questions, information on the diagnosis and criteria used to establish the diabetes type, the expected prognosis, and legal aspects are necessary. In addition, information on possible types of therapy (lifestyle, medication), necessary therapy intensity, and any psychological and social support is required. The potential benefits and harms of the various interventions (comparative effects of different drugs, of blood pressure reduction versus blood sugar reduction, etc.) must also be explained. The need for information will depend on the person concerned and may depend on the severity of the disease, possible concomitant diseases, whether the person is employed or not, gender and age, available resources, etc. The A^+CHIS recognizes the information needs and adapts the health information presentation in terms of the level of detail of the provided information, content, gender aspects, type of visualization or form of presentation, to meet specific needs. It will be ensured that the content is evidence-based. A key requirement of the A^+CHIS is that the presentation should take into account possible cognitive biases of its consumers, which requires tracking appropriately defined and maintained consumer profiles over the repeated uses. As an example, if a consumer primarily and repeatedly searches for information units about symptoms, possible side effects and complications of type 2 diabetes, and its medical treatment, which can be interpreted as confirmation bias, A^+CHIS could suggest information units about protective behavioral strategies as de-biasing strategy and to increase the consumers' self-efficacy [64]. To this end, cognitive psychology principles of health information need to be researched and reflected in the rules the A^+CHIS applies for adaptive health information presentation.

7.3 Visual Health Information and Visual Analytics for Healthcare

According to various studies, existing CHISs lack in readability and suitability [46, 57, 83, 93, 107, 114, 117]. In our vision, medical information should be visualized such as to be understood by consumers and serve their individual information needs [18], taking into account evolving information needs and health and visual literacy. Medical information comprises heterogeneous data types, e.g., textual descriptions of symptoms and treatments, networks of cause–effect relationships, time series data to show measurements, and other numeric data quantifying medical relationships, including uncertainties. Using indirect and direct human-computer interaction approaches, we can enable to infer automatically information consumers require. Dynamically updated user profiles can support to retrieve relevant medical information from appropriately structured knowledge databases. Based on knowledge on the effectiveness of different information visualization techniques for the satisfaction of different user interests and cognitive properties, we can present this information in interactive visualizations. The level of detail and visual presentation should be automatically tailored to the respective user's needs.

To this end, we can rely on information visualization (specifically, scalable effective techniques for the display of different types of data), user interaction techniques (specifically, graphical user interfaces with direct and indirect interaction modalities), and knowledge technologies (specifically, knowledge representation, recommender systems, and classification techniques). An A^+CHIS may rely on direct interaction modalities, like question-asking, or selection. It may potentially also rely on indirect ones, e.g., interaction log data analysis to predict interest and literacy. With this information, it can adapt the visual information display to the consumer.

Existing adaptation and recommender systems often control only one information modality (e.g., by recommending different products to consumers) or are focused on a single information presentation format (e.g., texts or audio/video files). A^+CHIS should simultaneously decide *what* to present, *how* to present it, and to *observe* how users interact with presented information. Based on this, we may maintain consumer profiles by adapting them to changing information needs over time.

7.3.1 Previous Work

7.3.1.1 Interactive Data Visualization and Health Data Visualization

In visualization, the goal is to find cognitively useful visual representations of data that enable us to understand complex data and make insightful decisions [70, 112]. Information visualization techniques are interactive, allowing to select, filter, and navigate data to support task-oriented data analysis such as exploration and hypothesis generation, comparisons, pattern searching, and the verification of dependencies [116]. Different visualization techniques have been proposed, depending on the type of data (e.g., time series, geospatial data, textual, or high-dimensional data) and application domains, including medical information. To date, effective visualization techniques have been proposed. For example, the LifeLine system was among the first to visually represent patient treatment histories and support interactive exploration [80]. Electronic health records enable novel visualization applications for patient data [86]. The KAVAGait approach [110] helps doctors inspect complex data derived during clinical gait analysis and supports diagnoses and patient treatment decisions. In [22, 33, 118], timeline-based visualization techniques are used to display patients' pathways from a clinical point of view, e.g., patient flow, summaries of individual periods in treatment, or treatment plans for diabetes patients. Also, icon-based and radial layout-based visualizations have been explored to visualize multidimensional health record data [19, 24]. An A^+CHIS should rely on such visualization approaches as a basis to proactively choose and adapt to the specific information needs, including consumers who are not educated information or interaction experts.

7.3.1.2 Visual Abstractions and Visual Literacy

Understanding advanced data visualizations and visual interactions depends, among others, on the visual literacy of a user. Visual literacy can be defined as the ability to recognize and understand ideas conveyed through visual representations (visible actions, symbols, or images) [1]. Previous work has focused on design choices and visual interactions aimed at making exploratory data analysis more comprehensible to novice users. In [87], a concept of teaching and learning unfamiliar visualizations by analogy was proposed that uses transformative morphing to explain unfamiliar visualizations by linking them to more familiar ones. In [20], a design space for storytelling with timelines was introduced that characterizes 14 different design choices along three dimensions: representation, scale, and layout.

VisGuides[1] is a discussion platform that collects visualization guidelines and allows expert discussion on guidelines and respective empirical results. Depending on the type of abstraction, cognitive load may occur. As an example, different visual abstraction methods for scatter plot diagrams, for example, include density-based [29, 63], cluster-based [58], and regression-based [92] abstractions, which convey different properties of data in scatter plots. Again, these are valuable approaches to convey health information, and an A^+CHIS should pro-actively choose, adapt, and present these base visualization techniques to adapt to different customer information needs and information processing abilities including cognitive properties and possible cognitive biases.

[1] VisGuides Forum on Visualization Guidelines. https://visguides.org/ (accessed July 11, 2020).

7.3.1.3 Adaptive Visualization for General and Medical Data

An essential aspect of data visualization is enabling insight into data. However, a key factor is that not all users have the same knowledge and understanding of visual data representations (see also Sect. 7.5.1.1). Two different user characteristics exist: long-term user characteristics (e.g., cognitive abilities and expertise) and short-term characteristics (e.g., cognitive load and attention). Both should be considered when designing information visualizations [105]. To increase the general effectiveness of visualizations, they should be adapted to users' individual visualization needs and abilities. Studies based on user characteristics, such as perceptual speed, verbal working memory, visual working memory, and user expertise, have been conducted to assess the effectiveness of visualization types [105]. A key challenge in adaptation is to do it automatically. Indirect interaction modalities like eye tracking can potentially be used. In [100], information on users' eye gaze patterns is used to predict user visualization needs. These may help to adapt the visualization to suit the identified task. Recent work applies data analysis to low-level user interaction signals. For example, in [76], a hidden Markov model is used to derive learning interest and predict relevant information items in a visualization developed from user interaction data. These are interesting approaches to adapting visualization but require careful set-up of user models and tailoring to specific domains. We aim to use both codified domain and user knowledge, as well as feedback loops, to develop user interest models and apply them to the adaptation of information displays.

In medical applications, adaptive visualizations may provide considerable benefits by optimizing insight into, e.g., medical histories, patient observations, lab results, clinical findings, etc. AdaptiveEHR [51] is a context-based framework that uses biomedical knowledge structures (ontologies) and graphical disease models to generate a tailored presentation of patient records based on patient information needs. In [69], adaptive visual symbols are presented to visualize personal health records and to summarize a patient's medical history with the desired complexity. An adaptive A+CHIS could include several health data visualizations as the basis. For instance, symbolic representations of medical events [22, 33], fact sheets [103], or visual depiction of a hierarchy of diseases [24] could be adapted for individual consumers and their needs. Furthermore, other visualization techniques like Sankey diagrams, generic network representations, or time-oriented visualizations could be used to support quantitative data analysis of medical data [45, 82, 111].

7.3.1.4 Knowledge Technologies and Medical Health Information

Integration of information retrieval systems in the form of a medical question-answering system [41, 73], or an intelligent chatbot [74, 84], can be valuable for adaptive systems. Chatbots [98] engage patients in a conversation about medical information needs. Also, question-answering systems can be useful to identify similar patients, patterns of diseases, and successful treatments and to provide specific answers to questions. Both question-answering systems and chatbots can be applied to ontologies to take queries expressed in natural language and return answers drawn from available semantic information [4, 59].

Knowledge techniques often rely on structured databases, with semantic information being associated with a domain, so that it can be processed automatically and without human intervention. Ontologies can represent knowledge as a set of concepts within a domain and to define relationships between the concepts. Different ontologies exist in the medical domain, e.g., the Unified Medical Language System,[2] the Bioportal Repository of Biomedical Ontologies,[3] the Disease Ontology,[4] or the OLS Repository for Biomedical Ontologies.[5]

In [89], an ontology was used to analyze information from online healthcare forums. The approach reveals the relationship between patient profiles and health-related terms extracted from their forum messages.

The ontology captures such patient profile data as age, gender, ethnicity and habits, and health-related information like diseases, side effects, and symptoms. In [30], an ontology-based model for diabetic patients is presented to aid doctors in diagnostic decision-making. An overview of the application and effectiveness of ontologies in e-Health applications is given in [43]. An A$^+$CHIS can rely on such ontologies to adapt the health information presentation, for example, by aggregating or expanding the level of detail of presented information to consumer information needs.

7.3.2 Research Challenges

In the course of a new research project by the co-authors, we aim to develop, implement, and evaluate novel A$^+$CHIS aiming to address the above motivated requirements. We specify the following guiding research challenges and questions for information visualization for A$^+$CHIS as follows:

- How can health information, including dependencies between health, preconditions, treatment, and behavior, be effectively visualized at different levels of abstraction, regarding both the information content and presentation form?
- Which direct and especially indirect feedback mechanisms are effective in recognize the evolving medical interest and visual literacy of consumers?
- How can the profiles of consumers be updated to reflect evolving interest and literacy?

[2] Unified Medical Language System. https://www.nlm.nih.gov/research/umls/Snomed/snomed_browsers.html (accessed July 11, 2020).

[3] Bioportal Repository of Biomedical Ontologies. https://bioportal.bioontology.org/ (accessed July 11, 2020).

[4] Disease Ontology. https://disease-ontology.org/ (accessed July 11, 2020).

[5] OLS Repository for Biomedical Ontologies. https://www.ebi.ac.uk/ols/index (accessed July 11, 2020).

- How can guidelines for effective visualizations be compiled in a knowledge database for adaptation of visual information?
- How can knowledge from different sources (health information, consumer profiles, and visualization guidelines) be efficiently compiled into knowledge databases with semantic representations?

7.4 Evidence-Based Health Information and Systems

CHIS has the task of providing laypersons with a comprehensive overview of diseases and thus increasing the health literacy in the population. Research addresses the question of what kind of CHISs are currently available internationally, how they are structured, how the medical content will be presented to consumers, and how far these aspects can contribute to the development of a new advanced, adaptive, and interactive consumer health information system (A^+CHIS).

For the development and testing of adaptive health information systems, real-life scenarios are required. We focus on diabetes mellitus type 2, as it is very relevant to the public health and affects a wide section of the population [54]. Hence, A^+CHIS would deliver great benefits in this area. Diabetes mellitus type 2 is a chronic disease, i.e., affected persons are confronted with it over a long period of time [10]. Therefore, the type of information required varies greatly depending on the specific situation and interest. With the progression of the disease and age, different information is necessary. Evidence-based medical data can be prepared with regard to different levels of detail and the individual needs of the users. Then, the information can be combined with new visualization concepts and techniques as well as with cognitive-psychological research to enable an interactive adaptive system to present the right information in the most appropriate form.

7.4.1 Previous Work

7.4.1.1 Health Literacy

One of the cornerstones of patient charters, e.g., the Austrian Patient Charter,[6] is the right to be informed about one's own health or illness. This information can only contribute to strengthening health literacy and promoting informed decision-making if it is comprehensive and understandable.

In the Health Literacy Survey-Europe (HLS-EU) project in 2011, health literacy was defined as "people's knowledge, motivation, and competences to access,

[6] Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz. Patientenrechte. https://www.gesundheit.gv.at/gesundheitsleistungen/patientenrechte/inhalt, 2020 (accessed July 23, 2020).

understand, appraise and apply health information in order to make judgments and take decisions in everyday life concerning health care, disease prevention and health promotion to maintain or improve quality of life throughout the course of life" [96].

The HLS-EU project was conducted among eight European countries (Austria, Bulgaria, Germany, Greece, Ireland, the Netherlands, Poland, and Spain), according to which the health literacy of the Austrian population is lower than in other European countries. At 56%, the percentage of people with inadequate or problematic health literacy is higher than the international average (48%) [95]. Low health literacy is associated with poorer health outcomes, higher rates of hospitalization, greater use of emergency care, and higher rates of mortality in the elderly [17].

"Health competence" among individuals requires the ability to read and understand health information and to be able to interpret it and use it for one's own good. It is well known that health literacy in the elderly, in poorer people, and in those with little school education, is lower than in younger, well-educated, and well-situated persons [17].

7.4.1.2 Consumer Health Information Systems

Shared decision-making and adequate information on health issues are not only in the interest of patients [42, 68] but also a legal requirement.[7] CHIS are tools that are commonly used to support informed decision-making.

Patient information is available from many sources, and its purpose is to provide patients with a comprehensive picture of their disease. This information should help patients understand their symptoms and develop a sense of not only benefits, risks, and side effects but also useless or even harmful interventions [88]. Existing CHISs aim at patient information and are particularly concerned with:

- General knowledge of health, diseases, their effects, and their courses
- Interventions to maintain health (prevention and health promotion)
- Early detection, diagnosis, treatment, palliation, rehabilitation, and follow-up care of diseases and associated medical decisions
- Care and coping with illness
- Daily life with an illness[8]

Health information can be provided in very different situations, for various target groups, and in a wide range of formats. This includes not only written information (in printed and digital form) but also audio and video formats and apps for mobile phones. Dynamic Internet formats such as interactive decision

[7] Bundesministerium für Gesundheit und Frauen. Gesundheitsziele Österreich - Richtungsweisende Vorschläge für ein gesünderes Österreich (Langfassung). https://gesundheitsziele-oesterreich.at/website2017/wp-content/uploads/2018/08/gz_langfassung_2018.pdf, 2017 (accessed July 23, 2020).

[8] Deutsches Netzwerk Evidenzbasierte Medizin e.V. Gute Praxis Gesundheitsinformation. https://www.ebm-netzwerk.de/de/medien/pdf/gpgi_2_20160721.pdf, 2016 (accessed July 23, 2020).

aids, which are targeted at a specific decision-making process and often within the context of a treatment, are also included. As people differ in terms of their abilities, whether they prefer visual or auditory information, and in the health topics that interest them, it is essential that health information is individualized. However, current CHISs generally present content statically and do not take into account that previous knowledge, the need for information, and the individual situation of patients can vary. A "one-size-fits-all" approach to providing information is followed, and CHISs provide health information unidirectionally, i.e., information flows only from the system to users. Furthermore, the patients themselves are rarely actively involved in the development process of CHIS. A positive example is "Stiftung Gesundheitswissen," a German non-profit foundation that only prepares health information materials once peoples' needs have been identified.[9] It also delivers information in several different formats such as in text and graphic form and provides multimedia options such as reality and animated films. Furthermore, the same health information is presented in a variety of ways in order to spread the information as widely as possible. When searching for health information, one encounters a variety of web-based or written materials published by different organizations and individuals that are of different quality, accuracy, and reliability. This in turn poses significant challenges for the user in selecting sources and, in particular, in assessing the credibility and trustworthiness of these sources [90]. Although the use of online formats is increasing [12], according to a report of the situation prior to 2016, doctors were still the most important provider of health information. The U.S. National Trends Survey, which has studied changing communication trends and practices in cancer care for more than a decade, also reported that doctors are still a more reliable source of health information than online tools, health authorities, and brochures[10] [77]. As far as we know, no currently available media channel in the health field uses adaptive and interactive CHIS. One should therefore conduct a systematic review to identify the media sources that are used to provide medical knowledge to patients and find out whether interactive health information tools are yet in use.

7.4.1.3 Quality of CHIS

Regardless of whether individuals can understand and interpret information, it is essential that available information materials are evidence-based and reliable. However, the quality of health information has several dimensions. In addition to the correctness of the content, these include the up-to-dateness and completeness of content, as well as such aspects as readability, appropriate detail, presentation, and

[9] Stiftung Gesundheitswissen. https://www.stiftung-gesundheitswissen.de/ (accessed July 23, 2020).

[10] Health information national trends survey. https://hints.cancer.gov/about-hints/learn-more-about-hints.aspx (accessed July 23, 2020).

accessibility. Several methodological papers have therefore been written on how to develop high-quality health information materials and how to assess existing health information [27, 66]. Unfortunately, the content of health information is often driven by commercial interests and rarely presents a balanced view.

In the course of a study in Styria, over 1000 print versions of health information materials from general practices have been collected. All information materials had considerable shortcomings and did not provide the balanced, comprehensive, and comprehensible information that would help patients raise their health literacy and make informed decisions [50]. The same picture is true of online health information. As shown in a recent systematic review [36] that included 153 cross-sectional studies, the Internet is not a source of reliable health information for non-professionals that have no education in medicine.

7.4.2 Research Challenges

Based on the above requirements discussion, we specify the following guiding research challenges and questions for evidence-based health information for A^{+}CHIS as follows:

- What CHISs are currently available? What are their characteristics and how can they inform the development of an A^{+}CHIS?
  - Which types of CHIS and media are currently used in practice?
  - How are current CHIS structured, and what is their level of information detail?
  - What are the differences between CHISs? What is the range of different content, concepts, and other characteristics?
- How can it be ensured that the contents of the A^{+}CHIS are trustworthy and uncertainties are sufficiently communicated?
  - What instruments and criteria are used to assess the methodological quality of current CHIS? What best practice rules can be deduced for adaptive visual CHIS?
  - How can advanced, comprehensive, evidence-based CHIS be designed and evaluated for type 2 diabetes mellitus?
  - Which approaches can support the filling of health information systems with quality-assured evidence-based content?

7.5 Cognitive Psychology of Health Information

Adaptive health information systems should reflect the cognitive dimensions of health information consumption, underlying knowledge, comprehension, and processing. The main objective is to facilitate in-depth processing without cognitive biases and misconceptions and the effective and sustainable learning of information units by users of A^+CHIS. This should lead to desirable health-related behavior, such as improved communication between patients and medical doctors and increased compliance.

To achieve this objective, an approach would be to obtain requirements by user requirement analysis, e.g., by semi-structured interviews with patients and relatives. Well-established instructional design principles that facilitate comprehension and learning processes should hereby avoid the occurrence of cognitive biases and misconceptions. The adaptation of presented health information would rely on advanced multidimensional adaptive, personalized, and interactive mechanisms, including feedback loops (in both directions) between users and the A^+CHIS to increase knowledge, motivation, and health-related behaviors in a non-intrusive manner. Non-intrusive assessments of previous knowledge, information needs, and the motives of users should be carried out.

Previous work on knowledge representation, non-intrusive assessment, adaptive, personalized, and interactive mechanisms, cognitive bias mitigation in visualizations, and multi-method evaluation approaches will be instrumental to this end. Design guidelines and principles aimed at facilitating comprehension and an effective and sustainable learning processes on the one hand, and design guidelines for mitigating cognitive biases and misconceptions in visualizations on the other, have so far been considered to be separate research areas. We may overcome this separation by synthesizing these two areas. Existing solutions and research into adaptive, personalized, and interactive mechanisms can be combined in an innovative way, with the aim of attaining ideal adaptation. This process will also rely on summative evaluation studies that compare non-adaptive and different degrees and forms of adaptive and interactive A^+CHIS with regard to comprehension and effective and sustainable learning processes.

7.5.1 Previous Work

7.5.1.1 Knowledge Representation

Research on knowledge representation in Cognitive Psychology has a long tradition (e.g., [56]). Although some overlap exists with computer science (e.g., [79]), with both disciplines using the same formats (such as mathematical expressions and procedural codes), it is important to differentiate between the two. In this area, we will focus on cognitive-psychological approaches (e.g., [5]) which in CHIS have to meet the following requirements: (a) be suitable for representing both medical and user knowledge, (b) be able to represent knowledge from different perspectives because of the multiple, adaptive aims of A^+CHIS, and (c) ensure to visually represent medical and user knowledge in a meaningful and transparent way.

At least the Formal Concept Analysis (FCA), the Knowledge Space Theory (KST), and a set of graph-based knowledge representations meet these requirements. The FCA (e.g., [15, 113]) visualizes complex information as concept lattices without loss of information. The KST (e.g., [7, 37, 40]) provides a formal basis for simultaneously structuring a domain of knowledge and the knowledge of individuals that is based on prerequisite relations. Finally, graph-based knowledge representations visualize the inherent structure of meanings (e.g., [3, 8, 97, 106]).

Almost all of these approaches have been successful applied in the provision of medical information and in medical education: for instance, for application of FCA [81], of KST [6], or by Graph-based knowledge representations such as Concept Maps [34], Conceptual Graphs [53], and Mind Maps [2]. The main challenge of developing A$^+$CHIS for domain and user knowledge representation is to research how to provide multiple perspectives on the content simultaneously.

7.5.1.2 Adaptive Assessment

In view of the huge amount of medical information and the variety of consumer needs, a selective presentation will be necessary. User-centered adaptive assessments must guide the selection procedure to suit individual information needs. Based on the different formats of knowledge representation, adaptive assessment procedures rely on the transitivity of different underlying structures/relations. Typical examples refer to (i) difficulty [85], (ii) prerequisites of information units [37], (iii) subordinated meanings (hyponymy and hypernymy), (iv) preferences, or (v) graphical knowledge representations [101]. In the medical domain, some examples of adaptive assessments of knowledge have already been proposed (e.g., [48, 60, 78]).

However, research into multidimensional adaptive assessments is lacking. A second challenge is to gradually replace formative multidimensional assessments by using non-intrusive, indirect adaptive assessment procedures "in the wild" (i.e., self-regulated interactivity, see, for example, [94]). A third challenge is posed by the need for an indirect adaptive assessment of users' needs with regard to the properties of the system in order to improve the usage and acceptance of A$^+$CHIS.

7.5.1.3 Interactivity

A$^+$CHIS is conceptualized for adaptive, interactive, multimodal Human-Computer Interaction (HCI). Although they are the two sides of the same coin, we must distinguish between HCI in computer science and in psychology. HCI is also a topic in psychology for a long time already (see e.g. [25]). Instead of a data-driven, bottom-up approach, we favor theory-driven top-down approaches for analyzing interaction data and supporting the user with respect to Self-Regulated Behavior and Learning (SRBL) [72, 119], which is used as general framework for modeling the "interactive human behavior in the loops." Within the SRBL framework, more specific approaches focus on (i) Information Seeking and Retrieval Processes [91], (ii) Process-Oriented Feedback [55, 71], (iii) Motivation [13], and

(iv) Microadaptivity [99]. The challenge is to integrate the different approaches for supporting the consumer in using A^{+}CHIS interactively for reaching his/her goals and information needs in operating with the system.

7.5.1.4 Identification and Mitigation of Cognitive Biases

In situations that are uncertain and complex or in case of time constraints, individuals often apply heuristics, or "rules of thumb," when making decisions, or when evaluating the value, importance, and meaning of information. Although often useful, such heuristics can lead to severe and systematic judgment errors, referred to as cognitive biases. In the literature, a wide range of cognitive biases have been identified, such as the confirmation bias, the framing effect, anchoring, or the Bayes-rate fallacy. Even if the history of cognitive bias research originated in the late 1960s (e.g., [109]), the mitigation of cognitive biases in visualizations is a relatively new and emerging research topic. Some suggestions on how appropriate visualization techniques could mitigate the effects of the aforementioned cognitive biases have been described in [16, 38]. An example from the context of medical information is that even experts have difficulties estimating the risk of treatments when confronted with conditional probabilities [49]. However, such Bayes-rate fallacies can be easily mitigated by showing frequencies rather than probabilities [44]. Interactive visualization techniques possess even greater potential to mitigate certain cognitive biases than static visualizations since the information can be shown from different perspectives and with different levels of detail.

7.5.1.5 Instructional Design

This refers to the creation of information units that ensure comprehension as well as an effective and sustainable learning process of users. Cognitive psychology focuses on the learning process from an information processing perspective. The working memory and its capacity limitations [11, 32] play a major role, for example, in case of the cognitive theory of multimedia learning (CTML) [61, 62] and the cognitive load theory (CLT) [104]. The CTML defines a set of principles for the design of information units, such as the multimedia principle, which states that pictures should be accompanied by explanatory text (narration) and vice versa. The CLT distinguishes between three different types of cognitive load: (a) intrinsic cognitive load is caused by the learning task itself (e.g., statistical information), (b) Germane cognitive load refers to activities that are required to foster learning, such as schema construction, and (c) extraneous cognitive load refers to cognitive activities that are irrelevant to learning and should therefore be avoided to prevent cognitive overload among consumers of the information units.

7.5.1.6 Evaluation of Adaptive Systems

In the context of adaptive health information systems, we favor a multi-method approach to formative and summative evaluation that combines qualitative and quantitative methods and statistical analyzes, as well as explorative and (quasi-) experimental study designs. This multi-method approach toward empirical research is also reflected in [9] and [14]. Evaluation activities that are particularly challenging include (i) the collection and monitoring of user requirements, (ii) the holistic examination of the impact of instructional design guidelines and principles on user comprehension, learning processes, and the avoidance of misconceptions and cognitive biases, and (iii) the identification and validation of ideal adaptation by comparing prototypes that use different degrees of adaptivity and personalization [75] and differing interactive mechanisms.

7.5.2 Research Challenges

Based on the above requirements discussion, we specify the following guiding research challenges and questions for cognitive psychology for A^+CHIS as follows:

- What instructional design principles should be applied in the construction of information units to ensure comprehension and an effective and sustainable learning process? What cognitive biases and potential misconceptions are involved and/or evoked when interacting with A^+CHIS and how can they be detected and mitigated?
- What advanced aspects of personalization, multidimensional adaptation, and interaction should be implemented in A^+CHIS to ensure the comprehension and learning process is effective and sustainable? How can these aspects be improved?
- What cognitive and motivational processes are involved in interacting with the A^+CHIS in a self-regulated manner and how can users' previous knowledge, information needs, and motives be assessed in a non-intrusive fashion? What kind of feedback loops between users and the A^+CHIS (in both directions) can be used to increase knowledge, motivation, and health-related behaviors?
- What are the requirements, motives, and information needs of individual users and different user groups, such as patients and relatives, when using standard CHIS? Are existing evaluation methods suitable for comparing non-adaptive and different degrees and forms of adaptive and interactive A^+CHIS with regard to comprehension and learning processes, motives, and user behavior?

7.6 Architecture and Machine Learning Methods for an Adaptive Visual Consumer Health Information System

In the following, we devise a system architecture which can implement an A$^+$CHIS. It is based on the following core components and makes use of knowledge databases encoding the domain knowledge on health information, customer description including cognitive profiles, and visualization rules and guidelines. An adaptation engine predicts user interest, delivers visual information, and manages user profiles. We also discuss the key role of Machine Learning in implementing the adaptation engine.

7.6.1 Overview of Proposed Architectures

The **Adaptation Engine** (Fig. 7.1 center) drives the selection, adaptation, and presentation of appropriate health information to consumers, based on their information needs, expectations, previous knowledge, etc. The engine queries semantically structured information from three specific **Knowledge Databases** (KDBs). Each KDB uses appropriate semantic structures such as concept maps and ontologies to store facts and relationships about domains. The **Medical KDB** (Fig. 7.1 top left) stores evidence-based knowledge about type 2 diabetes. One example is information on indications and treatment goals for blood sugar-lowering non-drug or drug therapy that takes into account the predispositions and behavior of patients,

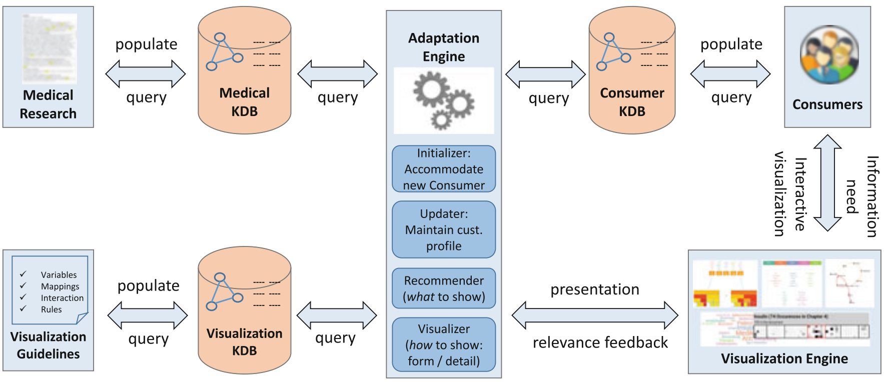

Fig. 7.1 System architecture of proposed A$^+$CHIS: an adaptation engine (center) automatically selects and adapts health information to suit consumers, based on knowledge databases of medical information (top left), consumer profile information (top right), and visualization guidelines (bottom left). A visualization engine (bottom right) presents the adapted information and collects consumer feedback, updating the consumer profile and further improving the user information provision

including information on the possible benefits and harms of intensifying treatment. Another example is criteria for the diagnosis of diabetes and the probability of specific disease courses in relation to patient compliance and behavior. This KDB may be populated using selected literature and research data from diabetes research. The **Consumer KDB** (Fig. 7.1 top right) stores profiles about consumers and describes medically and cognitively relevant characteristics such as biophysical properties like age, gender, medical history, and cognitive properties like pre-knowledge, interests, preferences and expectations, biases, uncertainties, medical literacy, etc. This KDB is populated directly when consumers provide data about themselves and indirectly by inferring interest, e.g., by comparing consumers, or analyzing consumer interactions with the system. The **Visualization KDB** (Fig. 7.1 bottom left) stores rules and best practices for the effective visual representation of data and information, depending on data characteristics, user tasks, and user visual literacy. Examples include effective diagram types to visualize probabilities in relation to the visual literacy of consumers, and rules on how to aggregate and dis-aggregate information, and thus to adjust the provided level of detail and representation of the information content. This KDB is populated based on existing visualization research and guidelines.

The specific tasks of the adaptation engine are to **initialize** the presentation of information to new consumers. We can assume that at the beginning, no profile information on new consumers is available. To initialize the consumer profile, e.g., explicit selection of topics by users, question-answering sessions, or contextual prediction of consumer profiles based on similarity to existing consumers are possible, among others. The tasks of the adaptation engine include keeping track of returning consumers by **updating** existing consumer profiles. The **Recommender** task comprises mechanisms to decide *what* information to show to users and at what level of *detail*. The **Visualizer** task determines *how* to present the information, e.g., numerically, textually, symbolically, diagrammatically, interactively, etc.

The **Visualization Engine** presents health information to the user, provides interaction mechanisms, and, actively and passively, collects **Relevance Feedback** about the consumer, e.g., to determine whether the presented information is helpful and relevant, not relevant or already known, etc. This feedback is used to further adapt and improve the health information and to update the consumer KDB. The visualization engine supports browser-based visualization environments that scale to different platforms, from mobile devices for on-the-go usage, to desktop settings at home, and to large touch-displays e.g., made available in practices and public spaces.

Note that this architecture is flexible with regard to the amount of information in its specific individual knowledge databases. For a first implementation, representative information in the KDBs should be ingested as far as possible. However, we expect it is not possible to obtain complete information (however measured) in each domain. For example, it may not be possible or helpful to attempt to gather all available medical knowledge on diabetes or all imaginable medically and cognitively relevant consumer characteristics. However, one may ensure that sufficient representative information is available in each KDB to enable the research, development, and evaluation of novel A^+CHIS with such an architecture. Information may be collected bottom-up, beginning with representative information provided in existing standard CHIS and gradually extended from the state of the art.

7.6.2 Machine Learning Approaches for Adaptation

7.6.2.1 Main Methods and Application Possibilities in an Adaptive CHIS

The adaptation engine has many tasks which require appropriate algorithms to operate. Machine Learning and Data Science [23] are rapidly expanding fields of technology, which provide a wealth of approaches that can be leveraged for adaption. Important tasks in Machine Learning are clustering, classification, and prediction. *Clustering* groups similar items, useful, e.g., to group information items on the same topic for recommending to users, or, to assign users of the system into similar interest groups. *Classification* assigns labels to new, previously unseen information items or users. It is useful, e.g., to classify the stage and information need of a user within a diabetes development scenario (see Sect. 7.2) or the occurrence of cognitive biases a user may exhibit (see Sect. 7.5.1.4). In *prediction*, one extrapolates information following an observed state. This is useful, e.g., to predict the next information items a user may be interested in, to be able to recommend it.

Also important to us are methods from Natural Language Processing (NLP) [35]. *Information retrieval* methods enable finding relevant information in response to user queries. Similarity can be computed over different information items, from which clusters of information can be determined for over viewing large amounts of information. NLP provides methods for *topic extraction* from text collections, and *text summarization*. The latter is useful to adapt the level of detail by which information is shown, e.g., either in full text form or in an aggregate or just keywords describing topics.

Recommender systems research [52] addresses methods to find matching information items for users to recommend, with many applications, e.g., in e-Commerce, social networks, and information search. The methods typically take into account the recommendation properties of the user, the application and usage context, and the information domain. Methods are based on recommending items among similar users (*collaborative filtering*), recommending similar or dissimilar items (*content-based filtering*), and/or modeling and taking into account knowledge about the application domain. Recently, *Health Recommender Systems* (HRSs) have emerged as an important application domain [108]. As the authors of that survey discuss, goals include improving the understanding of the medical condition, improving the health condition, and motivating a healthier lifestyle. The adaptive CHIS can incorporate such methods.

Our A^+CHIS relies in particular on the visual representation of information, and in the past, visualization systems have incorporated adaptation and recommendation methods to some extent. The Polaris system [102] implemented a rule-based approach to choose an appropriate visual representation, based on the specific dataset to show. The Voyager system [115] supports the interactive exploration of datasets. The user selects initial data variables of interest, and the system suggests to expand this selection by additional variables. This approach has been shown to stimulate interaction and obtain broader insight into the data. In the Draco framework [65], information visualization design knowledge is formalized and can be applied to automatically create visualizations for input datasets, to be explored by the users.

7.6.2.2 Discussion of Machine Learning Approaches

The above is just a selection of techniques from a much larger body of work in Machine Learning, Recommender Systems, and Visualization Automation. It can be used to start building an adaptive CHIS system. There are not only many interesting possibilities but also pitfalls in applying Machine Learning methods to adaptation. Often, an extensive amount of training data or formal modeling of knowledge like rules and ontologies is required. The amount and quality of this data is decisive for the effectiveness of the adaptation. Appropriate training data may not be sufficiently available due to cost, privacy concerns, or small user and/or expert base. In addition, typical problems in Machine Learning like data transformation and normalization, extraction of descriptors/feature vectors as input to the methods, and choice of parameters need to be solved. Hence, algorithms oftentimes do not work out of the box.

On the user side, we wish to provide the users with a good understanding of why the system makes certain adaptations and recommendations. Machine Learning methods in many cases work as a black box, with the user not being able to comprehend the decisions and how they relate to her or his data and requests. Also, the predictions made often come with uncertainties of varying degrees. Recent approaches try to include the users tight in the Machine Learning process by visual representations of the data, the algorithms, and the results [39] and hence improve trust in the results [28].

An adaptive CHIS system can be built step by step, integrating more adaptation methods over time and gradually evaluating these for the effectiveness, acceptance, and eventually, influence on the user health and understanding.

7.7 Conclusion

There is an urgent need for adaptive, personalized, and interactive Consumer Health Information Systems (CHIS), which provide suitable medical information to consumers, considering their current and evolving information needs, health literacy, age, gender, preferences, and knowledge state, etc. To design, implement, and evaluate such an A$^+$CHIS, a synthesis of profound expertise in the fields of information visualization, evidence-based health information, and cognitive psychology is required. The use case for A$^+$CHIS will be type 2 diabetes mellitus since it is highly relevant to public health and affects a broad section of the population. As outlined in Sect. 7.2, this use case is more complex than one might initially think: different patients with type 2 diabetes mellitus may have completely different information needs, which also affects the *what*, *how*, and *level of detail* of the information presented to the consumers.

In the course of a research project in Austria with three universities (Graz University of Technology, Graz University, and Medical University of Graz) lasting for 4.5 years, an A$^+$CHIS will be designed, developed, and evaluated. This A$^+$CHIS aims to overcome the restrictions of current static CHIS, by introducing innovative interactive information visualization techniques, evidence-based health information and principles of cognitive psychology to avoid cognitive biases and misconceptions, over-information or over-diagnosis, and to facilitate comprehension, health literacy, and desirable health-related behavior, including improved communication between patients and medical doctors and increased compliance.

Acknowledgments This work was funded by the Austrian Science Fund (FWF) as part of the project "Human-Centered Interactive Adaptive Visual Approaches in High-Quality Health Information" (A$^+$CHIS Grant No. FG 11-B).

References

1. J. Aanstoos. Visual literacy: an overview. In *32nd Applied Imagery Pattern Recognition Workshop, 2003. Proceedings.*, pages 189–193, 2003.
2. G. Abdel-Hamid. Mind maps as a new teaching strategy for medical students. *MOJ Anat & Physiol*, 3(3):00090, 2017.
3. F. Ackermann, C. Eden, and S. Cropper. *Getting started with cognitive mapping*. Banxia Software, 1992.
4. H. Al-Zubaide and A. A. Issa. Ontbot: Ontology based chatbot. In *International Symposium on Innovations in Information and Communications Technology*, pages 7–12, 2011.
5. D. Albert. *Knowledge structures*. Springer, 1994.
6. D. Albert, C. Hockemeyer, Z. Kulcsar, and G. Shorten. Competence assessment for spinal anaesthesia. In *Symposium of the Austrian HCI and Usability Engineering Group*, pages 165–170. Springer, 2007.
7. D. Albert and J. Lukas. *Knowledge spaces: Theories, empirical research, and applications*. Psychology Press, 1999.
8. D. Albert and C. Steiner. Representing Domain Knowledge by Concept Maps: How to Validate Them? In T. Okamoto, D. Albert, and T. H. . F. W. Hesse, editors, *2nd Joint Workshop of Cognition and Learning Through Media-Communication for Advanced e-Learning (JWCL)*, pages 169–174, Tokyo, Japan, 2005.
9. D. Albert and C. M. Steiner. Reflections on the evaluation of adaptive learning technologies. In *2011 IEEE International Conference on Technology for Education*, pages 295–296, 2011.
10. A. D. Association. Introduction: Standards of medical care in diabetes—2021. *Diabetes Care*, 44(Supplement 1):S1–S2, 2021.

11. A. D. Baddeley and G. Hitch. Working memory. In *Psychology of learning and motivation*, volume 8, pages 47–89. Elsevier, 1974.
12. E. Baumann and F. Czerwinski. *Erst mal Doktor Google fragen? Nutzung neuer Medien zur Information und zum Austausch über Gesundheitsthemen*, pages 57–79. Bertelsmann Stiftung, 01 2015.
13. M. Bedek, P. Seitlinger, S. Kopeinik, and D. Albert. Inferring a learner's cognitive, motivational and emotional state in a digital educational game. *Electronic journal of e-Learning*, 10(2):172–184, 2012.
14. M. A. Bedek, O. Firssova, E. P. Stefanova, F. Prinsen, and F. Chaimala. User-driven development of an inquiry-based learning platform: Qualitative formative evaluations in weSPOT. *Interaction Design and Architecture(s) Journal - IxD&A*, 23:122–139, 2014.
15. M. A. Bedek, M. D. Kickmeier-Rust, and D. Albert. Formal concept analysis for modelling students in a technology-enhanced learning setting. In *ARTEL@ EC-TEL*, pages 27–33, 2015.
16. M. A. Bedek, A. Nussbaumer, L. Huszar, and D. Albert. Methods for discovering cognitive biases in a visual analytics environment. In *Cognitive Biases in Visualizations*, pages 61–73. Springer, 2018.
17. N. D. Berkman, S. L. Sheridan, K. E. Donahue, D. J. Halpern, and K. Crotty. Low Health Literacy and Health Outcomes: An Updated Systematic Review. *Annals of Internal Medicine*, 155(2):97–107, 07 2011.
18. L. Biernatzki, S. Kuske, J. Genz, M. Ritschel, A. Stephan, C. Bächle, S. Droste, S. Grobosch, N. Ernstmann, N. Chernyak, et al. Information needs in people with diabetes mellitus: a systematic review. *Systematic reviews*, 7(1):1–21, 2018.
19. D. Borland, V. West, and E. Hammond. Multivariate visualization of system-wide national health service data using radial coordinates. In *2014 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, 11 2014.
20. M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner. Timelines revisited: A design space and considerations for expressive storytelling. *IEEE Transactions on Visualization and Computer Graphics*, 23(9):2151–2164, 2017.
21. M. Bunge, I. Mühlhauser, and A. Steckelberg. What constitutes evidence-based patient information? overview of discussed criteria. *Patient Education and Counseling*, 78(3):316–328, 2010. Changing Patient Education.
22. H. S. G. Caballero, A. Corvo, P. M. Dixit, and M. A. Westenberg. Visual analytics for evaluating clinical pathways. In *2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, pages 39–46, 2017.
23. L. Cao. Data science: A comprehensive overview. *ACM Comput. Surv.*, 50(3), 2017.
24. N. Cao, D. Gotz, J. Sun, and H. Qu. Dicon: Interactive visual analysis of multidimensional clusters. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2581–2590, 2011.
25. S. K. Card, T. P. Moran, and A. Newell. *The psychology of human-computer interaction. 1983.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.
26. A. Cawsey, F. Grasso, and C. Paris. *Adaptive Information for Consumers of Healthcare*, pages 465–484. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.
27. D. Charnock, S. Shepperd, G. Needham, and R. Gann. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *Epidemiol Community Health*, 53(2):105–111, 1999.
28. A. Chatzimparmpas, R. M. Martins, I. Jusufi, K. Kucher, F. Rossi, and A. Kerren. The state of the art in enhancing trust in machine learning models with the use of visualizations. *Comput. Graph. Forum*, 39(3):713–756, 2020.
29. H. Chen, W. Chen, H. Mei, Z. Liu, K. Zhou, W. Chen, W. Gu, and K. Ma. Visual abstraction and exploration of multi-class scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1683–1692, 2014.

30. L. Chen, D. Lu, M. Zhu, M. Muzammal, O. W. Samuel, G. Huang, W. Li, and H. Wu. OMDP: An ontology-based model for diagnosis and treatment of diabetes patients in remote healthcare systems. *International Journal of Distributed Sensor Networks*, 15(5):1550147719847112, 2019.
31. A. Coulter, V. Entwistle, and D. Gilbert. Sharing decisions with patients: is the information good enough? *BMJ*, 318:318–322, 1999.
32. N. Cowan. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, 24(1):87–114, 2001.
33. F. Dabek, E. Jimenez, and J. J. Caban. A timeline-based framework for aggregating and summarizing electronic health records. In *2017 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, pages 55–61, 2017.
34. B. J. Daley and D. M. Torre. Concept maps in medical education: an analytical literature review. *Medical education*, 44(5):440–448, 2010.
35. S. Dandapat. Nitin indurkhya and fred j. damerau (eds): Handbook of natural language processing (second edition) - CRC Press, Boca Raton, 2010, xxxiii + 678 pp, hardbound, ISBN 978-1-4200-8592-1. *Mach. Transl.*, 25(4):377–381, 2011.
36. L. Daraz, A. Morrow, O. Ponce, B. Beuschel, M. Farah, A. Katabi, M. Alsawas, A. Majzoub, R. Benkhadra, M. Seisa, J. Ding, L. Prokop, and M. Murad. Can patients trust online health information? A meta-narrative systematic review addressing the quality of health information on the Internet. *Journal of General Internal Medicine*, 34(9):1884–1891, 2019.
37. J.-P. Doignon and J.-C. Falmagne. Spaces for the assessment of knowledge. *International journal of man-machine studies*, 23(2):175–196, 1985.
38. G. Ellis. *Cognitive Biases in Visualizations*. Springer, 2018.
39. A. Endert, W. Ribarsky, C. Turkay, B. L. W. Wong, I. T. Nabney, I. D. Blanco, and F. Rossi. The state of the art in integrating machine learning into visual analytics. *Comput. Graph. Forum*, 36(8):458–486, 2017.
40. J.-C. Falmagne, D. Albert, C. Doble, D. Eppstein, and X. Hu. *Knowledge spaces: Applications in education*. Springer Science & Business Media, 2013.
41. M. Feng, B. Xiang, M. R. Glass, L. Wang, and B. Zhou. Applying deep learning to answer selection: A study and an open task. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pages 813–820, 2015.
42. A. Gaisser. Bedarf an Krebsinformation in Deutschland: was für wen und wie? *Forum*, pages 259–264, 2012.
43. M. Garshasbi, H. Asadi, and A. Asosheh. Application and effectiveness of ontology on e-health. In *7'th International Symposium on Telecommunications (IST'2014)*, pages 544–549, 2014.
44. G. Gigerenzer and A. Edwards. Simple tools for understanding risks: From innumeracy to insight. *Bmj*, 327(7417):741–744, 2003.
45. S. Guo, C. Lin, D. Gotz, B. Jin, H. Zha, L. Shu, and N. Cao. Understanding care plans of community acquired pneumonia based on Sankey diagram. In *2016 IEEE Workshop on Visual Analytics in Healthcare (VAHC)*, 2016.
46. F. Hill-Briggs and A. Smith. Evaluation of diabetes and cardiovascular disease print patient education materials for use with low-health literate populations. *Diabetes care*, 31:667–71, 05 2008.
47. I. Hirschberg, G. Seidel, D. Strech, H. Bastian, and M.-L. Dierks. Evidence-based health information from the users' perspective - a qualitative analysis. *BMC health services research*, 13:405, 10 2013.
48. C. Hockemeyer, A. Nussbaumer, E. Lövquist, A. Aboulafia, D. Breen, G. Shorten, and D. Albert. Applying a web and simulation-based system for adaptive competence assessment of spinal anaesthesia. In *International Conference on Web-Based Learning*, pages 182–191. Springer, 2009.
49. U. Hoffrage, S. Lindsey, R. Hertwig, and G. Gigerenzer. Communicating statistical information. *Science*, 290:2261–2262, 2000.

50. K. Horvath, N. Posch, R. Brodnig, J. Plath, and A. Siebenhofer-Kroitzsch. Qualität medizinischer Informationsbroschüren in den Hausarztpraxen der Steiermark. In *51. Kongress für Allgemeinmedizin und Familienmedizin, Der Mensch im Mittelpunkt? Hausärztliches Handeln zwischen Ansprüchen und Alltag*, 2017. Abstract.
51. W. Hsu, R. K. Taira, S. El-Saden, H. Kangarloo, and A. A. T. Bui. Context-based electronic health record: Toward patient specific healthcare. *IEEE Transactions on Information Technology in Biomedicine*, 16(2):228–234, 2012.
52. D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich. *Recommender Systems - An Introduction*. Cambridge University Press, 2010.
53. B. Kamsu-Foguem, G. Diallo, and C. Foguem. Conceptual graph-based knowledge representation for supporting reasoning in African traditional medicine. *Engineering Applications of Artificial Intelligence*, 26:1348–1365, 04 2013.
54. M. Khan, M. Hashim, J. King, R. Govender, and J. Alkaabi. Epidemiology of type 2 diabetes – global burden of disease and forecasted trends. *Journal of Epidemiology and Global Health*, 10, 11 2019.
55. M. D. Kickmeier-Rust and D. Albert. Personalized support, guidance, and feedback by embedded assessment and reasoning: What we can learn from educational computer games. In *IFIP Human-Computer Interaction Symposium*, pages 142–151. Springer, 2010.
56. D. E. Kieras et al. Knowledge representation in cognitive psychology,'. *Mathematical Frontiers of the Social and Policy Sciences*, 1981.
57. Y. Kweon and K.-W. Choi. Readability and suitability evaluation of educational materials on diabetes mellitus. *The Korean Journal of Health Service Management*, 8:161–174, 06 2014.
58. H. Liao, Y. Wu, L. Chen, and W. Chen. Cluster-based visual abstraction for multivariate scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 24(9):2531–2545, 2018.
59. V. Lopez, M. Pasin, and E. Motta. Aqualog: An ontology-portable question answering system for the semantic web. In *Proceedings of the Second European Conference on The Semantic Web: Research and Applications*, ESWC'05, pages 546–562, Berlin, Heidelberg, 2005. Springer-Verlag.
60. E. Lövquist, A. Aboulafia, D. Breen, G. Shorten, D. Zhang, and D. Albert. Designing a simulation-supported adaptive assessment system for spinal anaesthesia. In *Proceedings of the 11th IASTED International Conference Computers and Advanced Technology in Education (CATE2008)*, pages 316–321, 2008.
61. R. E. Mayer. Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, 41:31–48, 2005.
62. R. E. Mayer. *Multimedia Learning*. Cambridge University Press, 2 edition, 2009.
63. A. Mayorga and M. Gleicher. Splatterplots: Overcoming overdraw in scatter plots. *IEEE Transactions on Visualization and Computer Graphics*, 19(9):1526–1538, 2013.
64. S. Mohebi, L. Azadbakht, A. Feizi, G. Sharifirad, and M. Kargar. Review the key role of self-efficacy in diabetes care. *Journal of education and health promotion*, 2, 2013.
65. D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco. *IEEE Trans. Vis. Comput. Graph.*, 25(1):438–448, 2019.
66. B. Moult, L. Franck, and H. Brady. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health care information. *Health Expect*, 7(2):165–175, 2004.
67. I. Muehlhauser, M. Albrecht, and A. Steckelberg. Evidenzbasierte gesundheitsinformationen. *Zbl Arbeitsmed*, 64:334–337, 2014.
68. I. Mühlhauser and M. Lenz. [Does patient knowledge improve treatment outcome?]. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 102:223–230, 02 2008.
69. H. Muller, H. Maurer, R. Reihs, S. Sauer, and K. Zatloukal. Adaptive visual symbols for personal health records. In *2011 15th International Conference on Information Visualisation*, pages 220–225, 2011.

70. T. Munzner. *Visualization Analysis and Design: Principles, Techniques, and Practice*. AK Peters, 2014.
71. S. Narciss. Feedback strategies. *Encyclopedia of the learning sciences, volume F (6)*, pages 1289–1293, 2012.
72. S. Narciss, A. Proske, and H. Koerndle. Promoting self-regulated learning in web-based learning environments. *Computers in human behavior*, 23(3):1126–1144, 2007.
73. L. Nie, M. Wang, Y. Gao, Z. Zha, and T. Chua. Beyond text QA: Multimedia answer generation by harvesting web information. *IEEE Transactions on Multimedia*, 15(2):426–441, 2013.
74. M. Niranjan, M. S. Saipreethy, and T. G. Kumar. An intelligent question answering conversational agent using naive Bayesian classifier. In *2012 IEEE International Conference on Technology Enhanced Education (ICTEE)*, pages 1–5, 2012.
75. A. Nussbaumer, C. M. Steiner, and O. Conlan. Towards a multi-modal methodology for user-centred evaluation of adaptive systems. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, pages 219–220, 2019.
76. A. Ottley, R. Garnett, and R. Wan. Follow the clicks: Learning and anticipating mouse interactions during exploratory data analysis. *Comput. Graph. Forum*, 38(3):41–52, 2019.
77. S. R. Paige, J. L. Krieger, and M. L. Stellefson. The influence of ehealth literacy on perceived trust in online health communication channels and sources. *Journal of health communication*, 22(1):53–65, 2017.
78. M. G. Pandy, A. J. Petrosino, B. A. Austin, and R. E. Barr. Assessing adaptive expertise in undergraduate biomechanics. *Journal of Engineering Education*, 93(3):211–222, 2004.
79. A. Patel and S. Jain. Formalisms of representing knowledge. *Procedia Computer Science*, 125:542–549, 2018.
80. C. Plaisant, D. Heller, J. Li, B. Shneiderman, R. A. Mushlin, and J. Karat. Visualizing medical records with lifelines. In *CHI 98 Conference Summary on Human Factors in Computing Systems*, pages 28–29, 1998.
81. J. Poelmans, D. I. Ignatov, S. O. Kuznetsov, and G. Dedene. Formal concept analysis in knowledge processing: A survey on applications. *Expert systems with applications*, 40(16):6538–6560, 2013.
82. R. Preiner, J. Schmidt, K. Krösl, T. Schreck, and G. Mistelbauer. Augmenting node-link diagrams with topographic attribute maps. *Computer Graphics Forum*, 39(3):369–381, 2020.
83. J. Protheroe, E. Estacio, and S. Saidy-Khan. Patient information materials in general practices and promotion of health literacy: an observational study of their effectiveness. *The British journal of general practice : the journal of the Royal College of General Practitioners*, 65 632:e192–7, 2015.
84. B. R. Ranoliya, N. Raghuwanshi, and S. Singh. Chatbot for university related FAQs. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1525–1530, 2017.
85. G. Rasch. *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.* Nielsen & Lydiche, Oxford, England, 1960.
86. A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, and B. Shneiderman. Interactive information visualization to explore and query electronic health records. *Foundations and Trends in Human-Computer Interaction*, 5(3):207–298, 2013.
87. P. Ruchikachorn and K. Mueller. Learning visualizations by analogy: Promoting visual literacy through visualization morphing. *IEEE Transactions on Visualization and Computer Graphics*, 21(9):1028–1044, 2015.
88. S. Saenger, B. Lang, D. Klemperer, C. Thomeczek, and M.-L. Dierks. *Manual Patienteninformation. Empfehlungen zur Erstellung evidenzbasierter Patienteninformationen*, volume 25. Aerztliches Zentrum für Qualitaet in der Medizin, 01 2007.
89. H. Sampathkumar, X. Chen, and B. Luo. Ontology-based visualization of healthcare data mined from online healthcare forums. In *2015 International Conference on Healthcare Informatics*, pages 325–334, 2015.

90. L. Sbaffi and J. Rowley. Trust and credibility in online health information: a review and agenda for future research. *Journal of Medical Internet Research*, 19, 04 2017.
91. P. Seitlinger, T. Ley, D. Kowald, D. Theiler, I. Hasani-Mavriqi, S. Dennerlein, E. Lex, and D. Albert. Balancing the fluency-consistency tradeoff in collaborative information search with a recommender approach. *International Journal of Human–Computer Interaction*, 34(6):557–575, 2018.
92. L. Shao, A. Mahajan, T. Schreck, and D. J. Lehmann. Interactive regression lens for exploring scatter plots. *Computer Graphics Forum*, 36(3):157–166, 2017.
93. C. Shieh and B. Hosei. Printed health information materials: Evaluation of readability and suitability. *Journal of community health nursing*, 25:73–90, 04 2008.
94. V. J. Shute and Y. J. Kim. Formative and stealth assessment. In *Handbook of Research on Educational Communications and Technology*, pages 311–321. Springer New York, 2014.
95. K. Sørensen, J. M. Pelikan, F. Röthlin, K. Ganahl, Z. Slonska, G. Doyle, J. Fullam, B. Kondilis, D. Agrafiotis, E. Uiters, M. Falcon, M. Mensing, K. Tchamov, S. v. d. Broucke, and H. Brand. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). *European Journal of Public Health*, 25(6):1053–1058, 04 2015.
96. K. Sørensen, S. Van den Broucke, J. Fullam, G. Doyle, J. Pelikan, Z. Slonska, H. Brand, and (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health*, 12(1):80, Jan 2012.
97. J. F. Sowa. Architectures for intelligent systems. *IBM Systems Journal*, 41(3):331–349, 2002.
98. P. Srivastava and N. Singh. Automatized medical chatbot (medibot). In *2020 International Conference on Power Electronics IoT Applications in Renewable Energy and its Control (PARC)*, pages 351–354, 2020.
99. L. Stefanutti and D. Albert. Skill assessment in problem solving and simulated learning environments. *Journal of Universal Computer Science - J. UCS*, 9(12):1455–1468, 2003.
100. B. Steichen, G. Carenini, and C. Conati. User-adaptive information visualization: Using eye gaze data to infer visualization tasks and user cognitive abilities. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, IUI'13, pages 317–328, New York, NY, USA, 2013. Association for Computing Machinery.
101. C. M. Steiner, D. Albert, and J. Heller. Concept mapping as a means to build e-learning. *Advanced principles of effective e-learning*, pages 59–111, 2007.
102. C. Stolte, D. Tang, and P. Hanrahan. Polaris: a system for query, analysis, and visualization of multidimensional databases. *Commun. ACM*, 51(11):75–84, 2008.
103. H. Suominen, T. Schreck, G. Leroy, H. Hochheiser, L. Goeuriot, L. Kelly, D. L. Mowery, J. Nualart, G. Ferraro, and D. A. Keim. Task 1 of the CLEF eHealth Evaluation Lab 2014. In *Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15–18, 2014*, pages 1–30, 2014.
104. J. Sweller, J. J. Van Merrienboer, and F. G. Paas. Cognitive architecture and instructional design. *Educational psychology review*, 10(3):251–296, 1998.
105. D. Toker, C. Conati, G. Carenini, and M. Haraty. Towards adaptive information visualization: On the influence of user characteristics. In *User Modeling, Adaptation, and Personalization*, pages 274–285, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
106. B. Tony and B. Buzan. The mind map book: How to use radiant thinking to maximize your brain's untapped potential, 1993.
107. B. Tran, M. Singh, B. Lee, R. Rudd, and D. Singhal. Readability, complexity, and suitability analysis of online lymphedema resources. *Journal of Surgical Research*, 213, 03 2017.
108. T. N. T. Tran, A. Felfernig, C. Trattner, and A. Holzinger. Recommender systems in the healthcare domain: state-of-the-art and research issues. *J. Intell. Inf. Syst.*, 57(1):171–201, 2021.
109. A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.

110. M. Wagner, D. Slijepcevic, B. Horsak, A. Rind, M. Zeppelzauer, and W. Aigner. KAVAGait: Knowledge-assisted visual analytics for clinical gait analysis. *IEEE Trans. Vis. Comput. Graph.*, 25(3):1528–1542, 2019.
111. T. D. Wang, C. Plaisant, B. Shneiderman, N. Spring, D. Roseman, G. Marchand, V. Mukherjee, and M. Smith. Temporal summaries: Supporting temporal categorical searching, aggregation and comparison. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1049–1056, 2009.
112. M. Ward, G. Grinstein, and D. Keim. *Interactive Data Visualization: Foundations, Techniques, and Applications*. Taylor and Francis, 2015.
113. R. Wille. Formal concept analysis as mathematical theory of concepts and concept hierarchies. In *Formal concept analysis*, pages 1–33. Springer, 2005.
114. A. Williams, K. Muir, and J. Rosdahl. Readability of patient education materials in ophthalmology: A single-institution study and systematic review. *BMC Ophthalmology*, 16:133, 08 2016.
115. K. Wongsuphasawat, D. Moritz, A. Anand, J. D. Mackinlay, B. Howe, and J. Heer. Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE Trans. Vis. Comput. Graph.*, 22(1):649–658, 2016.
116. J. S. Yi, Y. ah Kang, J. T. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. Vis. Comput. Graph.*, 13(6):1224–1231, 2007.
117. O. V. Yurdakul, M. S. Kilicoglu, and F. Bagcier. Evaluating the reliability and readability of online information on osteoporosis. *Archives of Endocrinology and Metabolism*, 65:85–92, 02 2021.
118. Y. Zhang, K. Chanana, and C. Dunne. IDMVis: Temporal event sequence visualization for type 1 diabetes treatment decision support. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):512–522, 2019.
119. B. J. Zimmerman. Becoming a self-regulated learner: An overview. *Theory into practice*, 41(2):64–70, 2002.

Chapter 8
Design Cognition in Data Visualization

Paul C. Parsons

Abstract In this chapter I introduce the topic of design cognition and its relevance to data visualization. I outline two historically dominant paradigms of design cognition. The first, promoted by Herbert Simon in the 1970s, is the rational problem solving paradigm which is based on information processing psychology and problem solving theory. The second, promoted by Donald Schön in the 1980s, is the reflective practice paradigm which is based on constructivist philosophy and situated views of cognition. I outline some of their strengths and weakness and attempts to reconcile their differences. Underlying philosophical issues pertaining to cognition and epistemology are briefly discussed. I then examine implications of these two paradigms for four data visualization topics: defining, automating, modeling, and teaching data visualization design. In discussing these topics, possible avenues of future research are proposed.

8.1 Introduction

How do designers formulate and solve design problems? What kinds of cognitive processes do they rely on while doing so? These are the types of questions asked by the researchers studying design cognition. Rather than focusing on only methods, tools, or outcomes of designers, studies in design cognition investigate how and why designers think in the ways they do while designing. Design cognition has been studied across a wide variety of disciplines, including engineering [4], architecture [37], computer science [9], instructional design [59], and graphic design [76]. Across these disciplines, many aspects of cognition in design have been investigated, including, among others, episodic memory [35], fixation [62], chunking [43], bias [16], abductive reasoning [11], analogical reasoning [79], metacognitive monitoring and control [5], and recall [13]. A number of core strategies of design

P. C. Parsons (✉)
Purdue University, West Lafayette, IN, USA
e-mail: parsonsp@purdue.edu

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_8

thinking have been identified through empirical investigation, including conjecture-based problem formulation, problem-solution co-evolution, analogical reasoning, mental simulation, and fixated solution generation [5, 24].

Many cognitive structures and processes that are important for the use of visualizations are also important for their design. For instance, studies have shown that designers rely on chunking to ideate effectively [43], employ abductive reasoning during concept selection [19], are influenced by color in ways that bias their thinking while sketching [16], rely on shared mental models in collaborative settings [18], and struggle with fixation while generating ideas [12]. Topics such as cognitive bias, visual reasoning, fixation, and mental models are commonly seen in the visualization literature; however, they are almost exclusively focused on the cognition of users rather than the cognition of designers. In this chapter, I argue that design cognition is an important yet neglected area of study for data visualization.

8.1.1 Why Study Visualization Design Cognition?

It is easy to appreciate why the cognition of users is an important area of inquiry for data visualization. The many different ways of visualizing data and creating interactive interfaces have implications for how users interpret, understand, and act. Topics relating to visual marks and channels, color, mental models, uncertainty, biases, sensemaking, and others have received significant attention in the visualization literature. Indeed, the vast majority of literature at the intersection of visualization and psychology is focused on users, whereas the cognition of designers is largely an unexplored topic. The reason for this is unclear, especially since design cognition has been a research topic in multiple design disciplines for decades.

One reason for a lack of inquiry into design cognition may stem from a commonly held assumption within many scientific fields. The assumption is that the design consists largely of the *application* of scientific knowledge to instrumental problems [8, 66]. From this standpoint, the important forms of design knowledge include laws, principles, guidelines, patterns, and other objective forms of knowledge that can be codified, and the prototypical role of the designer is to know these and apply them to various design situations. For example, there may exist design guidelines about data types, visual marks and channels, color, visualization and interaction techniques, evaluation strategies, and other similar forms of knowledge. From this "application" perspective, the designers need to first know about these things and then determine how to apply them to particular problems. If they have the right training, experience, and access to guidelines when they need them, they can apply this existing knowledge in useful ways.

While this "application" view may be appealing, it has been largely abandoned by design scholars, as it has not held up well to empirical scrutiny [14, 36, 39]. While objective, codified knowledge certainly plays a role in design and in many cases may be necessary, it is not sufficient for good design [75]. Rather, designers rely on a host of personal and contextual factors—along with the more formal types of knowledge—to engage with the complexity of situations they face in the real world [56]. Buchanan [8] articulates how widespread this assumption has been, noting that "each of the sciences that have come into contact with design has tended to regard design as an 'applied' version of its own knowledge," emphasizing the mistake of viewing design as simply a "practical demonstration" of scientific findings. Thus, even if a robust program of research at the intersection of psychology and visualization is developed, if its scope is limited to users only—and especially if design is viewed merely as an application of research findings—we will likely fail to understand and influence design practice effectively.

It is useful to ask whether the cognition of data visualization designers really needs to be studied when findings from several other design disciplines already exist and could be translated into a visualization context. There is evidence that aspects of design cognition are common across different design disciplines, although significant differences have also been identified [81]. For instance, Akin [2] found significant differences in design cognition among engineers and architects, and Purcell and Gero [62] found differences between mechanical engineers and product designers. In a review paper examining design across numerous domains, Visser [81] affirms the existence of differences and similarities and speculates that these differences may have implications for the kinds of knowledge that designers rely on. From this evidence, it is reasonable to assume that design cognition in visualization will share similarities with other fields yet will also have its own unique characteristics. For instance, designers in other fields—even related fields like interaction design and graphic design—may not have to navigate issues involving data wrangling, visual mapping, perceptual and cognitive considerations, and interactivity, all of which are important for data visualization. Research on design cognition has significantly influenced theory, practice, and education in numerous design fields [33] and could similarly do so for data visualization. However, the particular facets of visualization design cognition that make it different from other design disciplines must be carefully examined as a part of such an effort.

8.1.2 Methods for Studying Design Cognition

While the focus of this chapter is not methodological, a brief overview of how design cognition has been studied may be beneficial. Previous research has heavily relied on "protocol studies" to elicit cognitive processes [15, 24]. This method, which is already well known to visualization researchers doing human subjects studies, involves asking designers to "think-aloud" while doing a design activity. These studies generate verbal protocols that can be transcribed and analyzed with the goal of uncovering aspects of thinking and reasoning. This kind of approach can be taken with individual designers who work alone on design problems or with teams of designers working together. Team-based protocols have been used to elicit socio-cognitive facets of collaborative design cognition [2].

Design cognition can also be studied in both controlled settings, such as a lab or workshop, and in less controlled settings, such as designers' everyday work environments. Studies in controlled settings can be beneficial, as they allow common design tasks to be given to participants and allow for the control of variables, including time spent, access to resources, and so on. Although empirical lab studies are commonly employed in design research (e.g., [10, 31]), they differ from realistic design contexts in a number of ways. For instance, lab studies may exclude factors that shape design work in commercial settings, including the effects of organizational culture, project timescales, project management, and workload. Lab studies may also present participants with relatively simple problems over short time periods, which are not often representative of real-world design tasks. As is often the case in experimental research, there is the risk of reducing both ecological and external validity [10]. For these reasons, it is beneficial to conduct studies in both controlled "lab" settings and "in the wild" of real-world practice. Studies can employ a range of methods, including protocol analysis, semi-structured interviews, diary studies, contextual observations, and co-design workshops.

8.2 Two Paradigms of Design Cognition

Two dominant paradigms have historically been used to describe the cognitive nature of design. The first was articulated by Herbert Simon in the 1970s and was inspired by information processing psychology and theories of problem solving. The second was articulated by Donald Schön in the 1980s and was inspired by constructivist philosophy and situated views of cognition and professional practice. These two paradigms present different perspectives on the cognitive activities involved in designing and have different implications for how design is studied and taught. In what follows, I will refer to these views often as the "rational problem solving" view and the "reflective practice" view. These two paradigms will be described next, followed by some of their implications in general and for data visualization in particular.

8.2.1 Design as Rational Problem Solving

The work of Herbert Simon has been highly influential in the study of design cognition [65], despite there being few references directly discussing design cognition among his nearly 1000 articles [80]. In particular, his book *The Sciences of the Artificial* [73] and his article "The Structure of Ill-Structured Problems" [74] serve as the foundation of his work on design cognition. Simon's view of design cognition was unsurprisingly influenced by his pioneering body of work on problem solving (see [54]). Simon was working on design cognition not long after the "cognitive revolution" of the 1950s within psychology (in fact, he was a central figure in the movement). The dominant view of the mind at the time was as a symbolic information processor. Simon was a proponent of this view, and he saw design cognition as a form of heuristic search carried out by an information processing system. In this view, the logic of design involves *finding* alternatives within a space of possibilities while using certain strategies to manage the complexity of the space.

Simon viewed the "shape of design" as essentially hierarchical. He embraced a systems perspective, in which the way to design for a complex problem is to decompose it into sub-problems or sub-functions. Once the complex problem is decomposed, "the design of each component can then be carried out with some degree of independence of the design of others" [73]. He viewed the design process as involving, "first, the generation of alternatives and, then, the testing of these alternatives against a whole array of requirements and constraints" [73]. He also had ideas for how to sequence activities within the design process that were inspired by how computer programs can engage in top-down programming and resource allocation. Stated inspirations for his theory of design include decision theory, control theory, dynamic programming, heuristic search and means-end analysis, resource allocation, and hierarchical decomposition.

Overall, Simon's view of design is formal, objective, and computational. He was inspired by utility theory and statistical decision theory as logical frameworks for rational choice among alternatives within a design space. He recognized this choice could not be optimized, instead advocating for satisficing [72] as the dominant heuristic. He thought of design problems mechanistically, as systems of interrelated parts that could be broken down, solved, and put back together again. The cognitive acts involved in these processes were essentially part of a heuristic search process, with the aim of deducing which of the available alternatives satisfies the given design criteria within a set of constraints.

8.2.2 Design as Reflective Practice

Donald Schön presented the most well-known alternative to Simon's problem solving view of design in his book *The Reflective Practitioner* [66] and a series of subsequent papers. Schön rejected the view of design as an information processing or heuristic search problem, instead positioning it as a form of *making* in which design cognition is fundamentally transactional in nature, unfolding as a conversation with the materials of the design situation [67, 68]. He drew on Nelson Goodman's notion of worldmaking [25], which posits that people are continuously making and maintaining the worlds that are matched with their professional knowledge. People have "particular, professional ways of seeing their world and a way of constructing and maintaining the world as they see it" [66]. According to Schön, the designer does not mainly search through a solution space; rather, the designer actively structures the space by framing it, determining which things to attend to, and imposing their view of the world on it—in this view, design is a much more constructive kind of enterprise than the view put forward by Simon.

Table 8.1 Two paradigms of design cognition compared. Adapted from Dorst and Dijkhuis [22]

	Rational problem solving	Reflection in action
Designer	Information processor	Person constructing reality
Design Problem	Ill defined, unstructured	Essentially unique
Design Process	Heuristic search	Reflective conversation
Design Knowledge	Laws, rules, procedures	Precedent, experience
Design Decisions	Rational, objective decisions	Personal, tacit judgments
Example/model	Optimization theory	Professional artistry

Schön argued that the problem solving view is accurate only when ends are fixed and clear, which is not typical of design problems. He pushed back against the "technical rationality" of the problem solving view, noting that the central *problem setting* work of the designer is not technical: "it is rather through the non-technical process of framing the problematic situation that we may organize and clarify both the ends to be achieved and the possible means of achieving them" [66]. This kind of framing work falls outside the scope of the problem solving view, yet it constitutes a large part of the cognitive work of designing. Designers "name and frame" problems, using their professional judgment to assess particular situations and identify the problem or opportunity to be addressed. These situations tend to be unique, complex, and dynamic—thus not well suited to the simple application of theoretical knowledge to the situation at hand.

One of Schön's critiques of the problem solving view is that it is too narrow and does not capture much of what actually happens in design. Beyond this scope issue, however, there is a subtler distinction in the nature of the cognitive acts being carried out. Simon's view suggests that designers plan and select from alternatives systematically, sometimes iteratively and in parallel, eventually finding a solution that fits the design criteria. Schön's view is that cognitive acts are much more contingent and situated. The designer engages in a process of "seeing-moving-seeing," consisting of action sequences where there are unintended consequences of each move that is made. From this perspective, the design process is essentially a conversational structure, where, as in a conversation between friends, there is no way to predict each turn the conversation will take. It is unpredictable yet still disciplined. It does not follow a pre-determined process, but rather reacts to the needs of the situation given the professional knowledge of the designer. Table 8.1, adapted from a similar table by Dorst and Dijkhuis [22], summarizes the two paradigms and some of their key differences.

8.3 Attempts at Integration

While design as problem solving and design as reflective practice have been widely acknowledged as two radically different, competing paradigms, there have been multiple attempts to integrate them—or at least see value in each. Dorst suggests that although these two paradigms "are on opposite sides of a deep schism that runs through science and philosophy" [21], they both can be valuable in understanding design. Both Simon and Schön did not pay enough attention to the structure of design problems, asserts Dorst, and by doing so we can see how both paradigms can describe a single design process [20]. Dorst posits that rational problem solving "is better for describing the more determined problem stretches of problem solving, and a variant of reflective practice, with its sensitivity to interpretation and situatedness, could be used to pinpoint the structure in the underdetermined episodes of design thinking, especially the moments of 'breakdown' (or 'reframing')" [20]. In one study, Dorst and Dijkhuis [22] attempted to describe an industrial design process using the two paradigms, with a focus on how closely they matched the *experiences* of the designers. They concluded that the rational problem solving view works well when design problems are clear-cut—but not otherwise. They also concluded that the reflective practice view does not offer as much descriptive precision and rigor, but it more closely matches the actual experiences of designers and provides a better description of both the design process and its content.

Visser has proposed a similar approach, noting that "design involves problem solving, but that design is not (only) problem solving" [80]. However, although Visser acknowledges the value in Simon's perspective, she also suggests that Simon "misrepresented" design in six key ways, including that his position overestimated the importance of problem decomposition, the importance of search, and the importance of means-end analysis as cognitive aspects of design. Visser suggests Simon's misrepresentations are due to his proclivity to view engineering as the prototypical design discipline, and in doing so he neglects the wide diversity of design traditions. Visser believes not only that situated perspectives on design cognition are essential, as in Schön's view, but also that Schön's work lacked some precision that has since been improved by others (e.g., [1]). Visser ultimately notes that while the problem solving approach does have value, the situated approach "has in principle the potential to propose a more appropriate view on design" [80].

More recently, Hatchuel has argued that Simon's attempts to develop a theory of design cognition were left unfinished [30]. Inspired by Simon's famous concept of bounded rationality, Hatchuel proposed the concept of "expandable rationality" as a paradigm that addresses some of Simon's shortcomings. This topic will be addressed in more detail in Sect. 8.4.2 in the context of automated visualization design.

8.3.1 Philosophical Considerations

Despite attempts to integrate the two paradigms in pragmatically useful ways, there are still fundamental differences between them that are not easy to rectify. These issues are related to much bigger philosophical arguments in cognitive science about the nature of symbolic information processing in complex real-world situations [77, 78] and to debates about epistemology in science and design. Each of these paradigms is built on an underlying philosophical view of knowledge and of the world. Simon's perspective is predominately positivist, as evidenced by his emphasis on empiricism, logic, and objectivity. He was influenced by computational metaphors of information processing psychology, which view the mind as essentially a disembodied information processor. Simon was either not aware of or dismissed contemporary work arguing that knowledge in general is essentially personal and tacit [60, 61] and that design knowledge in particular relies on social and political judgments [64]. Schön explicitly rejected the positivist epistemology of Simon and instead built his perspective from a constructivist orientation [34]. Schön embraced the personal and tacit nature of knowledge and explicitly rejected the idea that design knowledge could be fully codified in any objective manner [68].

There is no hope to rectify these epistemological issues here, and the philosophical debates are somewhat removed from data visualization research and practice. However, these differences do have implications for data visualization. For instance, one difference deals with how designers make decisions as they move through their process. I have previously interviewed data visualization practitioners to understand design practice as described in their own terms (see [55–58]). During these conversations, it was abundantly clear from their descriptions that judgments, rather than formal decisions, were essential cognitive acts that drove their design process. Practitioners described engaging in judgements all the time—often in overlapping, layered ways [56]—and engaging in logical decision-making processes and search strategies very rarely if at all. This reliance on professional judgment has been noted elsewhere—e.g., among data analysts [3], interaction designers [40], and instructional designers [27].

Simon's view of design cognition rejects judgment as a legitimate cognitive act. He instead views the cognitive acts of design as involving formal logic and rational search strategies that in theory and practice can be fully codified. He wrote in praise of computer programs that could fully represent complex design processes, where "there is no question . . . of the design process hiding behind the cloak of 'judgment' or 'experience'." Simon viewed judgment with suspicion, as if any decision-making processes that are rigorous should essentially be separable from the decision-maker and thus be amenable to codification in formal language. I have previously written about how a concept like chartjunk is used by practitioners in personal, contextually relevant ways [57]. It appears unlikely that the use of such a concept could be codified in a rule-based prescriptive manner. There is evidence that effective design relies on personal, tacit, situated knowledge—the kind of knowledge that cannot be articulated and codified. For instance, when discussing the notion of chartjunk with visualization practitioners, its conceptualization, interpretation, and application were all very much tied to the individual designer and their judgments about the design situation [57]. In recounted situations, designers made judgments that were effective, yet there was no apparent procedural explication of the entire process behind the design outcomes. The personal, situated, and tacit nature of design judgments has been demonstrated in other design fields, including instructional design [27] and interaction design [75].

On the role of judgment in design cognition, it appears that Simon was mistaken. Yet the positivist orientation in general and the rational problem solving view in particular are still very popular in STEM fields. Meyer and Dykes [47] have recently written about the positivist leanings of the visualization and computer science communities. Scholars in other fields have written about the need for recognizing epistemological issues in the use of data and visualizations (e.g., Drucker [23] in the humanities). Schön [67] pointed out the dominance of positivism in academia and in the professional schools, referring to the dominant epistemology as *technical rationality*. From this standpoint, the kinds of knowledge that are valued are general, objective, and abstract. Many academics in science and technology fields learn that these characteristics are the hallmarks of good research. Herein lies the consequence of adopting, even unconsciously, one of these underlying philosophical orientations. If the researchers are trained in an environment embracing technical rationality, they will likely strive to generate knowledge that is abstract, objective, and general. Even while doing empirical work, they may only see that which is abstract, objective, and general because that is what they have learned to recognize as valuable. The philosophical orientation can be reinforced even through empirical investigation because it acts as lens through which the researcher interprets the world.

The issue of philosophical orientation is of course a very general one, so what does it mean for visualization researchers? If the field is primarily positivist in orientation, researchers are likely to gravitate toward the rational problem solving view because it claims to be objective and systematic. They will then view visualization design primarily as a heuristic search problem carried out by designers as information processors. If researchers do not recognize design cognition as comprising personal and tacit forms of knowledge, they will not develop a sufficient understanding of how design is actually practiced and may not be able to train visualization designers effectively. While these are some generic implications of adopting one paradigm—most likely to be the rational problem solving paradigm in the visualization community—I attempt to elaborate some more specific implications in the following section.

8.4 Implications for Data Visualization

The way that design cognition is construed is not simply a matter of abstract intellectual debate and has many—often subtle—consequences for visualization research, practice, and training. In the following sections, I discuss some implications for defining, automating, modeling, and teaching data visualization design.

8.4.1 Defining Design for Data Visualization

Addressing many of the challenges and differences discussed in this chapter rests on determining what kind of activity design is considered to be. The question of what constitutes design is inevitable but not easy to answer. There are the well-known generic answers, like Simon's devising "courses of action aimed at changing existing situations into preferred ones" [73]. Nelson and Stolterman promote design as a "third way," distinct from science and art, describing it as "the ability to imagine that-which-does-not-yet-exist, to make it appear in concrete form as a new, purposeful addition to the real world" [53]. Regarding visualization specifically, van de Moere and Purchase draw from multiple definitions of design to emphasize the creative aspects of design and the personal role of the designer in shaping the design process [50]. In their influential paper on design study methodology, Sedlmair et al. [70] define design as the "creative process of searching through a vast space of possibilities to select one of many possible good choices from the backdrop of the far larger set of bad choices." This definition is closely aligned with the rational problem solving view of Simon described above. The constructive, situated view of designing is not well articulated in the core visualization literature, although there are instances of this view on visualization design in the digital humanities (e.g., [17, 23]).

Providing an agreed-upon definition of design for data visualization is likely an impossible goal. One possible way to achieve consensus is to focus on the guiding values or ideals of the discipline, as has been done recently for interaction design [32]. I will not attempt to do that here but will suggest that any agreement about the nature of design cognition needs to rest on some degree of consensus about the nature of design broadly construed. For instance, if design is limited to the selection of visualization and interaction techniques from a known set of possibilities, implications are different from the view where design includes the influence of personal competencies and philosophical commitments, the role of tacit knowledge, the reliance on deep patterns of personal experience, and the real-world concerns of managing clients, expectations, software tools, deadlines, budgets, and so on.

Any definition of design should be informed by the cognitive and other considerations that go into the work of doing design. Whether design cognition is fundamentally rational problem solving or reflective practice necessarily influences the definition of design. Is design primarily about discovery, as Simon's view might suggest, or about constructing the reality of the situation, as Schön's view suggests? Design theorists Nelson and Stolterman reject the problem solving view of design, noting its focus being limited to "that-which-is (description and explanation)" instead of "that-which-ought-to-be (ethics and morality), and …that-which-is-desired (desiderata)" [53]. Adopting any view of visualization design in general, and design cognition in particular, will influence the view of other important topics like automated design, design models, and design education, each of which will be discussed in the following sections.

8.4.2 Automated Visualization Design

One topic for which design cognition has deep implications—in ways that may not be immediately obvious—is automated visualization design. Since the early days of AI, the researchers have had a desire to model human cognition computationally, with the goal of either fully or partially automating cognitive processes and human knowledge. One particular instance of this is in modeling design knowledge and processes with the goal of automating design. There are implications for the success of this vision, however, based on the true nature of design cognition. For instance, if human cognition operates fundamentally in a symbolic information processing mode—which is the basic thrust of information processing psychology and Simon's theory of problem solving—then it should be possible to replicate cognitive processes computationally. However, if symbolic information processing is not the right metaphor or is at least not the right one for design cognition, it may not be possible to fully automate the relevant knowledge and cognitive processes for design.

Within the visualization literature, the goal of automating design has a long history. The most well-known early work was from Mackinlay on his presentation tool APT [42], which viewed the design of graphical representations as fundamentally a search problem aiming to optimize effectiveness and expressiveness. Subsequent research in this space has informed the design of systems like Tableau [41], SAGE [49], Voyager [83], Draco [51], and numerous others in recent years.

These automated tools appear to be framed around two visions. One vision is to assist analysts in understanding their data. Because analysts do not have expertise in visualization design, it can be difficult for them to create useful visualizations. As a result, they will often turn to default charting options in tools like Excel, which may be unhelpful or even misleading. If visualizations instead can be recommended to them based on characteristics of the data and the analyst's goals, the analysis situation can be greatly improved. This is the stated intention behind techniques like Tableau's "Show Me" [41] and more recent work on query languages like CompassQL [84]. Another vision for these automated tools is to assist the designers in creating visualizations based on codified design knowledge. This is a perspective taken by recent systems like Draco[51], in which the aim is to formally model design knowledge with hard and soft constraints over logical facts. This vision embraces the "application" view described previously in Sect. 8.1.1—researchers conduct empirical studies to generate design knowledge, which then gets applied in practice. One stated motivation for creating these systems is that there is a gap that needs to be filled between the researcher-generated knowledge and its application in practice.

There is no doubt that automated visualization design tools are useful, especially as aids to analysts who do not have adequate visualization design knowledge. In such cases, however, the "design" work that is being done is narrow with respect to the whole range of considerations that go into real-world design practice [55]. The extent to which visualization design can be automated depends in part on the nature of design cognition, design knowledge, and design practice. If the complexity, uncertainty, and messiness of real-world design is not considered, computational tools can automate only small parts of the design process. Here, we see again implications of which paradigm(s) of design cognition is embraced. If design cognition is fundamentally symbolic information processing, as in Simon's view, design knowledge and processes are separable from the designer and can be codified as such. Based on the reflective practice view, however, an expectation for fully automated design is unrealistic even in principle. Unless artificial intelligence can develop the subtle appreciative and imaginative abilities of humans, there is no hope for design to be fully automated. On this matter, Schön [68] argues that computers would need to achieve *phenomenological and functional equivalence* with humans to be able to reproduce essential aspects of design cognition, including the continual, subjective appreciation of a situation and the envisioning of future design worlds. Previous work by Alspaugh et al. [3] has surfaced similar concerns from professional data analysts about the role of automation in data wrangling and visualization.

A perspective that may be helpful here is Hatchuel's concept of "expanded rationality" [30], briefly described previously in Sect. 8.3. Hatchuel argues that true design problems involve infinite and non-countable sets, for which heuristic search is not an appropriate strategy. Bounded rationality does not help with these kinds of situations because true design situations are infinitely expandable. For instance, when a client approaches a designer and says "help me see something useful in my sales data," there is no bounded problem space that can be computationally exhausted. The concept of "useful" is infinitely expandable, and it is the task of the designer to frame the problem space and make the design task manageable. The designer may interview the client and other stakeholders, for instance, and determine that what is useful is to see the growth of certain market segments in relation to political or natural events. This is very much in line with Schön's emphasis on the active, constructive nature of framing the problem to be addressed. Even in principle, infinite time and computing resources could not explore the space. Hatchuel's theory argues that a situation is a "real design problem" only if the initial concepts allow for unexpected expansion. The designer makes use of their creative, imaginative, and appreciative abilities to do this work. If there is no opportunity for expansion, Hatchuel argues there is no real *design* problem—only a regular problem to be solved. This view may provide conceptual support for the issue of automation in design, as it offers a useful description of what distinguishes design from typical problem solving.

Aside from the cognitive processes involved in design, a central topic for understanding the possibilities of automation in design is *design knowledge*. If the design involves knowledge that is—even in part—fundamentally irreducible, it is not possible to codify. There is much evidence that design knowledge is holistic and personal in the ways that Polanyi described scientific knowledge [60]. Certainly some design knowledge is objective and able to be codified, but if critical pieces of knowledge are personal and tacit, there is little hope to fully automate design.

The nature of design knowledge has been written about extensively by others and is beyond the scope of this chapter.

When it comes to automated visualization design, there are at least two key topics that must be encountered. First, if design cognition is expansive in nature, computational search processes are not sufficient for fully automating design—even in principle. Second, if design knowledge is personal and tacit, it cannot be fully codified. Some aspects of design can be automated, especially those that deal with heuristic search through spaces of known solutions and those that deal with objective kinds of knowledge like laws and principles. However, it is important to recognize that these cover only a portion of what is involved in real-world design practice. The important question is not whether design can be fully automated, but rather: what are the cognitive processes and types of knowledge that can be codified and how should tools work in concert with designers to leverage computational and human strengths in design?

8.4.3 Visualization Design Models and Frameworks

Similar to the topic of automated visualization design is the view of design models and frameworks in the visualization literature. Numerous frameworks and models have been proposed to describe the design process and to provide researchers and designers with advice and guidance. For instance, popular decision models include the Nested Model [52] and its Blocks and Guidelines extension [48]. Popular process models include the nine-stage framework in the Design Study Methodology [70], the Design Activity Framework [45], and others [26, 44, 69, 71].

Here, we can again ask what role the underlying paradigms of design cognition have played in the development of these models and frameworks. Most appear to be closer in spirit to Simon's view of design cognition—they are intended to explicate knowledge, support decision-making, and enable a rational search process through a space of known possibilities. They do not appear to align as much with Schön's reflective practice view, in which we would expect to see descriptions of the situated, transactional nature of design—e.g., problem framing and setting, imagination, judgment, and contingent means of navigating the design situation. We would also expect to see descriptions of reflection in action taking place. These existing models generally do value reflection, but mainly as a mechanism for making contributions back to the research community and not as an essential cognitive act that aids movement through the design process. For instance, the nine-stage framework [70] has as its 8th stage "reflection." By this is meant reflection *after* the design process, on what was done and how it relates to the research landscape, rather than *during* design, as an essential and continual feature. In their paper "Reflection on Reflection in Applied Visualization Research," Meyer and Dykes [46] have noted how there is a "bias towards post-study reflection," recommending "a more structured and purposeful approach to reflection throughout the entire design process." Additionally, many of these models and frameworks acknowledge the ill-structured and wicked nature of design problems. However, Simon acknowledged this as well [74] but still promoted the rational problem solving view. Referring to design as ill-structured, wicked, messy, complex, or iterative does not indicate which paradigm is being adopted. My goal here is not to criticize existing frameworks, but rather to raise the question of how they relate to paradigms of design cognition.

The models and frameworks discussed above have not explicitly engaged with paradigms of design cognition to motivate their development, perhaps due to a lack of awareness, as it is largely an unknown topic within the visualization literature. However, these models and frameworks rely on assumptions about cognitive aspects of design, whether they are explicitly acknowledged or not. These assumptions are of course influenced by the primary philosophical paradigms within the visualization field, which Meyer and Dykes [47] have recently argued is dominated by positivism. Furthermore, these models have largely been developed for the academic research community. Here there may be somewhat of a self-fulfilling prophecy taking place. If a research community adopts one paradigm of design cognition—unconsciously or not—it will value models that conform to that paradigm. For example, a positivist community will value objective, empirical, abstract knowledge; because that is what is valued, researchers will attempt to develop abstract, general models; and these models will then be evaluated on the same positivist criteria that led to their development. Finally, it is important to note that these models and frameworks have been generated and used within the research community, and their applicability and relevance to real-world practice is uncertain [55]. The cognitive acts of researchers doing design work and practitioners doing design work likely not only have some differences but also likely share foundational elements. However, the nature of these relationships has not been investigated and could be a topic for future inquiry.

8.4.4 Visualization Education

The view of design cognition that is adopted by educators has significant implications for data visualization pedagogy. For instance, consider the "application" view described previously in Sect. 8.1.1. Within this perspective, there is an implicit hierarchy of knowledge. The foundation of the hierarchy is basic, scientific knowledge. This knowledge then gets mixed with more applied or concrete types of knowledge that are connected to specific problems that need to be solved. The way that many educators teach in universities essentially follows this model. First, students attend lectures that cover the theory, then they subsequently "apply" the knowledge through labs or assignments. This model makes sense if design really is an application of basic knowledge to specific problems. It fits squarely within the positivist landscape of the majority of STEM disciplines. The view of the cognitive work that is taking place also aligns with Simon's view of design, where designers are engaging in heuristic search activities in an attempt to select and apply a good combination of components from the solution space. If this model is not accurate, however, the pedagogical approach may not be so effective. Plenty has been written on the topic of removing lectures and making classes more "active," often by using "flipped" modes of instruction and various "hands-on" activities within the classroom. These approaches may make instructors feel modern, but if they are not appropriate for the intended activity the students are training for, they are ultimately not very useful. Sometimes lectures are the appropriate mode of instruction, especially if students really do simply need to learn basic, abstract knowledge. The important question is what kind of activity is design? Is it essentially a form of application or is there more to it? And what is the appropriate mode of instruction for training designers?

I believe the contents of this chapter have demonstrated that visualization design is not only about application (although it may involve application), and it is not only a rational problem solving activity (although it may involve it too). Explicating a pedagogy of data visualization is much beyond the scope of this chapter, but a few points can be made here. If data visualization design is not simply about application, if its cognitive aspects involve making, expansive thinking, and the envisioning of future worlds, and if its activity is fundamentally transactional in nature, then students need to be supported in these activities with the right instructional models. One such pedagogical model is that of the design studio, a model that has a long history in art and design. Design studios position the role of an instructor as more of a coach than a lecturer, where students learn by doing, and the coaches provide demonstrations, critiques, and just-in-time instruction as means of formative feedback. Studio pedagogy tends to be more constructionist-oriented, aligning with the reflective practice view of design and not as strongly with the application view (although, as discussed in Sect. 8.3, it is important to remember that these are not mutually exclusive views in their entirety). In the studio, students engage with the complexity and messiness of design, typically relying more on trial-and-error and just-in-time learning than repetition and reinforcement toward the correct application of abstract principles. The goal of studio pedagogy is to prepare students to handle the complexity, uncertainty, and messiness of real-world practice rather than providing them with prescriptive procedures to follow or abstract theory to apply. In Schön's view, design students need tools for reflection that allow them to appropriately face each unique design situation with all its complexity and richness under consideration. Based on this, Stolterman [75] suggests that design education should focus on training designers to be "prepared-for-action" and not "guided-in-action."

Here the implications of adopting one paradigm of design cognition should again be apparent. If the cognitive acts of designing are akin to rational problem solving, where logic and procedural thinking are valued, the design studio and all of its messiness may not be as effective as a traditional lecture and lab format. But if the cognitive acts of designing are more transactional in nature, where designers need to "converse" with the materials of a particular situation, and receive just-in-time instruction and critique, the design studio appears to be a much better fit. Although much more can be said about various approaches to design pedagogy, such discussion goes beyond the scope of this chapter (see [6, 7, 28, 29, 38, 63, 82] for more depth on studio-based pedagogy in different design disciplines).

8.5 Summary

Design cognition has not received much attention in the visualization literature. In this chapter, I introduced some aspects of design cognition and described their relevance for data visualization. The discussion has been built largely on the two historically dominant paradigms of rational problem solving and reflective practice. These paradigms provide different pictures of what design cognition is, and I have attempted to describe implications for four data visualization topics: defining, automating, modeling, and teaching data visualization design. I posit that the visualization community embraces the rational problem solving view of design more than the reflective practice one, although future research can investigate this claim in more detail. Embracing the reflective practice paradigm may lead to more comprehensive knowledge of data visualization design and open new opportunities for research and teaching. However, even without embracing a particular paradigm, any focus on design cognition will likely be valuable for better understanding and teaching data visualization design.

References

1. R. S. Adams, J. Turns, and C. J. Atman. Educating effective engineering designers: The role of reflective practice. *Design Studies*, 24(3):275–294, 2003.
2. O. Akin. Variants in design cognition. In C. Eastman, W. Michael McCracken, and W. C. Newstetter, editors, *Design Knowing and Learning: Cognition in Design Education*, page 105–124. Elsevier, 2001.
3. S. Alspaugh, N. Zokaei, A. Liu, C. Jin, and M. A. Hearst. Futzing and moseying: Interviews with professional data analysts on exploration practices. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):22–31, Jan 2019.
4. C. J. Atman, R. S. Adams, M. E. Cardella, J. Turns, S. Mosborg, and J. Saleem. Engineering design processes: A comparison of students and expert practitioners. *Journal of Engineering Education*, 96(4):359–379, 2007.
5. L. J. Ball and B. T. Christensen. Advancing an understanding of design cognition and design metacognition: Progress and prospects. *Design Studies*, 65:35–59, Nov 2019.
6. E. Blevis, Y.-k. Lim, E. Stolterman, T. V. Wolf, and K. Sato. Supporting design studio culture in HCI. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, pages 2821–2824, 2007.
7. E. Boling, R. A. Schwier, C. M. Gray, K. M. Smith, and K. Campbell. *Studio teaching in higher education: Selected design cases*. Routledge, 2016.
8. R. Buchanan. Wicked problems in design thinking. *Design Issues*, 8(2):5–21, 1992.
9. J. M. Carroll. Scenarios and design cognition. In *Proceedings IEEE Joint International Conference on Requirements Engineering*, page 3–5, 2002.
10. P. J. Cash, B. J. Hicks, and S. J. Culley. A comparison of designer activity using core design situations in the laboratory and practice. *Design Studies*, 34(5):575–611, Sep 2013.

11. C. L. Cramer-Petersen, B. T. Christensen, and S. Ahmed-Kristensen. Empirically analysing design reasoning patterns: Abductive-deductive reasoning patterns dominate design idea generation. *Design Studies*, 60:39–70, 2019.
12. N. Crilly. Creativity and fixation in the real world: A literature review of case study research. *Design Studies*, 2019.
13. D. P. Crismond. *Investigate-and-Redesign Tasks as a Context for Learning and Doing Science and Technology: A study of naive, novice and expert high school and adult designers doing product comparisons and redesign tasks*. PhD thesis, Harvard University, 1997.
14. N. Cross. Science and design methodology: A review. *Research in Engineering Design*, 5(2):63–69, June 1993.
15. N. Cross. Designerly ways of knowing: Design discipline versus design science. *Design Issues*, 17(3):49–55, 2001.
16. A. Damle and P. J. Smith. Biasing cognitive processes during design: the effects of color. *Design Studies*, 30(5):521–540, Sep 2009.
17. S. B. Davis and O. Vane. Design as externalization: Enabling research. *Information Design Journal*, 25(1):28–42, Dec 2019.
18. A. Dong, M. S. Kleinsmann, and F. Deken. Investigating design cognition in the construction and enactment of team mental models. *Design Studies*, 34(1):1–33, Jan 2013.
19. A. Dong, D. Lovallo, and R. Mounarath. The effect of abductive reasoning on concept selection decisions. *Design Studies*, 37:37–58, Mar 2015.
20. K. Dorst. Exploring the structure of design problems. In *International Conference on Engineering Design (ICED)*, page 10, 2003.
21. K. Dorst. *Frame Innovation: Create New Thinking by Design*. MIT Press, Mar 2015.
22. K. Dorst and J. Dijkhuis. Comparing paradigms for describing design activity. *Design Studies*, 16(2):261–274, Apr 1995.
23. J. Drucker. Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(1):1–21, 2010.
24. C. M. Eastman. New directions in design cognition: Studies of representation and recall. In *Design Knowing and Learning*, page 147–198, 2001.
25. N. Goodman. *Ways of Worldmaking*. Hackett Publishing, 1978.
26. S. Goodwin, J. Dykes, S. Jones, I. Dillingham, G. Dove, A. Duffy, A. Kachkaev, A. Slingsby, and J. Wood. Creative user-centered visualization design for energy analysts and modelers. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2516–2525, 2013.
27. C. M. Gray, C. Dagli, M. Demiral-Uzan, F. Ergulec, V. Tan, A. A. Altuwaijri, K. Gyabak, M. Hilligoss, R. Kizilboga, K. Tomita, and et al. Judgment and instructional design: How ID practitioners work in practice. *Performance Improvement Quarterly*, 28(3):25–49, 2015.
28. C. M. Gray, P. Parsons, A. L. Toombs, N. Rasche, and M. Vorvoreanu. Designing an aesthetic learner experience: UX, instructional design, and design pedagogy. *International Journal of Designs for Learning*, 11(1):41–58, 2020.
29. S. Greenberg. Embedding a design studio course in a conventional computer science program. In *Creativity and HCI: From experience to design in education*, pages 23–41. Springer, 2009.
30. A. Hatchuel. Towards design theory and expandable rationality: The unfinished program of Herbert Simon. *Journal of Management and Governance*, 5(3–4):260–273, 2002.
31. N. V. Hernandez, J. J. Shah, and S. M. Smith. Understanding design ideation mechanisms through multilevel aligned empirical studies. *Design Studies*, 31(4):382–410, Jul 2010.
32. K. Höök and J. Löwgren. Characterizing interaction design by its ideals: A discipline in transition. *She Ji: The Journal of Design, Economics, and Innovation*, 7(1):24–40, 2021.
33. H. Jiang and C.-C. Yen. Protocol analysis in design research: a review. In *Proceedings of the International Association of Societies of Design Research Conference (IASDR 2009)*, 2009.
34. E. A. Kinsella. Constructivist underpinnings in Donald Schön's theory of reflective practice: echoes of Nelson Goodman. *Reflective Practice*, 7(3):277–286, 2006.
35. B. Lawson. The context of mind. *Designing in context*, page 133–148, 2001.
36. B. Lawson. *How Designers Think*. Routledge, 2006.
37. B. R. Lawson. Cognitive strategies in architectural design. *Ergonomics*, 22(1):59–68, 1979.

38. P. Little and M. Cardenas. Use of "studio" methods in the introductory engineering design curriculum. *Journal of Engineering Education*, 90(3):309–318, 2001.
39. J. Löwgren. Applying design methodology to software development. In *Proceedings of the conference on Designing interactive systems processes, practices, methods, & techniques - DIS '95*, pages 87–95, Ann Arbor, Michigan, United States, 1995. ACM Press.
40. J. Löwgren and E. Stolterman. *Thoughtful interaction design: A design perspective on information technology*. MIT Press, 2004.
41. J. Mackinlay, P. Hanrahan, and C. Stolte. Show me: Automatic presentation for visual analysis. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1137–1144, 2007.
42. J. D. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics*, 5(2):110–141, 1987.
43. X. Mao, O. Galil, Q. Parrish, and C. Sen. Evidence of cognitive chunking in freehand sketching during design ideation. *Design Studies*, 67:1–26, Mar 2020.
44. N. McCurdy, J. Dykes, and M. Meyer. Action design research and visualization design. In *Proceedings of the 6th Biannual Workshop on evaluation and BEyond - methodoLogIcal approaches for Visualization (BELIV)*, page 10–18. ACM Press, 2016.
45. S. McKenna, D. Mazur, J. Agutter, and M. Meyer. Design activity framework for visualization design. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2191–2200, 2014.
46. M. Meyer and J. Dykes. Reflection on reflection in applied visualization research. *IEEE Computer Graphics and Applications*, 38(6):9–16, 2018.
47. M. Meyer and J. Dykes. Criteria for rigor in visualization design study. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):87–97, 2020.
48. M. Meyer, M. Sedlmair, P. S. Quinan, and T. Munzner. The nested blocks and guidelines model. *Information Visualization*, 14(3):234–249, 2015.
49. V. O. Mittal, J. D. Moore, G. Carenini, and S. Roth. Describing complex charts in natural language: A caption generation system. *Computational Linguistics*, 24(3):431–467, 1998.
50. A. V. Moere and H. Purchase. On the role of design in information visualization. *Information Visualization*, 10(4):356–371, Oct 2011.
51. D. Moritz, C. Wang, G. L. Nelson, H. Lin, A. M. Smith, B. Howe, and J. Heer. Formalizing visualization design knowledge as constraints: Actionable and extensible models in Draco. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):438–448, Jan 2019.
52. T. Munzner. A nested model for visualization design and validation. In *IEEE Transactions on Visualization and Computer Graphics*, volume 15, page 921–928, 2009.
53. H. G. Nelson and E. Stolterman. *The Design Way: Intentional Change in an Unpredictable World*. The MIT Press, second edition, 2012.
54. A. Newell and H. A. Simon. *Human Problem Solving*. Prentice-Hall Englewood Cliffs, NJ, 1972.
55. P. Parsons. Understanding data visualization design practice. *IEEE Transactions on Visualization and Computer Graphics*, 28(1), 2022.
56. P. Parsons, C. M. Gray, A. Baigelenov, and I. Carr. Design judgment in data visualization practice. In *IEEE Visualization Conference (VIS), short papers*, Sep 2020.
57. P. Parsons and P. Shukla. Data visualization practitioners' perspectives on chartjunk. In *IEEE Visualization Conference (VIS), short papers*, Sep 2020.
58. P. Parsons, P. Shukla, and C. Park. Fixation and creativity in data visualization design: Experiences and perspectives of practitioners. In *IEEE Visualization Conference (VIS), short papers*, 2021.
59. R. S. Perez and C. D. Emery. Designer thinking: How novices and experts think about instructional design. *Performance Improvement Quarterly*, 8(3):80–95, 1995.
60. M. Polanyi. *Personal knowledge: Towards a post-critical philosophy*. University of Chicago Press, 1958.
61. M. Polanyi. *The Tacit Dimension*. University of Chicago Press, 1966.
62. A. Purcell and J. S. Gero. Design and other types of fixation. *Design Studies*, 17(4):363–383, Oct 1996.

63. Y. J. Reimer and S. A. Douglas. Teaching HCI design with the studio approach. *Computer Science Education*, 13(3):191–205, Sep 2003.
64. H. W. Rittel and M. M. Webber. Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169, 1973.
65. N. F. Roozenburg and K. Dorst. Describing design as a reflective practice: Observations on Schön's theory of practice. In *Designers*, pages 29–41. Springer, 1998.
66. D. A. Schön. *The Reflective Practitioner: How Professionals Think In Action*. Basic Books, 1983.
67. D. A. Schön. Designing: Rules, types and worlds. *Design Studies*, 9(3):10, 1988.
68. D. A. Schön. Designing as a reflective conversation with the materials of a design situation. *Research and Engineering Design*, 3(3), 1992.
69. K. Sedig and P. Parsons. Design of visualizations for human-information interaction: A pattern-based framework. *Synthesis Lectures on Visualization*, 4(1):1–185, 2016.
70. M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2431–2440, 2012.
71. B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pages 1–7, 2006.
72. H. A. Simon. *Administrative Behavior*. Macmillan Inc., 1947.
73. H. A. Simon. *The Sciences of the Artificial*. MIT Press, 1969.
74. H. A. Simon. The structure of ill structured problems. *Artificial intelligence*, 4(3–4):181–201, 1973.
75. E. Stolterman. The nature of design practice and implications for interaction design research. *International Journal of Design*, 2(1):55–65, 2008.
76. C. Stones and T. Cassidy. Seeing and discovering: how do student designers reinterpret sketches and digital marks during graphic design ideation? *Design Studies*, 31(5):439–460, Sep 2010.
77. L. A. Suchman. *Plans and Situated Actions: The Problem of Human-Machine Communication*. Cambridge University Press, 1987.
78. A. H. Vera and H. A. Simon. Situated action: A symbolic interpretation. *Cognitive science*, 17(1):7–48, 1993.
79. W. Visser. Two functions of analogical reasoning in design: a cognitive-psychology approach. *Design Studies*, 17(4):417–434, 1996.
80. W. Visser. Designing as construction of representations: A dynamic viewpoint in cognitive design research. *Human–Computer Interaction*, 21(1):103–152, Mar 2006.
81. W. Visser. Design: one, but in different forms. *Design Studies*, 30(3):187–223, 2009.
82. M. Vorvoreanu, C. M. Gray, P. Parsons, and N. Rasche. Advancing UX education: A model for integrated studio pedagogy. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 1441–1446. ACM, 2017.
83. K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, and J. Heer. Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):649–658, 2015.
84. K. Wongsuphasawat, D. Moritz, A. Anand, J. Mackinlay, B. Howe, and J. Heer. Towards a general-purpose query language for visualization recommendation. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, HILDA '16. ACM, 2016.

Chapter 9
Visualization Psychology: Foundations for an Interdisciplinary Research Program

Amy Rae Fox and James D. Hollan

Abstract What might a discipline of Visualization Psychology look like? If research on the psychological aspects of visualization were to coalesce, in the sense of a Lakatosian research program, what refutation-resistant theoretical commitments would magnetize its "hard core"? In this chapter, we argue that any interdisciplinary inquiry concerned with psychological aspects of visualization should situate its phenomena in the broader context of external representation, as a (triadic) semiotic activity achieved via information processing in a distributed cognitive system.

9.1 Introduction

Our goal in this chapter is *not* to provide a grand unified theory of visualization, nor to review all relevant work in the social and behavioral sciences. Rather, we offer a conceptual framework: a series of theoretical premises we argue should form the foundation of any interdisciplinary inquiry concerned with psychological aspects of visualization. We start by addressing the virtue of a hypothetical Visualization Psychology, arguing that the phenomenon of visualization is a fertile laboratory for exploring human cognition, that engineering and design-driven research can be improved via appropriate grounding in theories of perception and cognition, and that well-structured collaborations across disciplinary boundaries can foster a virtuous cycle beneficial to both traditions of research. In Sect. 9.3, we argue that such inquiry should situate visualization in the broader context of external representation (Sect. 9.3.1) as a (triadic) semiotic activity (Sect. 9.3.2) involving information processing (Sect. 9.3.3) in a distributed cognitive system (Sect. 9.3.4). In Sect. 9.4, we illustrate how this framework can be applied in both empirical and

A. R. Fox (✉) · J. D. Hollan
University of California San Diego, La Jolla, CA, USA

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_9

theoretical contexts, before concluding with a discussion of the role of Psychology in the history (and future) of Visualization.

9.2 Why Visualization Needs Psychology

The first workshop on Visualization Psychology was held in conjunction with the IEEE VIS conference in 2020, with the following Call for Papers (CFP) [71]:

> Before 2010, each VIS conference typically featured 0–2 papers on empirical studies. The VisWeek 2010 in Salt Lake City became a turning point, and since then more and more empirical study papers have been presented at VIS. Between 2016 and 2019, there were some 60 empirical study papers in VIS/TVCG tracks. Many young talents who are knowledgeable in both VIS and psychology emerged in the VIS community, while many colleagues in psychology are authoring and co-authoring such papers and attending VIS conferences. It is therefore timely to ask the two communities: is there a need for Visualization Psychology as a new interdisciplinary subject?
>
> There are many branches of applied psychology, such as clinical psychology, counseling psychology, educational psychology, forensic psychology, health psychology, industrial–organizational psychology, legal psychology, media psychology, music psychology, occupational psychology, sports psychology, and so on. Almost all of these are widely recognized academic subjects and have their own conferences and journals. Since interactive visualization and visual analytics encompass most human-centric processes in data science and real-world data intelligence workflows, many will argue for the necessity and feasibility for developing Visualization Psychology in a coherent and organized manner.
>
> This is the first workshop that will enable the experts in VIS and psychology to define the scope of this new subject of Visualization Psychology collectively and stimulate new research directions and activities in both fields. The goals of the workshop are:
>
> - To provide researchers in VIS with a significant platform to develop their theories and experiments in addition to acquiring knowledge from psychology
> - To broaden the scope of empirical research in VIS to involve more cognitive aspects in addition to considering visualization a vision or perception problem
> - To enable researchers in psychology to explore VIS as a rich playground and carry out research beyond the existing molds

(continued)

- To enable the development of the young talents in VIS and psychology through the development of a new interdisciplinary subject and the platforms for research communication, publications, and collaboration

This CFP solicits an intersection between two communities: “Psychology” and “VIS.” In this context we can pragmatically identify the VIS community as scholars affiliated with publishing venues such as the VIS conference series[1] and the journal IEEE Transactions on Visualization and Computer Graphics. This is a community intellectually and institutionally grounded in the discipline of Computer Science, closely related to (for others, a subset of) Human–Computer Interaction (HCI). The term “visualization research” then is used to refer to work in the VIS community, historically centered in engineering and design perspectives: developing tools to solve problems.

Psychology is a much older, more expansive discipline of science, dating back to the mid-nineteenth century. Let us say for the sake of argument that our goal is to engage scholars of psychology already publishing in VIS venues and others whose work is sufficiently relevant to visualization phenomena. For this, we might constrain “the psychology community” to be scholars of the cognitive, perceptual, or educational branches of experimental psychology, as well as vision science, learning science and cognitive science, whose phenomena of interest include human interaction with visual-spatial representations of information.[2] Psychological research in this sense will include work published in venues outside VIS (such as journals and conferences of the Cognitive Science Society, Psychonomic Society, Association for Psychological Science, and International Society for the Learning Sciences, among others). For brevity, we use the term psychologist as a placeholder for members of this more diverse disciplinary milieu.

What might the goals of this new interdisciplinary community be? The following claims are made explicit in the VisPsych CFP and offer a first approximation of what a Visualization Psychology might hope to accomplish:

1. Visualization research should be informed by psychological theories.
2. Visualization research should emphasize cognitive as well as perceptual factors.
3. Visualization phenomena offer a rich playground for further developing psychological theory.

[1] Self-identified as the “premier forum for advances in visualization and visual analytics,” VIS is sponsored by the IEEE (The Institute of Electrical and Electronics Engineers) Computer Society and Technical Committee (special interest group) on Visualization and Graphics (TCVG).

[2] One might also find research detailing interaction with graphics in other applied branches of Psychology—the use of multimedia graphics in the courtroom, for example—however the theories, models, and frameworks governing the basic science of such occurrences would likely come from cognitive, educational, or perceptual psychology.

The Role of Psychological Theory We suggest that the first point is true by virtue of epistemic relevance: the explanatory power and design impact of visualization research is improved when grounded in psychological theory, just as human interaction with a computer is better explained by theories of human psychology than formalisms governing the algorithms of the machine. For brevity, we use the term *psychological* as an umbrella for human aspects of interaction with visualizations; for example, how a reader perceives, forms a judgment from, or solves a problem with a visualization. This is in contrast to non-psychological questions, such as defining the algorithm for transforming a set of data into a particular representational form or how that computational system is engineered to afford input/output interaction. The latter questions may be required to enable the visualization phenomenon but neither necessitate nor explain human interaction with it. In this way research in visualization is like research in human–computer interaction. Psychological theories are needed to inform the design and evaluation of computational systems and to understand the dynamics of human interaction with them, but so too are contributions from the formal/mathematical science and engineering of computing. This is to say that VIS need not be subsumed *into* Psychology. Like HCI, visualization is a rich theoretical and empirical subject matter for interdisciplinary collaboration.

Situating Perception and Cognition The claim that empirical research in visualization should include cognitive in addition to perceptual theory is also trivially true, insofar as we are concerned with "cognitive" phenomena or behavior (i.e., beyond perceptual judgments). This is a question of levels of analysis and scope of phenomena. More often than not, empirical research in VIS (particularly investigations that center on the efficacy of some type of visualization or interactions with a visualization system) should be concerned with *cognitive* rather than *perceptual* phenomena. Accepting that the theoretical boundaries between perception and cognition are fuzzy, if we adopt an information-processing perspective from mainstream Cognitive Science, we can reasonably construe perception as some subset of cognition, concerned with stimulus-driven behavior, while the term cognition implies "higher order" processing, the influence of prior knowledge, or "what one does with" perceptual input. An empirical study with a task operationalized to measure constructs approximating perceptual processing is likely aimed at building and testing basic theory in perception, rather than evaluating the efficacy of a particular visualization.

This point is exemplified by the widespread misapplication of classic graphical perception studies by Bell Labs statisticians William Cleveland and Robert McGill (see [18–20]). When presented as stimulus a simplified statistical graph (e.g., a pie or divided bar chart, each with two segments marked with a dot), experimental subjects were asked to indicate "what percentage the smaller is of the larger": a *perceptual judgment*. The accuracy of subjects' responses (with respect to mathematical ground-truth) was evaluated and used to derive a ranking of relative accuracy for graphical encodings. From these results, one could conclude it is more effective to represent the quantitative difference between two values as a bar chart, rather than a pie chart. Unfortunately, this work has been generalized by some to the design heuristic, "bar charts are better than pie charts." If humans were perceptual computers with no prior knowledge, expertise, beliefs, or other individual differences, that *might* be the end of the story. However, a body of research in *graph comprehension* has demonstrated that if you use a different task in your study, for example, asking the graph reader to extract a specific value from the graph, to use it to make a decision, or perhaps a forecast, then the accuracy rankings do not necessarily hold (e.g., [63, 64]). This apparent contradiction arises from the insight that different task-demands require different "readings" of a graph: a (perceptual) judgment of relative size is different than extracting a data value which is different from detecting a trend, and so on. The more complex the behavioral task, the more "higher order" (i.e., resource-intensive, implicating prior knowledge) processing is required. While it would be appropriate to apply perceptual accuracy heuristics to design, for example, a simple graphic in a newspaper illustrating the quantitative month-over-month change in some economic report, it would be insufficient to rely solely on these heuristics to guide design of an interactive visual analytics system. Perceptual guidance for achieving simpler tasks is a useful starting point but does not encompass knowledge-driven interactions. Basic research on graph comprehension has clearly demonstrated that the effectiveness of graphical encodings arises not from the interaction of data and forms, but rather, data, forms, *individuals*, and *tasks*. This is not to say that perceptual processing is not relevant to complex cognitive activities or that there are no perceptual questions left to be answered. One of the most challenging, and in our view promising, areas ripe for theoretical development is along these fuzzy boundaries: exploring the factors that govern how stimulus-driven and knowledge-driven processes are integrated to produce behavior.

Visualization and the Virtuous Cycle The claim that visualization phenomena offer opportunities for advancing psychological research can be demonstrated from evidence. Grammars and frameworks (especially [9, 54, 74]) designed by Computer Scientists and implemented as libraries and interactive systems have made computer-based data visualizations accessible for researchers as tools for data analysis and presentation. For those whose research involves empirical study of human–information interaction, these are also tools for generating stimuli. The situations in which the stimuli might be used—for example, studying how a graph is used to make a decision, how a student leverages a chart and accompanying text to learn a concept, or how an analyst uses an interactive system to make a forecast—are all enabled by technologies borne of Computer Science-based visualization and computer graphics research. In turn, research on the human aspects of how and why and to what effect individuals interact with visualizations provides guidance for the appropriate design of visualization systems. Technology inspires new human activity, which offers the psychologist new subjects of inquiry. As is often the case with technology-driven endeavors, Psychology and Computer Science stand in relationship as a virtuous cycle: a positive feedback loop where progress in each stands to both improve in quality and volume the progress of the other. The relationship between Psychology and VIS is the relationship between Psychology and broader Human–Computer Interaction—appropriate considering that VIS grew from and is largely considered a part of HCI. There is, in fact, so much "psychology" *in* HCI and VIS and it is challenging to know where (and if) we should draw meaningful boundaries. We return to this issue in Sect. 9.5.

9.3 Elements of a Framework

> Lakatos's idea is to construct a methodology of science, and with it a demarcation criterion, whose precepts are more in accordance with scientific practice. (...) Instead of an individual falsifiable theory which ought to be rejected as soon as it is refuted, we have a sequence of falsifiable theories characterized by a shared hard core of central theses that are deemed irrefutable—or, at least, refutation-resistant—by methodological fiat. This sequence of theories constitutes a research program.—Musgrave and Pigden [44]

One way to conceptualize the structure of a Visualization Psychology is in terms of a *research program* in the tradition of post-positivist philosopher of science of Imre Lakatos [76]. Lakatos was skeptical of Kuhn's normative conception of science as progressing via successive stages where one research paradigm (i.e., a framework for approaching one's subject matter) is replaced by another. Lakatos characterized the practice of science as altogether messier, with multiple competing paradigms operating in parallel, in nonlinear cycles of progression (making theoretical and empirical progress) and degeneration (stagnating, and/or questioning core claims). For Lakatos, a research program was characterized not by a singular method, model, or theory, but rather a collection of basic (and by convention irrefutable) assumptions shared by its community. This *hard core* of theoretical commitments is surrounded by a protective *auxiliary belt* of hypotheses that constitute the work of science. Investigators rely on the shared language and lenses of the hard core to generate hypotheses in the auxiliary belt that might be shaped into theories or broken down and replaced. Progress is made so long as the auxiliary belt grows: theoretically, by extending the scope of theory to new empirical domains, or empirically, by finding corroborating evidence for theoretical claims (see [3, 76]).

What central theses—theoretical propositions resistant to refutation—might a Visualization Psychology have at its *hard core*? We propose a minimum of four central tenets. Individually, these ideas are not falsifiable theories, but rather perspectives and frameworks that have arisen from and give rise to empirically testable hypotheses.

1. **Visualization is external representation.** Visualization (as artifact) and visualization (as process) belong to the broader class of *external representation*.
2. **Meaning is constructed.** Interacting with a visualization is not a passive transmission of meaning (e.g., "extracted" from the artifact), but rather an active, interpretive semiotic process where knowledge is constructed.

3. **Information is processed.** Visualization is most effectively construed as the transmission of information across components of a system, via transformation between representational states.
4. **Cognition is distributed.** Intelligent action with a visualization is a function of a distributed cognitive system comprised of human actors and material artifacts situated in relation to their spatio-temporal environment.

We describe these perspectives in Sects. 9.3.1–9.3.4 and in Sect. 9.4 demonstrate how they can be applied in both empirical and theoretical research settings.

9.3.1 Visualization is External Representation

The language of representation is slippery and self-referencing. Shown a collection of marks on surfaces, you might label some as art, or pictures, others as diagrams, maps, or schematics, some charts, plots, or graphs, and others also as graphs, but you might use air quotes and call them "graph-theory graphs." Some you will identify as writing and others, like writing but not—some peculiar or particular system of notation. The linguistic labels you apply to each marking likely depend on your disciplinary background and are neither exhaustive nor mutually exclusive. Which of these are visualizations? (Fig. 9.1).

9.3.1.1 *On* Visualization

Let us start with definitions put forth in prominent VIS texts. In their venerated compilation of papers and essays, Card et al. [11] define Information Visualization as *"The use of computer-supported, interactive, visual representations of abstract data to amplify cognition"* (pg. 7). Stephen Few offers a functional definition, characterizing data visualization as *"an umbrella term to cover all types of visual representations that support the exploration, examination, and communication of data. Whatever the representation, as long as it's visual, and whatever it represents, as long as it's information, this constitutes data visualization"* [24, pg.12].

Such inclusive specifications may be effective for teaching but are less suitable guides for scientific inquiry. From this heuristic, we might conclude the words on this page constitute a visualization—but they would not be considered so by most visualization practitioners. Why? Because visualizations are somehow *graphic* in nature; from Ware, *"a graphical representation of data or concepts"* [72, pg.2]. Ware highlights how the term has transitioned in conventional meaning from "constructing a visual image in the mind" to *"an external artifact supporting*

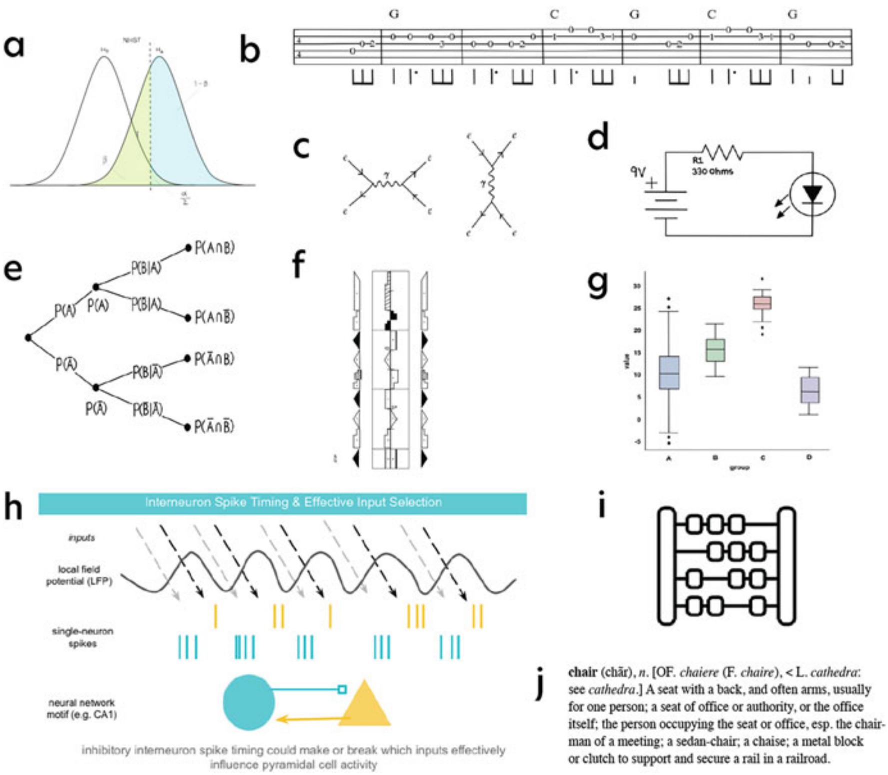

Fig. 9.1 A group of visual-spatial external representations. (**a**) a conceptual diagram indicating key concepts in null hypothesis significance testing; (**b**) portion of the song 'You Are My Sunshine' in guitar tabs notation; (**c**) Feynman diagram for an interaction between an electron and anti-electron with exchange of a photon; (**d**) schematic of a circuit depicting a 9V battery in configuration with a single resistor and LED; (**e**) tree diagram used in solving Bayesian reasoning problems; (**f**) Laban notation representing a ballet exercise; (**g**) boxplot depicting mean, interquartile range and outliers for 4 groups; (**h**) a figure from a neuroscience presentation that combines multiple representations of related phenomena to orient readers to both the research method and analysis of results; (**i**) an icon of an abacus-note that the object the icon represents would also be considered an external representation of number; (**j**) image of the words in a dictionary definition of the word chair (inspired by the conceptual art piece 'One and Three Chairs' by Joseph Kosuth)

decision making." Bertin (in English translation [5]) also refers to graphic representation, distinguishing this from relational imagery (e.g., art, photography) and mathematics (e.g., symbolic notation). In a more recent text, Munzner pragmatically characterizes the purpose of *computer-based visualization systems* as providing *"visual representations of data sets designed to help people carry out tasks more effectively"* [43, pg.1].

If we start with the assertion that a visualization (noun) is a type of representation, then to arrive at a useful definition we should characterize the properties of what representations can be admitted to this type. An externally available representation is accessed via some sensory modality; that visualizations are subject to *vision* is the only shared property of the aforementioned definitions. Similarly, we lack clarity as to whether visualizations need to be *graphic* (or indeed, what being graphic entails), if they need to be interactive, generated by a computer, whether they can refer to any information, or only "data," and perhaps only that which is deemed "abstract." Alternatively, we can try to infer a shared conceptualization of such terms based on how a community organizes itself. VIS[3] explores the sometimes fluid distinction within these properties in the organization of annual conference tracks, paper types, and sessions. Whether the referent of a visualization is abstract data or has some physical/geometric invariant is the (historical) distinction between the Information Visualization (InfoVIS) and Scientific Visualization (SciVIS) conferences. If the purpose of an artifact is to support an interactive, analytical process, then it would likely be called a visualization and fall into the Visual Analytics (VAST) conference. If the referent is more "conceptual" than data-driven, research is more likely to be published outside of VIS, such as in the (multidisciplinary) *International Conference on the Theory and Application of Diagrams*, and if the use of the representation is primarily for learning, then the research is likely evaluated in either disciplinary education (e.g., Chemistry Education, Math Education) or Learning Science. The number of paper types (and submissions) at VIS implicating computer systems, prototypes, and algorithms suggests a strong preference toward the computer as a presentation medium or "physical substrate." Though there is exciting growth in the topic of data *physicalization* and exploration of alternative sensory modalities for representing data, this area has yet to emerge as a large enough topic to warrant its own conference session in the past five years. Research on data *sonification* or *tactilization* are more likely to be found in the broader ACM SIG-CHI or topical journal like ACM Transactions on Applied Perception.

Definitions, as terminology, serve as tools for communicating and conceptualizing one's subject matter [10]. We draw on these definitions of Information and Data Visualization, not in critique of their notable contributions, but rather to call attention to a puzzling inconsistency in the foundation of the field. Our objects of inquiry are altogether over-specified and under-defined. Which of the artifacts in Fig. 9.1 are visualizations? We argue that to the visualization psychologist, it should not really matter. They are *all* instances of the larger class: external representations. Just as psycholinguists are concerned with the psychological and neurobiological factors that enable humans to acquire, use, comprehend, and produce *language* (not English, or "languages using the roman alphabet," or "languages written from left-to-right"), visualization psychologists should be concerned with the factors that enable humans to make use of *external representations* (not just the "graphic,"

[3] Referring to the annual IEEE combined conferences on Information Visualization (InfoVIS), Scientific Visualization (SciVIS), and Visual Analytics Science and Technology (VAST).

"data-driven," or "computer generated" variety). In this sense, designers and engineers of visualization systems have the luxury of specialization. But insofar as we believe that the interaction with visualization relies on general purpose cognitive mechanisms, psychologists do not. To understand how these artifacts function—to study how they are used by humans to construct meaning in support of complex cognitive activities—we must climb up the ladder of abstraction.

9.3.1.2 *On* External Representation

> The power of the unaided mind is highly overrated. Without external aids, memory, thought, and reasoning are all constrained. But human intelligence is highly flexible and adaptive, superb at inventing procedures and objects that overcome its own limits. The real powers come from devising external aids that enhance cognitive abilities. (...) It is things that make us smart.—Norman [47, pg. 43]

The term *external representation* came to prominence in the late 1970–80s, as the new discipline of Cognitive Science emerged from information-processing psychology with a common focus on the existence and nature of mental representation (see [8, 38, 46, 49]). But when the researchers focused solely on the mental, they needed unnecessarily complex mechanisms to explain behavior. Although AI and the mental imagery debate would ensure that mental representation remained a focus of mainstream Cognitive Science, the need to distinguish *internal* from *external* meant the birth of a new research area.

The complexity of external representation, however, was not immediately appreciated. In his treatise on cognitive representation, Palmer argued that mental representations were *"exceedingly complex and difficult to study,"* so one might start with the examination of *"noncognitive"*[4] representations, as they are *"simple, and easy to study"*[5] [48, pg. 262]. Subsequent elaboration of representational systems demonstrated there is much to explore with respect to the nature and function of such "noncognitive" structures (see [37, 55]).

Like research on visualization, however, empirical work on external representation was lacking in the explicit definition of terms. A study on problem solving with a diagram might refer to the diagram as an external representation and rely on the reader to draw the same antonymic implication as Palmer: an external representation is a representation that is *not* internal. The sensory modality, encoding media, presentation substrate, and communicate purpose are left under-specified, allowing the term to serve as a category for *things that can be perceived, that refer to other things.* Such things might be presented via any medium, in any encoding structure,

[4] Palmer reserves the qualifier *cognitive* for internal representations, designating the external as "noncognitive." Following a distributed cognitive perspective, we would characterize *both* as cognitive representations and prefer the term "mental" to describe those representations not perceivable to others.

[5] More "accessible" is perhaps the more generous characterization.

via any sensory modality, referring to anything (real or imagined), for any purpose. Zhang and Norman explicitly described external representations as *"knowledge and structure in the world, as physical symbols (e.g., written symbols, beads of abacuses, etc.) or as external rules, constraints, or relations embedded in physical configurations (e.g., spatial relations of written digits, visual and spatial layouts of diagrams, physical constraints in abacuses, etc.)*" [77, pg.3].

We refine this definition:

> *An **external representation** (noun) is the form of information, purposefully encoded as structures in material artifacts that serve a semiotic function as part of an interpretive process.*

Information is encoded externally via forms and structures that can be described along a continuum of implicit to explicit, depending on how much effort, or inference, is required in their use (see [34, 35]). We remove reference to knowledge in the world, preferring the constructivist premise that knowledge does not exist *in* the environment but is actively constructed by the individual via interaction *with* their environment. The kinds of constraints and structures described by Zhang and Norman are constituent parts of representations and of how they work. Most importantly, we clarify the scope of external representations as being constructed by some actor, for some purpose, thus grounding external representation in the context of communication—though broadly construed. (Many of the external representations we construct are meant for communication not with others, but our future selves.) Here, we admit visualization as a subset of external representation: an active construction of meaning via the exchange of information between actor and artifact. What is crucial is that we orient ourselves equally toward the artifact and the interactive process: representation as noun and representation as verb.

Despite a dearth of precise terminology in the proceeding decades, researchers[6] took up the challenge of discovering how humans *think with things*, studying how various forms of external representations (ERs) influence thinking for various ends. An early focus was the fashion in which graphic/diagrammatic ERs influence thinking in contrast to natural language, such as in problem solving [37, 77], learning [2], design [21], and scientific discovery [14]. Distinctions were drawn between encoding structures: the sentential/propositional (symbols), and graphic/diagrammatic (images), where the latter class was taken up by its own interdisciplinary community

[6] Particularly in Cognitive Science, Learning Science/Educational Psychology, and disciplinary education like Math, Chemistry, and Physics.

in the early 2000s.[7] Educational Psychologists and Learning Scientists turned their attention to multimodal and multimedia representations [41, 56]. A particularly impactful contribution was made by Michael Scaife and Yvonne Rogers in [55], wherein they proposed a "new agenda" for research on graphical representations and in considerable detail and sophistication demonstrate how such research promises to improve the design of future technologies while simultaneously advancing theories of cognition. By the late 2000s, sufficient interest across allied disciplines warranted a special issue of the journal TopiCS in Cognitive Science, dedicated to visual-spatial representations, with milestone contributions on visual analytics [25], graph comprehension [62], and diagrams [13], as well as reviews of how visual-spatial representations serve as tools for thinking [70] and corresponding implications for design [28]. These are indicative of the work we believe should be at the theoretical core of visualization psychology research.

Thus, we have moved from the study of computer-generated interactive data graphics to any externalization of thought. What we are left with, it seems, is a Goldilocks problem. The idiomatic conception of visualization is too narrow and an exhaustive conception of external representation too broad. Fortunately, there are dimensions along which this metaphorical problem space can be surveyed. We might think of these dimensions as ranges along which we can attune our attention, progressively expanding or narrowing our scope of inquiry depending on the state of theoretical and technological advancement.

On Encoding Medium Though we have noted the lack of precision in defining the scope of visualizations, there has been no lack of effort in cataloging [27] and taxonomizing them, from general descriptive frameworks [6, 15, 50, 65, 69] to those concerned with specific domains of data [1, 4, 7]. Two particularly useful (and under-appreciated) are those of Engelhardt [23] who offers an atomic, generative framework deserving of its characterization as a language of graphics and Massironi [40] who offers both a taxonomy and an evolutionary timeline. While most taxonomies deal with some intersection of graphical structure and data type (e.g., geographic maps, relational networks), the more common distinction in the cognitive and learning science literature is the continuum from descriptive to depictive, roughly analogous with symbolic to analog, or propositional to graphic. These terms refer to a semiotic modality (also medium), which indicates the degree of convention in the relation between a representation and thing to which it refers. While the poles of a depictive–descriptive continuum can be easily identified, there lays betwixt a murky medium. At what point of abstraction does an icon become a symbol? When it is no longer identifiable as its referent without convention? In whose judgment? We are more accurate in describing our scope of inquiry as multimedia than "primarily graphic." We propose that while origins of VIS as a field lie in the distinction of graphics from text, fundamental questions about

[7] The International Conference on the Theory and Application of Diagrams is a biennial gathering held since 2000, attended by a cross-section of Philosophers, Psychologists, Mathematicians, and Computer Scientists.

framing, persuasion, and even comprehension rely on understanding the function of text *alongside* graphics. It is rarely the case that external representations of the visual graphic variety are not accompanied by some form of linguistic proposition or sentential notation. Indeed, a visualization without a title and labels may be worth no words at all.

On Sensory Modality External representations can be constructed for any sensory modality, though by far the most attention has been paid to the visual. Deservedly so, as visuals are the most pervasive information artifacts, and the sensory modality about which we have the most understanding. Though we are surely far from exhausting the wellspring of questions to be asked about visual representations, we suggest that we accept within our scope multi-sensory representation. From a theoretical stance, this requires broader inclusion of expertise across perceptual psychology, though the applications are consequential. In an increasingly visualization-driven world, equality and accessibility demand informationally equivalent tools for those without visual perception. Notably, we can trace this view back to the inception of visualization in HCI:

> It should be noted that while we are emphasizing visualization, the general case is for *perceptualization.* It is just as possible to design systems for information *sonification* or *tactilization* of data as for multiple perceptualizations. Indeed, there are advantages in doing so. But vision, the sense with by far the largest bandwidth, is the obvious place to start, and it would take us too far afield to cover all the senses here.—Card et al. [11, pg.7]

On Representational Purpose or *Communicative Context* VIS texts describe the purpose of visualization as being to "amplify" cognition [11, 24, 72] though research in Cognitive Science suggests the story is more nuanced (see [36, 47]). External representations *enable* cognition and can change the nature of the task we are performing. This is not to say that one cannot think without external representations, but rather, there are certain kinds of thinking that are not possible without the right representations to think them.

The most generic purpose is to simply record: to offload from internal memory to external cognition. In terms of communication, to inform—for example, the boxplot in a manuscript, where one aims to inform the reader of some aspects of the underlying information—in a clear and simple manner.[8] But one might design that artifact differently if one intends for you to explore the data, undertake an analysis, or make a decision, a plan, or a forecast. An author might change their strategy if they want to strongly persuade you or, alternatively, want you to use the representation to learn. There are entire systems of diagrams designed for solving particular kinds of problems, and the design of representations to support conceptual change is the focus of specific subdisciplines in STEM education. We use the term *communicative context* to refer to the "cognitive activities" the designer of a representation intends the user to perform. The structure of these activities

[8] Note that clarity and simplicity *do not* imply truth. The designer of a representation has a voice that is echoed in every design decision, from what information to include to how to encode it.

has not been taxonomized, though a compelling framework for their hierarchical, emergent structure is detailed by Sedig and Parsons [60]. The relevant insight is that certain parameters of a representation, such as the computational efficiency, or relative explicitness of certain aspects of data, need to be tuned in accordance with the task the reader is expected to perform. Bertin (in English translation [5, pg. 183]) writes *"A graphic is never an end in itself; it is a moment in the processes of decision making."* To this, we add *"... or reasoning, or learning, or problem solving, or sensemaking, or analyzing, or planning, or forecasting... "* The graphic in the moment is thus deeply intertwined with the individual, their situation, and task contexts.

9.3.2 Meaning Is Constructed

> All meaningful phenomena (including words and images) are signs. To interpret something is to treat it as a sign. All experience is mediated by signs, and communication depends on them.—Chandler [12, pg. 23]

If external representations are things purposefully constructed to refer to other things, then understanding their referential function falls squarely within the realm of *semiotics*. Semiotics is the study of signs, where a sign is construed as "something which stands for something else"—*aliquid stat pro aliquo* [12]. Note this is a larger class of phenomena than external representations which we have (pragmatically) constrained as being purposefully constructed. Signs, conversely, can be naturally occurring: a trail of footprints in the snow or mud puddles following a heavy rain. The crux of the semiotic puzzle is that to *be* a sign, is to be *interpreted*. Phenomena become signs when meaning is assigned to them. You may have the intuition that to implicate semiotics is to open a Pandora's box where terms like *represent* and *signify* become so complex they risk losing any consistent meaning—and you would be right.[9] Our task is to introduce the elementary constructs of a particular semiotic approach that can be productively applied to understanding the function of external representations in distributed cognitive systems.

Imagine you encounter a line graph in a newspaper. Your job as a reader is to develop an understanding (interpretant) of what the graph (the representamen) indicates about some state of the world (the referent). The terms referent, representamen, and interpretant are drawn from American philosopher Charles Sanders Pierce, and his general account of the relations that govern representation, reference, and the meaning of signs [30]. Peirce's basic claim is that a "sign" consists of three

[9] "No treatment of semiotics can claim to be comprehensive because, in the broadest sense (as a general theory of signs), it embraces the whole field of signification, including "life, the universe, and everything," regardless of whether the signs are goal-directed (or interpreted as being so)" [12, pg. xvi].

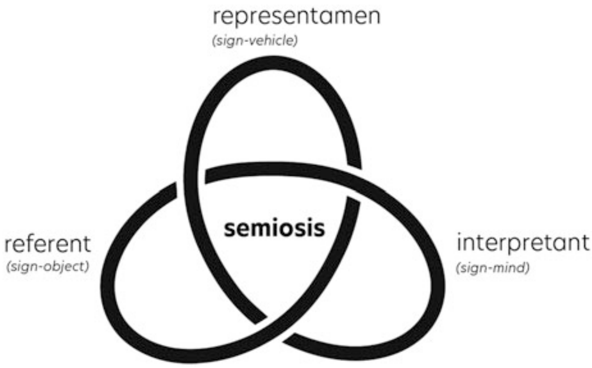

Fig. 9.2 The three components of a Peircean sign (referent, representamen, and interpretant) are irreducibly triadic

parts: (1) an object (referent) that is the thing being signified, (2) an element that signifies (representamen): that which does the referring, and (3) the interpretant: understanding that is made of the referent-representamen relation. Importantly, the entire triadic relation is referred to as a sign or representation and the dynamics of the relation semiosis or signification. Though Peirce's own terminology changed over the development of his ideas, to avoid confusion, we choose here three terms not commonly employed outside of semiotics: *referent* (also, sign-object, or signified), *representamen* (also, sign-vehicle, signifier), and *interpretant* (also, sign-mind, understanding) (see Fig. 9.2). The labels we colloquially apply to the material substances that comprise external representations—representation, sign—are in semiotic terms explicitly *not* equated with the material component of the sign. That is to say, the "representation" is not *the* representation, but only a *part* of it. The sign-relations are *irreducibly* triadic, and while we might for sake of analysis wish to isolate the relation between sign-object and sign-vehicle (for example, how a designer chooses to encode some information) or sign-vehicle and sign-mind (for example, how a reader interprets the encoding), their function is only constituted as a property of all three. This is perhaps more intuitive in psychological terms: constructing meaning is a combination of top-down (knowledge-driven) and bottom-up (stimulus-driven) interpretative processing. To examine how a reader interprets an encoding, we must consider their interaction with the encoding, and prior knowledge of the information being encoded.

Peirce's triadic semiotic is significant to the psychology of visualization in two ways. First, it makes explicit the constructive nature of meaning. Peirce's interpretant brings into the signifying function someone or something that does the interpreting: an intelligent process that constructs the translations between signifying elements of the representamen, in order to arrive at some approximation of the referent. In this way, the relation between the "thing" and the "representation" is not a direct and determined mapping, but entirely subjective, based on the interpretation of the observer. Second, Peirce's semiosis is dynamic, relying not on the entirety of that which acts as the representamen, but only on the elements relevant in signifying. Later accounts elaborate on subdivisions in the referent and interpretant that pertain to stages of processing in an unfolding chain of meaning [30]. This aspect has a distinctly cognitive appeal, as it suggests a distribution of meaning-making between the observer and environment; one that occurs via a process in time, not contained solely within artifacts or minds. In the context of cognition, together these features of Peirce's approach are consistent with what we know about the influence of prior knowledge and individual differences in the determination of meaning.

9.3.3 Information Is Processed

> "There is no information without information vehicles. Information vehicles are the carriers of information, the physical material in which the information-for-the-interpreter is encoded."—Nauta [45]

In an age of grounded, embodied, and extended cognition, it is rather fashionable to discount information-processing psychology as outdated. However, there is a difference between studying psychological phenomena *as* the processing of information and studying *the phenomenon* of information processing. The classical conception of information-processing regards the mind as a computational system manipulating symbols to enact representational states. The information-processing psychologist might seek to explain all psychological phenomena through this lens—behavior resulting from the propagation of representations, disregarding the influence of the body, modal systems, or environment. Contemporary theories that situate cognition beyond the mental are extraordinarily applicable to human interaction with external representations. But so too are some constructs from information processing. In a Visualization Psychology, we are directly concerned with how humans interact with information via representations. To the extent that we rely on the notion of information, we cannot escape the notion of its processing. Importantly, we are not proposing that to adopt an information-processing view of visualization requires commitment to a computational theory of mind, nor any strictly sentential/propositional symbol manipulation in the brain. One problem with information-processing models of cognition was that they paid "scant regard" to the external world of artifacts and information (see [53]). By exploring phenomena that require processing of multimedia (i.e., text and graphic) information, we expect that the Visualization Psychologists can improve on these theories by directly addressing the interface between external and internal information, especially in the construction of meaning.

9.3.4 Cognition Is Distributed

> It does not seem possible to account for the cognitive accomplishments of our species by reference to what is inside our heads alone. One must also consider the cognitive roles of the social and material world. But how shall we understand the relationships of the social and the material to cognitive processes that take place inside individual human actors? This is the problem that distributed cognition attempts to solve.—Hutchins [32, pg. 2071]

As behavioral scientists, we are concerned not only with the design and efficacy of external representations but also with their mechanisms: how and why they function (or not). These functions are enacted between the artifact(s) and person(s), embodied and situated in their environments and complex social structures. This complexity demands a distributed perspective of cognition, one that extends functions of the mind beyond the individual's skin and skull (see [16, 17]) and distributes them through time and space via material artifacts and members of society (see [31, 32]). Unlike traditional theories, *distributed cognition* extends the reach of what is considered "cognitive" beyond the individual to encompass interactions between people and with resources and materials in the environment.

The applicability of a distributed cognitive perspective to research in visualization [39] and human–computer interaction more broadly [29] has been successfully argued, and corresponding methods of cognitive ethnography are now widely accepted in VIS and HCI publications. Through cognitive ethnographic techniques (e.g., interviewing, participant observation, in-situ recording), a researcher can determine *what things mean* to the participants in an activity and to document *the means by which* these meanings are created. In this way, cognitive ethnography yields data for exploring cognitive mechanisms, while also feeding distributed cognitive theory by adding to the corpus of observed phenomena the theory should explain.

A distributed perspective on cognition is particularly relevant to the psychology of visualization because it not only provides an overarching framework for investigating representations and representational processes but actively encourages integration of ethnographic and experimental approaches as well. While the study of cognition *in the wild* can answer many kinds of questions about the nature of human cognition in real workplaces, the richness of real-world settings places limits on the power of observational methods. This is where well-motivated experiments are necessary. Having observed phenomena in natural settings, the researchers can set about designing more constrained experiments to systematically explore specific aspects of observed situated behaviors. Importantly, distributed cognition does *not* require that every aspect of a cognitive system be examined in every interaction: levels of analysis still apply. But a distributed cognitive perspective does require that the most highly operationalized inquiries of basic processes are contextualized as only parts of a more complex system of factors that taken together, explain behavior.

In every area of science and technology, the choices made about units of analysis have crucial consequences. Boundaries are often a matter of tradition in a field. Sometimes the traditionally assumed boundaries are exactly right for investigating a specific issue. For other phenomena, however, the boundaries either span too much or, more frequently, too little. The failure to reevaluate the unit of analysis as sciences advance and technology changes can fundamentally restrict development. A common critique of distributed cognition in psychological traditions is the *necessity* of extending the unit of analysis to the environment. From Wilson, for example, *"The fact that causal control is distributed across the situation is not sufficient justification for the claim that we must study a distributed system. Science is not ultimately about explaining the causality of any particular event. Instead, it is about understanding fundamental principles of organization and function"* [75, pg. 630]. We obviously disagree with this claim and argue that insofar as the function of the mind is to control real-time action in dynamic environments, any sufficient understanding of its organization requires theoretical and methodological approaches that directly address the environment as an active participant in cognition. Fortunately, today the lens of distributed cognition is part of an emerging zeitgeist that appreciates the central importance of closing the divide between computationally focused disciplines and disciplines concerned with understanding people and sociotechnical systems.

9.4 *On Doing* Visualization Psychology

We propose the following definition for *Visualization Psychology*:

> ***Visualization Psychology*** *is a scientific research program at the intersection of computing, behavioral and social sciences. It is characterized by the application of theories of perception, cognition and behavior to predict and/or explain the nature of human interaction with visualization systems, and by the use of visualization phenomena to inform theories of perception, cognition, and behavior.*

This definition emphasizes that (1) VisPsych should be a *scientific* endeavor: though it may involve close collaboration with designers and engineers, the intellectual goal of the research is generating knowledge, and (2) the flow of insight in VisPsych should be bidirectional: benefitting from and contributing to work in engineering or design-oriented aspects of visualization. Research in Visualization Psychology can contribute to the design and evaluation of visualization systems, while the design and engineering of visualization systems can provide sites of inquiry for both basic and applied psychological research. It has elements of both basic and applied science, employing basic theories to explain specific (visualization) phenomena, the outcomes of which may serve as data for (re)constructing basic theory. In this sense, much Visualization Psychology might be most accurately characterized as *use-inspired basic research*, à la Stokes [67] Pasteur's quadrant conception of a two-dimensional relation between basic and applied research. Finally, (3) we have purposefully characterized VisPsych as a *program* of research, rather than a singular discipline or distinct community.

According to Lakatos, the measure of a research program is its ability to make both theoretical and empirical progress. Theoretical progress is made by building on foundational tenets to develop theory and apply it to new empirical domains. Empirical progress is made by evaluating theory. Across the VIS and visualization-adjacent literature in psychology/learning science/cognitive science we find diverse examples of such progress.

Extending Theory to New Domains An example of theoretical progress can be found in the research agenda of Cognitive Psychologist Priti Shah, whose early research applied psychological theory from reading comprehension to the emerging topic of graph comprehension. Shah and colleagues built upon Pinker's [52] information-processing theory of graph comprehension by re-construing perception of a graph as *reading* and invoking constructs from the prevailing Construction-Integration [33] theory of text and discourse comprehension (see [61]). From this application came new testable hypotheses about the role of prior knowledge and individual differences in comprehension and the temporal dynamics of information processing in graph comprehension more broadly. Although this work was not explicitly situated in the context of distributed cognition or triadic semiotics, it is consistent with both lenses insofar as it situates graph comprehension as *discourse* between the designer and the reader of a visualization, differentially influenced by factors acting on the display, the individual, and task(s).

Evaluating Theory in New Domains A powerful example of the virtuous cycle between basic psychological and applied-visualization oriented research can be found in the recent movement to (re)connect research in visualization with vision science. While some of the earliest empirical work in visualization was concerned with visual perception (e.g., [18, 66]), modern interdisciplinary research in vision science offers both new methods (see [22]) and theoretical constructs. *Ensemble perception*, for example, refers to how the visual system extracts summary statistical information from groups of similar objects, ostensibly as a way of dealing with spatial and temporal processing constraints [73]. Szafir and colleagues have applied the theoretical arguments for ensemble perception to the domain of data visualization and argued for how it may serve as mechanism for some of the most common perceptual tasks we perform when interpreting visualizations, including identifying outliers, detecting trends, and estimating means [68]. While the development of ensemble perception as a construct is not necessarily grounded in distributed or semiotic perspectives, its application to visualization *is*: first, by providing an account for how differing interpretations can arise from the same visual stimuli (an explicit acknowledgement of the triadic nature of semiotic discourse), and second, by positioning the research as a *contribution* rather than *determinant* of design heuristics (an implicit acknowledgement that ensemble perception is a valuable piece rather than sole factor in the puzzle of visualization behavior).

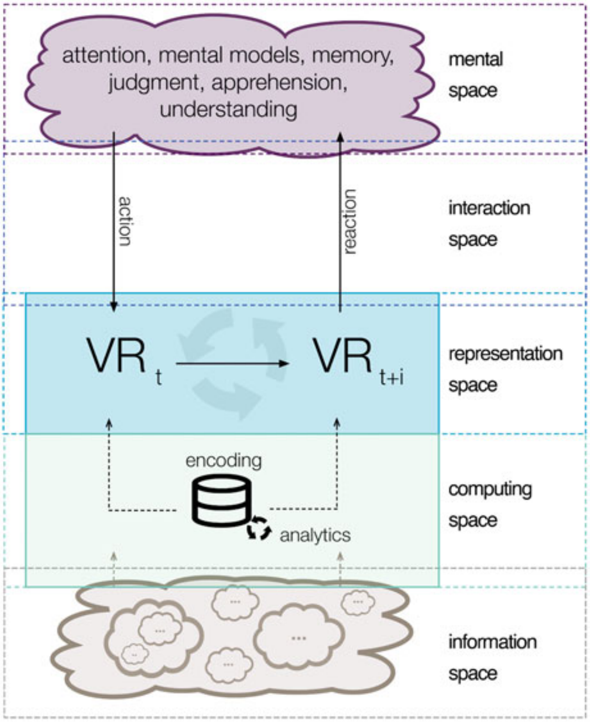

Fig. 9.3 The human–information interaction epistemic cycle (adapted from original draft with permission of author, Paul Parsons). We cannot overemphasize the importance of conceptualizing these spaces as metaphorical and not simultaneously reifying the layers as physical systems with linear exchanges of information. In practice, information processing emerges dynamically, simultaneously across the material components that constitute the system. This diagram can be construed as a snapshot of this dynamic processing, linearly unfolded in time from left to right

This contextualization is crucially important in ensuring basic theory is applied appropriately in design-driven research.

Model Building to Support Innovation An exemplar for progress that supports research in both basic and applied dimensions is the EDIFICE framework[10] developed by Sedig and Parsons. As a conceptual model, it provides a structure for thinking about the processing of information (such as goal-directed interaction with a visualization) distributed through the components of a cognitive system. In Fig. 9.3, we find five (metaphorical) spaces that together form a *human–information interaction epistemic cycle* (see [57–59]).

The *information space* consists of the set of information with which users might interact and the *computing space* its storage and manipulation (i.e., machine computation). In the *representation space*, encoded information is made available for perception. (The "space" of representation is an abstraction, but is reified in computers as "the interface.") The *interaction space* affords exchange of information via action and perception: where the interpreter performs actions and receives reactions. In the *mental space* exist the mind and mental operations that contribute to

[10] Epistemology and Design of human–InFormation Interaction in Cognitive activitiEs.

but importantly do not entirely constitute the construction of knowledge. The model is clearly grounded in the perspectives of information processing and distributed cognition. Though it was conceived in the context of interaction with complex visualization tools, its abstractions can be fruitfully applied to the wider space of multimodal and multimedia external representations. Most importantly, it makes explicit that the design of a visualization tool is a communicative act between designer and user.

The EDIFICE framework offers a productive nomenclature for designating which components of a distributed cognitive system we might be addressing in the context of a particular research project, allowing us to more accurately characterize limitations and desired integrations for future work. For example, a new visualization system that uses machine learning to recommend graph encodings would primarily involve the design of algorithms in the *computing space* and resultant productions in the *representation space.* A user-study of such a tool would involve measuring the outcome of operations in the *mental space* when an individual interacts with the application (via the *interaction space)*. Most importantly, the framework serves as tool for thinking about how the processing of information is distributed across a system of human-visualization interaction: a problem of substantial importance to designers and researchers alike. The authors have applied the framework to describe the relative distribution of information processing across machine and human actors [51], to characterize the construct of interactivity [60], and as the backbone for a pattern-language to aid conceptualization of novel visualization designs [57, 58].

9.5 The History and Future of Visualization Psychology

In *From Tool to Partner: The Evolution of Human–Computer Interaction* [26], Jonathan Grudin provides a comprehensive history of HCI. But this is not a commentary on the growth of a discipline, rather he illustrates how HCI (as a topic of study) emerged as a practice across communities in computer science, human factors, information systems, and information science. This is a telling editorial choice, revealing how entrenched institutional structures in academic disciplines interact with the moderately more pliable boundaries of professional societies to endow structural support to emerging subjects of inquiry that necessitate cross-disciplinary contribution. The cover illustration for the volume (penned by Susie Batford) can be read as deeply metaphorical. Over an undulating sea rise distinct mountain peaks, bearing the labels of various computing-related fields, including MIS (management information systems), HF (Human Factors), CS (Computer Science), and LIS (Library and Information Science). Running down and over and across the peaks, ostensibly nourishing rich research ecosystems, are bright blue rivers fed by an enormous raining cloud—labeled *psychology*.

One can imagine a similar scene for a history of Visualization. Research involving the creation, systematization and situated use of (primarily, though not entirely, graphic) visual-spatial representations of information is taking place across the sciences and humanities. Such research is enabled by *both* computing technology and theories of human behavior. By virtue of its name, the VIS community claims epistemic authority over visualization and serves as a pragmatic "home base" for technological innovation. But basic psychological theory rains upon disciplinary peaks like chemistry, physics and mathematics, education, communication studies, and even philosophy. Scholars in these disciplines are not merely *using* visualizations as tools in their work, but doing work that centers (representation) design, development, and evaluation, as well.

Through their Call for Papers, the organizers of the 2020 VisPsych workshop articulated a vision for a new subject, one that would catalyze an interdisciplinary community in pursuit of new research directions of benefit to both VIS and psychology. We begin this chapter by detailing the grounds on which we agree with this premise: that visualization (as a phenomenon) is a fertile laboratory for exploring human cognition, that engineering and design-driven research in visualization can be improved via appropriate grounding in psychological theory, and that well-structured collaborations across disciplinary boundaries can foster a virtuous cycle of mutual benefit. Where we diverge from this vision, is in characterizing the subject as *new*. Rather, we see the intersection of visualization and psychology as tracing back to the origins of human–computer interaction. Furthermore, relevant study of external representations permeates beyond the present institutional boundaries of VIS. We believe that the psychology of visualization is so fundamental to our progress that a call for a new interdisciplinary community should both catalyze a dedicated research program *and* re-center and expand the boundaries of visualization as a field.

Writing from the hallowed halls of Xerox PARC in the late 1990s, Stuart Card, Jock Mackinlay, and University of Maryland colleague Ben Shneiderman compiled what was to become the first *de facto* textbook for a burgeoning field—*Readings in Information Visualization: Using Vision to Think* (1999). Compiled a decade after the NSF-sponsored report that spawned the formal discipline [42] this now-venerated collection of papers and essays documented the state of VIS research at the close of its "foundational period," its table of contents betraying its continued entanglement with human–computer interaction, human factors, and computer graphics communities. As a technology, visualization opened new frontiers for presenting data in multiple dimensions with real-time interactions that the newly affordable PC platforms could render. Visualization was a tool for exploring the new information structures digital computers afforded, for supporting user interaction within the document-application paradigm of the time, and for conceptualizing and building the very graphical user interfaces we take for granted today. And there at the very beginning of visualization, there was *the psychology of external representation*. Card, Mackinlay and Shneiderman saw fit to begin their introductory chapter with a narrative of cognition outside the mind, describing how visual external representations like Arabic numerals, slide rules, and navigational charts could be used to support computation distributed through the environment.

But VIS was and would remain first and foremost a constituent of Computer Science. Like HCI and Human Factors, the early contributions of Psychology would be primarily psychophysics and empirical measures of "usability." While these areas are not to be dismissed, in the interceding decades, scientists have come to embrace perspectives that ground cognition in a body, situated in an environment, distributed through an ecosystem. There is a milieu in which these perspectives intersect and inform research as disparate as how expert mathematicians invent notations for new concepts, how animations of 3D models help or inhibit learning in chemistry, and how multiple modalities can be leveraged to engage diverse audiences in museums. These questions too are about humans interacting with representations of information; they are *like* but not *quite* VIS material. We believe that as a field, visualization should re-center itself in this space, taking a step back from Computer Science and toward social and behavioral sciences more broadly, "zooming out" from the interactive, abstract, computer-based caveats of (traditional) visualization to the first principles that apply across these phenomena. If we shift our focus from *visualization as a method of computing* to *external representation as a tool for thinking*, we find a framework for giving structure to the factors that exert causal influence on the phenomena we study; concepts that considered in isolation appear idiosyncratic may in fact be part of a more predictable, coherent whole.

Acknowledgments We offer thanks to Pamela Riviere for offering a diagram of neural data in Fig. 9.1, to Arvind Satyanarayan, Paul Parsons, David Kirsh, and Oisin Parkinsoncoombs for productive discussions on these topics, and to anonymous reviewers for productive feedback.

References

1. W. Aigner, S. Miksch, W. Muller, H. Schumann, and C. Tominski. Visualizing time-oriented data. a systematic view. *Computers & Graphics*, 31(3):401–409, June 2007.
2. S. Ainsworth. DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3):183–198, June 2006.
3. W. Bechtel. *Philosophy of Science: An Overview for Cognitive Science*. Psychology Press, 1988.
4. F. Beck, M. Burch, S. Diehl, and D. Weiskopf. A Taxonomy and Survey of Dynamic Graph Visualization. *Computer Graphics Forum*, 00(00):1–27, 2016.
5. J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press, Madison, WI, 1983.
6. A. Blackwell and Y. Engelhardt. A Meta-Taxonomy for Diagram Research. In *Diagrammatic Representation and Inference*, pages 47–64, 2002.
7. T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. Visualization of Eye Tracking Data: A Taxonomy and Survey. *Computer Graphics Forum*, 36(8):260–284, 2017.
8. M. Boden. *Mind As Machine: A History of Cognitive Science Two-Volume Set*. Oxford University Press, 2006.
9. M. Bostock, V. Ogievetsky, and J. Heer. D^3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), Dec. 2011.

10. M. Cabré Castellví. *Terminology: Theory, Methods, and Applications*, volume 1. John Benjamins Publishing, 1999.
11. S. Card, J. Mackinlay, and B. Shneiderman. *Readings in Information Visualization. Using Vision to Think.* Morgan Kaufmann Publishers, San Francisco, CA, 1999.
12. D. Chandler. *Semiotics: the basics*. The basics. Routledge, New York, NY, third edition, 2017.
13. P. C. Cheng. Probably good diagrams for learning: Representational epistemic recodification of probability theory. *Topics in Cognitive Science*, 3(3):475–498, 2011.
14. P. C. Cheng and H. A. Simon. Scientific discovery and creative reasoning with diagrams. In S. M. Smith, T. T. Ward, and R. A. Finke, editors, *The Creative Cognition Approach*, pages 205–228. The MIT Press, 1995.
15. E. Chi. A taxonomy of visualization techniques using the data state reference model. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, pages 69–75, 2000.
16. A. Clark. *Being There: Putting Brain, Body and World Together Again*. MIT Press, 1997.
17. A. Clark and D. Chalmers. The Extended Mind. *Analysis*, 58(1):7–19, 1998.
18. W. S. Cleveland and R. McGill. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.
19. W. S. Cleveland and R. McGill. An Experiment In Graphical Perception. *International Journal of Man-Machine Studies*, 25(5):491–500, Nov. 1986.
20. W. S. Cleveland and R. McGill. Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data. *Journal of the Royal Statistical Society Series A-Statistics in Society*, 150:192–229, 1987.
21. E. de Vries and C. Masclet. A framework for the study of external representations in collaborative design settings. *International Journal of Human-Computer Studies*, 71(1):46–58, Jan. 2013.
22. M. A. Elliott, C. Nothelfer, C. Xiong, and D. A. Szafir. A Design Space of Vision Science Methods for Visualization Research. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1117–1127, Feb. 2021.
23. Y. Engelhardt. *The Language of Graphics*. PhD thesis, University of Amsterdam, 2002.
24. S. Few. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press, 2009.
25. B. Fisher, T. M. Green, and R. Arias-Hernández. Visual Analytics as a Translational Cognitive Science. *Topics in Cognitive Science*, 3(3):609–625, 2011.
26. J. Grudin. *From Tool to Partner: The Evolution of Human-Computer Interaction*, volume 10 of *Synthesis Lectures on Human-Centered Informatics*. Morgan & Claypool Publishers, 2017.
27. R. L. Harris. *Information Graphics: A Comprehensive Illustrated Reference*. Oxford University Press, 1999.
28. M. Hegarty. The Cognitive Science of Visual-Spatial Displays: Implications for Design. *Topics in Cognitive Science*, 3(3):446–474, 2011.
29. J. Hollan, E. Hutchins, and D. Kirsh. Distributed cognition: Toward a new foundation for human-computer interaction research. *Transactions on Computer-Human Interaction (TOCHI)*, 7(2):174–196, June 2000.
30. J. Hoopes. *Peirce on Signs: Writings on Semiotic by Charles Sanders Peirce*. The University of North Carolina Press, Chapel Hill, 1991.
31. E. Hutchins. Cognition in the Wild. *MIT Press*, pages 1–5, 1995. ISBN: 0262581469.
32. E. Hutchins. Cognition, Distributed. In *International Encyclopedia of the Social & Behavioral Sciences*, pages 2068–2072. Elsevier, 2001.
33. W. Kintsch. The role of knowledge in discourse comprehension: a construction integration model. *Psychological Review*, pages 163–182, 1988.
34. D. Kirsh. When is Information Explicitly Represented? In *The Vancouver Studies in Cognitive Science*, pages 340–365. Oxford University Press, 1990.
35. D. Kirsh. Implicit and Explicit Representation. *Encyclopedia of Cognitive Science*, pages 478–481, 2006.

36. D. Kirsh. Thinking with external representations. *AI and Society*, 25(4):441–454, 2010.
37. J. Larkin and H. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 99:65–99, 1987.
38. P. H. Lindsay and D. Norman. *Human information processing: an introduction to psychology*. Academic Press, New York, 1972.
39. Z. Liu, N. Nersessian, and J. Stasko. Distributed cognition as a theoretical framework for information visualization. *IEEE transactions on visualization and computer graphics*, 14(6):1173–80, 2007.
40. M. Massironi. *The Psychology of Graphic Images*. Psychology Press, 2001.
41. R. Mayer. Cognitive Theory of Multimedia Learning. In *The Cambridge Handbook of Multimedia Learning*, pages 31–48. Cambridge University Press, 2005.
42. B. H. McCormick, T. A. DeFanti, and M. D. Brown. Visualization in Scientific Computing. *Computer Graphics*, 6(21):81, Nov. 1987.
43. T. Munzner. *Visualization Design and Analysis*. A.K. Peters visualization series. CRC Press, Taylor & Francis Group, 2014.
44. A. Musgrave and C. Pigden. Imre Lakatos. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2016.
45. D. Nauta. *The Meaning of Information*. Mouton, The Haugue, 1972.
46. U. Neisser. *Cognitive Psychology*. The Century psychology series. Appleton-Century-Crofts, New York, 1967.
47. D. A. Norman. *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1993.
48. S. E. Palmer. Fundamental Aspects of Cognitive Representation. In E. Rosch and B. B. Lloyd, editors, *Cognition and Categorization*, pages 259–302. Erlbaum, Hillsdale, NJ, 1978.
49. S. E. Palmer and R. Kimchi. The Information Processing Approach to Cognition. In *Approaches to Cognition: Contrasts and Controversies*, page 38. Lawrence Erlbaum Associates, 1986.
50. P. Parsons and K. Sedig. Common Visualizations: Their Cognitive Utility. In *Handbook of Human Centric Visualization*, pages 671–691. Springer, 2014.
51. P. Parsons and K. Sedig. Distribution of Information Processing While Performing Complex Cognitive Activities with Visualization Tools. In *Handbook of Human Centric Visualization*, pages 671–691. Springer, 2014.
52. S. Pinker. Theory of Graph Comprehension. In R. Freedle, editor, *Artificial Intelligence and the Future of Testing*, pages 73–126. Erlbaum, Hillsdale, NJ, 1990.
53. Y. Rogers. When the external entered HCI: Designing effective representations. In T. Erickson and D. W. McDonald, editors, *HCI Remixed: Reflections on Works That Have Influenced the HCI Community*, pages 275–280. MIT Press, Cambridge, MA, Jan. 2008.
54. A. Satyanarayan, D. Moritz, K. Wongsuphasawat, and J. Heer. Vega-Lite: A Grammar of Interactive Graphics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):341–350, 2017.
55. M. Scaife and Y. Rogers. External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, pages 185–213, 1996.
56. W. Schnotz. An Integrated Model of Text and Picture Comprehension. In *The Cambridge Handbook of Multimedia Learning*, pages 49–69. Cambridge University Press, 2005.
57. K. Sedig and P. Parsons. Interaction Design for Complex Cognitive Activities with Visual Representations: A Pattern-Based Approach. *AIS Transactions on Human-Computer Interaction*, 5(2):84–133, 2013.
58. K. Sedig and P. Parsons. *Design of Visualizations for Human-Information Interaction: A Pattern-Based Framework*. Synthesis Lectures on Visualization. Morgan & Claypool Publishers, 2016.
59. K. Sedig, P. Parsons, and A. Babanski. Towards a Characterization of Interactivity in Visual Analytics. *Journal of Multimedia Processing and Technologies, Special Issue on Theory and Application of Visual Analytics*, 3(1):12–28, 2012.

60. K. Sedig, P. Parsons, M. Dittmer, and R. Haworth. Human-Centered Interactivity of Visualization Tools: Micro and Macro-level Considerations. In *Handbook of Human Centric Visualization*, pages 671–691. Springer, 2014.
61. P. Shah. Graph Comprehension: The Role of Format, Content and Individual Differences. In M. Anderson, B. Meyer, and P. Olivier, editors, *Diagrammatic Representation and Reasoning*, pages 173–185. Springer London, London, 2002.
62. P. Shah and E. G. Freedman. Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, 3(3):560–578, 2011.
63. P. Shah, E. G. Freedman, and I. Vekiri. The Comprehension of Quantitative Information in Graphical Displays. In P. Shah, editor, *The Cambridge Handbook of Visuospatial Thinking*, pages 426–476. Cambridge University Press, New York, NY, 2005.
64. P. Shah and J. Hoeffner. Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, 14(1):47–69, Mar. 2002.
65. B. Shneiderman. The Eyes Have It: A Task By Data Type Taxonomy for Information Visualization. *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343, 1996.
66. D. Simkin and R. Hastie. An Information-Processing Analysis of Graph Perception. *Journal of the American Statistical Association*, 82(398):454–465, 1987.
67. D. E. Stokes. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Brookings Institution Press, 1997.
68. D. A. Szafir, S. Haroz, M. Gleicher, and S. Franconeri. Four types of ensemble coding in data visualizations. *Journal of Vision*, 16(5):11–11, Mar. 2016.
69. M. Tory and T. Moller. Rethinking Visualization: A High-Level Taxonomy. In *IEEE Symposium on Information Visualization*, pages 151–158, Oct. 2004.
70. B. Tversky. Visualizing Thought. *Topics in Cognitive Science*, 3(3):499–535, 2011.
71. VisPsych Committee. Visualization Psychology — 2020 IEEE VIS Workshop, 2020.
72. C. Ware. *Information Visualization: Perception for Design*. Elsevier Inc., San Francisco, CA, 2004.
73. D. Whitney and A. Y. Leib. Ensemble Perception. *Annual Review of Psychology*, page 25, 2017.
74. L. Wilkinson. *The Grammar of Graphics*. Statistics and Computing. Springer-Verlag, New York, 1999.
75. M. Wilson. Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625–36, Dec. 2002.
76. J. Worrall and G. Currie, editors. *The Methodology of Scientific Research Programmes: Philosophical Papers*, volume 1. Cambridge University Press, Cambridge, 1978.
77. J. Zhang and D. Norman. Representations in Distributed Cognitive Tasks. *Cognitive science*, 18(1):87–122, Jan. 1994.

Chapter 10
Visualization Psychology for Eye Tracking Evaluation

Maurice Koch, Kuno Kurzhals, Michael Burch, and Daniel Weiskopf

Abstract Technical progress in hardware and software enables us to record gaze data in everyday situations and over long time spans. Among a multitude of research opportunities, this technology enables visualization researchers to catch a glimpse behind performance measures and into the perceptual and cognitive processes of people using visualization techniques. The majority of eye tracking studies performed for visualization research are limited to the analysis of gaze distributions and aggregated statistics, thus only covering a small portion of insights that can be derived from gaze data. We argue that incorporating theories and methodology from psychology and cognitive science will benefit the design and evaluation of eye tracking experiments for visualization. This book chapter provides an overview of how eye tracking can be used in a variety of study designs. Furthermore, we discuss the potential merits of cognitive models for the evaluation of visualizations. We exemplify these concepts on two scenarios, each focusing on a different eye tracking study. Lastly, we identify several call for actions.

10.1 Introduction

Eye tracking experiments in visualization research provide insights into how people interpret and interact with visualizations. In contrast to classic performance analysis, the analysis of gaze behavior provides information about the distribution of visual attention over time. Eye tracking further helps understand visual strategies employed in interpreting a visualization or in working with a complex visual analytics system. In addition, machine learning, statistics, visualization research,

M. Koch · K. Kurzhals (✉) · D. Weiskopf
University of Stuttgart, Stuttgart, Germany
e-mail: Maurice.Koch@visus.uni-stuttgart.de; Kuno.Kurzhals@visus.uni-stuttgart.de; Daniel.Weiskopf@visus.uni-stuttgart.de

M. Burch
University of Applied Sciences Graubünden, Chur, Switzerland
e-mail: Michael.Burch@fhgr.ch

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_10

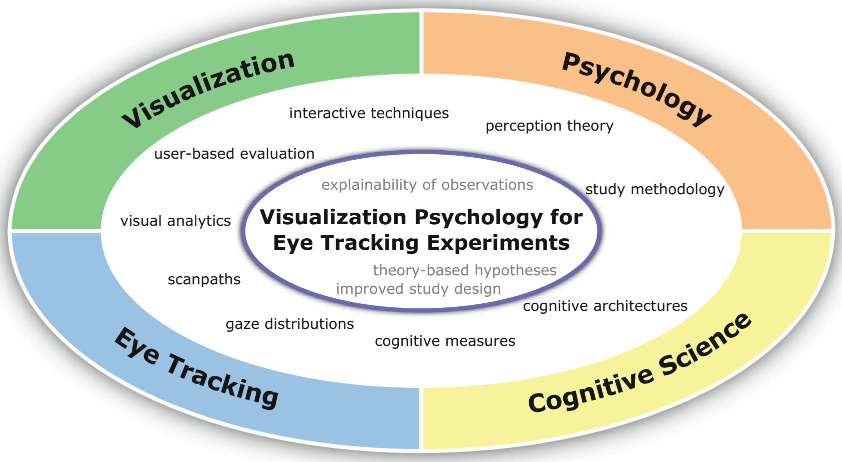

Fig. 10.1 Visualization psychology for eye tracking experiments incorporates expertise from psychology and cognitive science to improve the evaluation of visualization techniques by study methodology, theory integration, and cognitive architectures. Figure reprinted from Kurzhals et al. [26]

and data science in general contributed a multitude of new techniques [5, 11] to expand the spatio-temporal analysis of eye tracking data, verify results, and formulate new hypotheses. By combining such state-of-the-art analysis techniques with expertise from psychology, cognitive science, and eye tracking research, as depicted in Fig. 10.1, the design and insights gained from eye tracking experiments in visualization can be significantly improved. However, evaluation in visualization still lacks concrete guidance on such interdisciplinary research. One part of the problem is the increasing disconnect between psychology and visualization research. For example, in visual analytics, there is less focus on individual visualizations but on the processes that the tool is meant to support. Such processes often can be related to different scenarios, such as visual data analysis and reasoning and collaborative data analysis [30], to name a few. Although visualization research has become more process-centered on a conceptual level, evaluation today still mostly involves usability testing and benchmarking based on completion time and error metrics. For this reason, we advocate that the visualization community broadens their scope toward evaluation methodologies that better capture the dynamics of complex tool interactions. In a similar sense, we advocate that cognitive psychologists actively participate in that endeavor by focusing their study on higher level cognition. Fisher et al. [14] even call for translational research that bridges pure science and design, with the hope to better support knowledge transfer between both fields. A major inspiration for this work has been Kurzhals et al. [26], who advocated for more interdisciplinary research between the fields of psychology, cognitive science, and visualization. In this book chapter, we exemplify how the eye tracking modality could be beneficial to a broader scope of empirical studies, beyond classical laboratory experiments.

10.2 Study Designs

In the following, we describe how different study designs commonly found in visualization evaluation [10] can benefit from eye tracking methodology. Eye tracking has become popular in the evaluation of visualizations, and there is a wide variety of methods and metrics to evaluate the performance of visualization [15]. Kurzhals et al. [27] reviewed 368 publications that include eye tracking in a user study and identified three main approaches to evaluate visualizations: evaluating the distribution of visual attention, evaluating sequential characteristics of eye movements, and comparing the viewing behavior of different participant groups. Their review also shows that user studies with eye tracking have become more common in recent years.

However, the use of eye tracking in evaluation methods has been narrow in the sense that it is predominantly used in laboratory experiments but infrequently found in in-the-wild studies. Laboratory experiments offer great control and precise results but are primarily suited to study individual factors with predefined hypotheses. In this section, we outline the current practice of using eye tracking in visualization research, mostly in the context of controlled experiments. Furthermore, we outline how eye tracking could be beneficial beyond laboratory experiments. For this, we include a discussion of in-the-wild studies.

10.2.1 Controlled Experiments

Eye tracking has become increasingly popular in laboratory experiments. In visualization research, controlled experiments have been mostly conducted for summative evaluation, such as usability testing and benchmarking. However, such studies often fail to relate their findings to the underlying cognitive processes.

Here, we showcase just a few selected eye tracking studies in visualization with a strong focus on cognitive aspects, such as reasoning, memorability, and perception.

Huang et al. [20] studied how link crossings in graph drawings affect task performance. Participants were asked to find the shortest path between two specified nodes for each drawing. Their eye tracking experiment revealed that link crossings, contrary to the common belief, only have minor impact on graph reading performance, especially at angles of nearly 90°. Instead, the extra time spent on certain drawings was due to the tendency of subjects to prefer certain paths at the beginning of the search task. It was observed that subjects tend to follow links that are close to a (imaginary) straight line between the target nodes. This can increase the search time if no such links exist in the graph drawing, and alternative graph lines must be considered. This behavioral bias during the initial search process in graph drawings was termed geodesic-path tendency. Körner et al. [28, 29] found that this behavior can be explained by studying to which extent search and reasoning processes in graph comprehension are performed concurrently. The two main process involved in such a task are first detecting both specified nodes in the graph (search) and next finding the shortest path between those two nodes (reasoning). Assuming that these processes occur in parallel, subjects would not show this kind of bias toward certain links in graph drawings as described by geodesic-path tendency. Körner et al. conducted eye tracking experiments and found that these two graph comprehension processes indeed are mostly performed sequentially. This means that subjects can only rely on local information of the graph drawing to perform reasoning during the search task.

Borkin et al. [6] studied the memorability of visualizations and how well they are recognized and recalled. Their experiments consist of three phases: encoding, recognition, and recall. In the encoding phase, subjects were exposed to 100 different visualizations sampled from the MassVis dataset. After the encoding phase of 10 seconds per image, subjects were exposed to the same images plus unseen filler images as part of the recognition phase. In both phases, eye fixations were collected to examine the elements in visualizations that facilitate the memorability. In the last phase, subjects were asked to describe correctly identified images as best as possible to understand what elements were easily recalled from memory. In the encoding and recognition phases, eye fixations were analyzed with heatmaps to find what parts of the visualization draw initial attention to subjects. During encoding, subjects tend to perform visual exploration, and fixations are distributed across the image. This pattern can be observed on most images. Fixations during the recognition phase are distinct between most recognizable images and least recognizable images. It was shown that in the most recognizable visualizations, fixations are more biased toward the center of the image and are generally less widely distributed. This means that relatively few fixations are needed to recall easily recognizable images from memory, whereas less recognizable images require more contextual information. Their study also shows that participant descriptions are of higher quality for visualizations that are easily recognizable even with a reduced amount of encoding time (such as one second). Interestingly, prolonged exposure does not change the fact that some visualizations stay more recognizable.

Hegarty et al. [18] studied how saliency of task-relevant and task-irrelevant information on weather maps impacts task performance. Mean proportion of fixation time was measured to study the level of attention on task-relevant or task-irrelevant information before instructions and after instructions. On the one hand, it was reported that fixation time significantly increases on task-relevant areas after instructions were given, which shows that attention is strongly driven by top-down influences. On the other hand, visual salient regions do not draw attention to participants, unless they correspond to task-relevant areas. These results emphasize that visual salience does not necessarily facilitate task performance, unless participants are sufficiently guided by top-down processes toward task-relevant information.

The aforementioned visualization studies exemplify that eye tracking has become an established modality to study cognitive processes. Furthermore, many of these results are directly applicable to the visualization community

10.2.2 In-the-Wild Studies

As the complexity of visual artefacts increases, it becomes harder to provide holistic assessments of the effectiveness of complex visualization tools. Field studies offer more realism by assessing systems within their natural environment like at the domain expert's work place. In such settings, it is easier to study processes, like sense-making since they tend to be highly context-sensitive. Thus, such processes are more difficult to capture in controlled experiments that usually impose tight protocols [30]. Many researchers believe that visualization evaluations could benefit from more qualitative research, for example, by employing ethnographic techniques [13, 38]. In general, social science methods should receive more attention in the community since individual assessment techniques often fail to capture contextual factors [31].

Ethnographic techniques have been advocated by Shneiderman et al. [38] in the form of multi-dimensional in-depth long-term case studies (MILCs). MILCs are performed in-field, in a domain expert's natural working environment, and thus they are unobtrusive and guarantee more realistic results. Data collected in MILCs is mostly qualitative and consists of interviews, log books, user maintained diaries, and usage statistics obtained from the visualization tool. Field studies are often based on ethnographical participant observation methods, interviews, surveys, and automated logging of user activity [38], i.e., they are predominantly qualitative research in terms of data collection and analysis. Qualitative evaluation often involves thematic analysis and manual coding, and both are inherently subjective processes [10]. There are multiple problems associated with a primarily quantitative data collection and analysis approach. First, data collection and analysis are tedious processes that often involve a lot of manual work. In terms of data analysis, software tools like computer-assisted qualitative data analysis software (CAQDAS) [3] improve the efficiency of thematic analyses and assist coding, but only to a limited extent. This problem gets exacerbated in long-term studies where a large amount of diverse data is collected. For this reason, many MILCs come only with a few interviews and observations, and during the study, data collection is sparse, at most it consists of user interface logs that are automatically recorded (in practice, even logging is very uncommon except for Shneiderman's MILC study [38]).

The usage of physiological sensors is in particular challenging in ethnographic studies, where the property of unobtrusiveness must be obeyed (interference by study coordinators needs to be kept minimal). This is hardly achievable with stand-alone eye tracking devices and electroencephalogram (EEG), which are highly

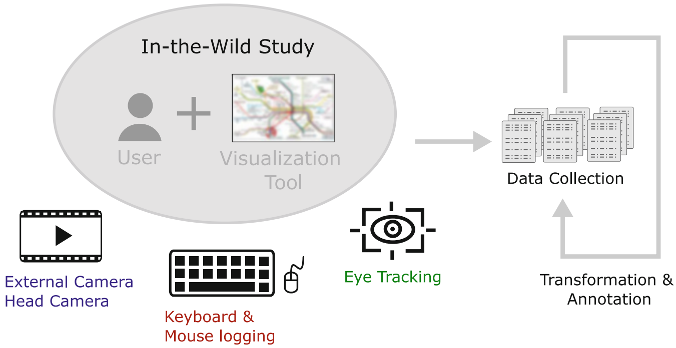

Fig. 10.2 Conceptual overview of data collection in in-the-wild studies. Different modalities such as camera views, keyboard and mouse logging, and eye tracking can be combined to generate a data-rich description of the study. The collected data can be transformed or extended semantically with labels provided by human annotators. The small visualization inset shows a figure reprinted by permission of Taylor & Francis Ltd from Netzel et al. [33]

invasive and lack mobility. Furthermore, such physiological sensors often require external supervision and careful setup. This naturally restricts what and how data is collected in ethnographic studies. However, in regard to eye tracking devices, we have seen technological progress toward mobile devices that are less invasive and require almost no external supervision. In this way, eye tracking could act as a quantitative modality that does not interfere with ethnographic requirements like unobtrusiveness. Figure 10.2 illustrates the basic idea of collecting data from multiple sources and semantically and/or algorithmically extending it in subsequent steps.

Whether a modality is considered invasive depends not only on the modality itself but also on the situational context. For example, think-aloud protocols can be elicited either naturally, or they can be imposed externally on request (by a study coordinator), which could negatively affect reasoning processes [2]. Think-aloud might also negatively interfere with the natural eye movement, for example, during attending the screen. To compensate this issue retrospective think-aloud [27] of screen recordings accompanied by eye tracking data was suggested [12]. In general, it is important to detect these attention shifts, which also occur naturally without external stimulation and revalidate the recorded eye movements. Transferring our studies to virtual reality (VR) could provide non-invasive access to physiological sensors that are readily available in VR headsets. This could go beyond eye tracking and further include tracking head/body movements and interface interactions.

The previously discussed scope of in-the-wild studies is on individuals but can be easily extended to collaborative settings as well. In that regard, pair analytics [2] provides an interesting approach to studying social and cognitive processes for the evaluation of visual analytics tools. Pair analytics studies the interaction between two human subjects, the subject matter expert and the visual analytics expert, and visual analytics tools. The visual analytics expert and subject matter expert collaborate to solve a specific domain goal, but both have different responsibilities and roles in that process. The subject matter expert (driver role) is the domain expert that has the contextual knowledge but not the expertise to use the visual analytics tools, whereas the visual analytics expert (navigator role) lacks the domain knowledge but the technical expertise to translate the verbal requests from the subject matter expert to tool commands. The dialog between the subject matter expert and the visual analytics expert makes the mental models and cognitive processes explicit and thus captures important cues of the collaborative process. Compared to classical think-aloud protocols, verbalization during collaborative processes occurs naturally. Aligning the rich data from think-aloud protocols with eye-movements from the subject matter expert and the visual analytics expert could be a good starting point for in-depth analysis on social and cognitive processes. Kumar et al. [25] have proposed a similar type of study, but in the context of pair programming. Data from eye tracking data and other modalities, like recorded video, are time-synchronized. Having discussed the merits of in-the-wild studies in the evaluation of visualizations, we also need to address the inherent difficulties of conducting those studies. As Shneiderman et al. [38] already mentioned, it is necessary for researchers and participants to allocate a considerable amount of time into such studies. For example, Valiati et al. [40] performed multiple longitudinal case studies, each took about three to four months. This complicates recruiting participants, in particular, when domain experts are needed. It needs to be emphasized that this requires an intense level of collaboration and devotion from both the researchers and the domain experts.

10.2.3 Bridging Between Quantitative and Qualitative Research

The aforementioned study designs can be roughly classified as being either qualitative or quantitative. Quantitative evaluation, often in laboratory experiments, follows statistical frameworks to make precise inferences about predefined hypotheses. Qualitative evaluation potentially provides a richer understanding of the studied phenomena. This includes field studies, observational studies, and interviews [10].

Study designs that encompass data collection, analysis, and inferences techniques from both methodological paradigms can potentially offset their individual shortcomings. The commonly found dichotomy in quantitative and qualitative inquiry is too narrow. This motivates the research field of mixed methods, which uses methods from both disciplines to provide a better understanding of the studied phenomena [23]. One of the hallmarks of mixed methods is to achieve integration by bringing qualitative and quantitative data together in one study [32]. This integration can occur at different levels such as integration at the study design level, methods, and interpretation/reporting. An example of integration at study level is an explanatory sequential design where the quantitative phase informs the follow-up qualitative phase. For example, a controlled study design with eye tracking could be conducted to quantitatively evaluate the performance on a visual search tasks with two different visual representations. A follow-up qualitative phase could be justified for several reasons. For example, a group of participants could strongly deviate in performance. The follow-up qualitative phase could try to identify the root of this cause by performing a retrospect think-aloud protocol where the respective participants comment on their played-back eye movements. Think-aloud can also be performed concurrently to eye tracking experiments, which would correspond to a convergent mixed methods design.

Integration at the other two levels is more concerned with mixed-data analysis, and it is considerably more challenging and less explored [32, 41]. Common strategies of mixed-data analysis include data transformation, typology development, extreme case analysis, and data consolidation [8]. Data consolidation is one of the greatest challenges of mixed-data analysis since it merges two datasets, which goes beyond linking. The difference is that both data sources remain clearly identifiable after data linking while consolidation leads to a genuine new piece of information. These techniques are not necessarily distinct, for example, data transformation could be an important prepossessing step for data consolation. Data transformation encompass two data conversion directions, either quantified data is transformed to qualitative data (qualitizing) or vice versa (quantizing) [41]. A common way to perform quantization is by counting codes in an thematic analysis. In that way, quantitative methods like inferential statistics can be applied indirectly to qualitative data. Qualtizing can be seen as a semantic transformation of the original quantitative data. This could add a semantic link to quantitative measurements, which is usually not present in such measurements beforehand. For example, gaze data in its raw form is just a trajectory in 2D space without any semantic link to the underlying stimulus. For static stimuli, this semantic link is easy to provide since there is a one-to-one correspondence between gaze location and stimuli location. However, such a direct correspondence is not present in dynamic stimuli where the underlying scene varies over time. Providing additional semantics to gaze data with underlying dynamic stimuli, for example, by labeling time spans according to the participant's activity, would increase the usefulness of these measurements. This form of data consolidation by annotation of quantitative data can improve the credibility of those measurements and thereby improve the quality of subsequent mixed-data analysis steps.

10.3 Explainability of Observations

As already outlined in the previous section, building semantic links between gaze data and contextual factors, like scene information or activity labels, can aid the data analysis and thereby the explainability of observations.

Areas of Interest

Scanpaths can be transformed to qualitative data by mapping each fixation to a label, which uniquely identifies an area of interests (AOIs). The usefulness of such a representation depends on the semantics of AOIs. For example, AOI grids automatically generated for static stimuli do not provide much semantic details since an AOI hit is just still an indicator of spatial position (spatial quantization) but does not provide semantic information w.r.t the underlying visual entity. A similar problem occurs for AOI induced by automatic clustering of gaze data, where regions with strong accumulation of gaze positions are defined as AOIs. In contrast to such automatically generated AOIs, manually AOIs defined based on semantics (images on web pages, axes on graphs, etc.) can provide more detailed information.

Interpretation and Data Analysis
In Sect. 10.2, we have mentioned the challenges in data collection and analysis in the context mixed-methods research. These kinds of challenges are particularly relevant for in-the-wild studies, such as the previously described long-term field studies in pair analytics. It is challenging to integrate data from heterogeneous data sources, such as eye tracking and other physiological sensors, as well as handwritten or verbal protocols. An interesting approach toward these problems is visual data analysis, sometimes referred to as *visualization for visualization (Vis4Vis)* [42]. The vision behind Vis4Vis is to use visualizations to analyze and communicate data from empirical studies. In the context of eye tracking studies, visual analysis tools have been shown to support the evaluation of studies. For example, Blascheck et al. [4] provide a comprehensive overview of visualization techniques for eye tracking data. Some visual analysis approaches have been proposed that integrate eye tracking data with other data modalities, such as think-aloud protocols and interaction logs. Blascheck et al. [3] proposed a visual analytics system that allows interactive coding and visual analysis of user activities. Such approaches could be considered as a first step toward visual analysis of data-rich empirical studies with multiple data modalities. Nonetheless, there is still the need for more scalable visual representations and automatic analysis techniques to better support the analysis of data from long-term empirical studies.

10.4 Cognitive Architectures

One of the overarching goals of empirical studies in visualization is to formulate guidelines and heuristics that inform the design of future visualizations. However, many psychological phenomena only apply to specific aspects of the evaluation, like Gestalt Laws, but visualization consists of multiple perceptual and cognitive aspects combined. Thus, guidelines and heuristics on system level would be preferable. However, since they typically involve higher level cognitive tasks, they are more influenced by individual factors, such as knowledge, cultural background, and cognitive capabilities. Computational models have the potential to generalize across a wide range of individuals [27] and can provide methods to accurately predict the effectiveness of visual designs [19]. As shown in Fig. 10.3, such simulation could be performed on multiple levels. On the most fundamental level one, simulation

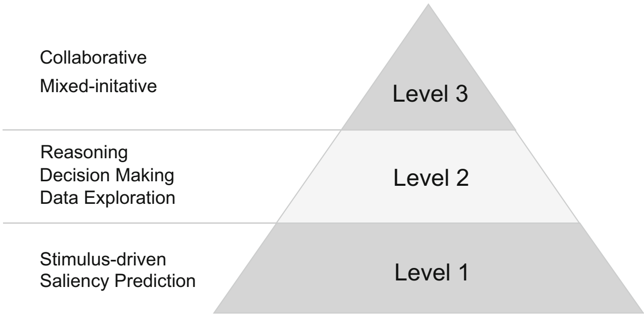

Fig. 10.3 Cognitive simulation can be performed on multiple levels. Each layer corresponds to one class of tasks. Each level depends on its lower levels. For example, simulation of collaborative settings with multiple individuals performing a common task requires successful simulation of cognitive tasks (levels 1 and 2) for individuals

of human cognition boils down to perceptual simulation that is often highly driven by the stimulus or more general bottom-up influences. Early work on that level has been proposed by Itti and Koch in the context of visual saliency prediction [22]. In general, cognitive simulation on higher levels has been less explored, mostly due to its complexity and the lack of formal descriptions. Nonetheless, computational models based on cognitive architectures have been proposed to automate the evaluation of visualizations on the level of reasoning and decision-making. One example of the application of cognitive architectures like ACT-R [1] is *CogTool* (see https://www.cogtool.org), which is deployed for the initial validation of web designs. Eye fixations can play an important role as a means to train and validate cognitive models. For example, Raschke et al. [36] propose a cognitive model based on ACT-R that simulates visual search strategies. Their motivation is to build a simulation tool similar to *CogTool* that allows automatic, thus non-empirical, evaluation of visualizations. In contrast to *CogTool*, which is based on an extended version of Keystroke-Level Model [9], their model is trained on eye fixations. Although their work does not provide any concrete implementation, other researchers have demonstrated that models based on ACT-R can simulate eye movements on simple line charts with high confidence [35]. Their model even provides vocal output and, thus, is able to simulate graph comprehension with results close to human level. From a technical viewpoint, cognitive architectures like ACT-R have some limitations that prevent their adoption to more complex tasks. For example, Heine et al. [19] advocate the use of probabilistic models, like Dynamic Bayesian networks, in the context of modeling human cognition. Probabilistic models could provide a unified mathematical model toward human cognition and allow to describe variation of factors that are not explicitly modeled. This is a strong advantage over ACT-R that depends on explicit rule-based modeling, which does not scale well for sophisticated visualizations.

10.5 Example Scenarios

Visualization evaluation could benefit from the aforementioned study designs, the explainability of observations, and cognitive architectures. We exemplify this, based on two previous eye tracking studies: one on the design of metro maps [33] and one on the evaluation of parallel coordinates plots [34]. We discuss how these studies could be enhanced and extended by adopting ideas from the previous sections of this chapter.

10.5.1 Overview of Scenarios

Scenario 1: Metro Maps

Investigating the readability of metro maps is a challenging field of research, but the gained insights are valuable information on how to find design flaws, enhance the design, and make the maps more understandable to travelers [7]. Netzel et al. [33] compare color-coded and grayscale public transport maps with an eye tracking study. The major outcome is that color is an important ingredient to reduce the cognitive burden to follow lines. Eye tracking was essential in this study to understand the strategies participants applied to solve a route finding task between a start and a target station (Fig. 10.4). The analysis showed that color maps led to much longer saccades, and it was hypothesize that colored lines made participants feel safe and, hence, the route finding tasks could be answered faster and more reliably. In contrast, in grayscale maps, the participants' eyes moved with significantly smaller saccades to trace a line reliably, which was due to missing color that would otherwise have helped to visually and perceptually separate the metro lines from each other. A practical result of this eye tracking experiment for the professional map designer is that color is crucial for route finding tasks, and

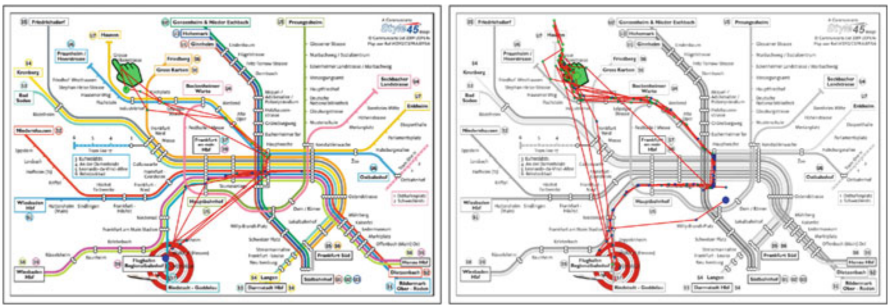

Fig. 10.4 Scenario 1: Metro maps in color (left) and in gray scale (right) have been compared for solving a way finding task from a start (hand) to a target location. Eye tracking was measured to identify differences in the reading behavior of both conditions. Figure reprinted by permission of Taylor & Francis Ltd from Netzel et al. [33]

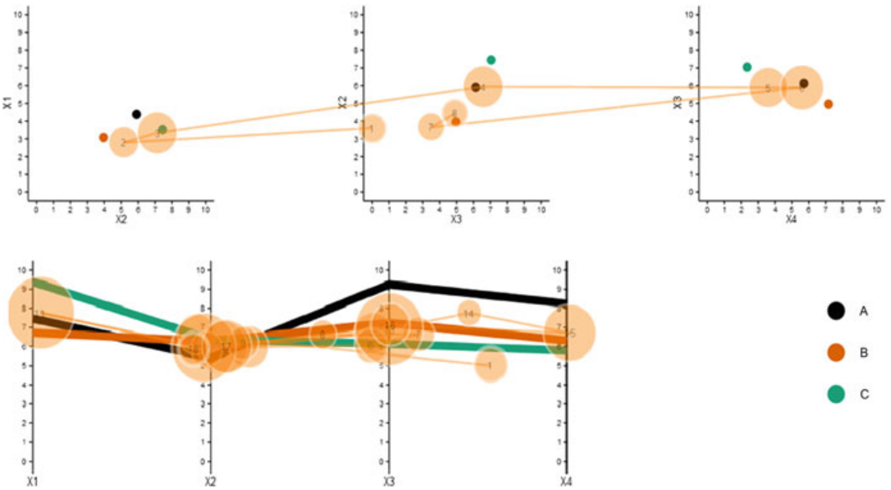

Fig. 10.5 Scenario 2: Eye tracking was applied to compare how people compare distances between three points (A, B, and C) in scatterplots (top) and parallel coordinate plots (bottom). Figure reprinted by permission of Elsevier from Netzel et al. [34]. Copyright 2017 Zhejiang University and Zhejiang University Press. Licensed under the CC BY-NC-ND 4.0 license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

hence the much cheaper printed variants in gray scale would obviously be counter-productive for the business, although the costs are much lower.

Scenario 2: Scatter and Parallel Coordinates Plots
The second example of a study investigates the assessment of relative distances between multidimensional data points with scatterplots and parallel coordinates plots [34] (Fig. 10.5). The authors performed an eye tracking study and showed that scatterplots are efficient for the interpretation of distances in two dimensions, but participants performed significantly better with parallel coordinates when the number of dimensions was increased up to eight. With the inclusion of eye tracking, it was possible to identify differences in the viewing of the two visualization types considering fixation durations and saccade lengths. The authors further introduced a visual scanning model to describe different strategies for solving the task. With the help of eye tracking, a bias toward the center (parallel coordinates plot) and the left side (scatterplots) of the visualizations could also be measured, which is important for the design of such plots considering where participants will potentially spend most of their attention. However, understanding clear visual attention patterns like following a line as described in the former eye tracking study is not possible here since either the diagram consists of crowds of points (scatterplot) or a lot of crossing and partially occluding polylines (parallel coordinates plot). Hence, the reading behavior is more complex and harder to model than in Scenario 1.

10.5.2 Potential Extensions

In-the-Wild Studies

As described in Sect. 10.2.2, studies in the wild provide a higher realism for experimental outcomes. For Scenario 1, this is highly desirable because the interpretation of metro maps is a task performed by many people in everyday situations. For the sake of controllability, stimuli and task were adjusted to fit to a laboratory setting: People were watching metro maps on a screen with start and goal clearly highlighted. The situation in a real metro station would differ significantly. Numerous confounding factors such as distractions by other people, no clear identification of start and goal, and other potential stress-inducing factors might influence the results of how people look at such a map.

Scenario 2, in contrast, involves visualization techniques (i.e., parallel coordinate plots) that are less known to people. An application in the wild would presumably take place with domain experts and data scientists rather than a more general audience of students, as it was the case in the conducted study. Furthermore, the set of performed tasks would be extended in comparison to the lab study. However, for the hypotheses of the original experiment, the expertise of the participants was not the determining factor since the study aimed to analyze general behavior. For measurements over longer time periods, the experts could potentially show additional behavior patterns and learning effects, while general behavior aspects should not change.

Collaborative Studies and Pair Analytics

The investigation of metro maps in Scenario 1 is often an individual task but is in real life also performed collaboratively. Similar to the application of the task in the wild, the analysis of collaborative task solving has the potential to reveal details on how decision-making is performed. Scenario 2 can be imagined for typical analysis tasks involving domain and visualization experts. In both scenarios, the dialog between participating people provides valuable information on a qualitative level. Scenario 1 provides the possibility to perform a symmetrical setup where both persons have the same prerequisites and solve the task together. In Scenario 2, the integration of the visualizations in a visual analytics framework has the potential to focus more on a pair analytics approach where people with different fields of expertise (i.e., domain and visualization expert) work together to solve the task.

Furthermore, measuring the gaze behavior of both persons indicates periods when they potentially share visual attention and when they might be confused, e.g., searching for the region the other person is talking about. Hence, eye tracking helps evaluating not only the visualization at hand but also the interaction between persons.

Mixed Methods

Qualitative and quantitative evaluation combined provide a more comprehensive understanding of the research topic than each method on its own. Scenarios 1 and 2 mainly focused on the quantitative evaluation of traditional performance measures and established eye tracking metrics. However, with respect to the analysis of visual strategies, both studies included visual analysis for the qualitative assessment of recorded scanpaths. We argue that such observations will become more important for experiments whenever eye tracking is involved. Furthermore, additional data (e.g., think-aloud, interaction logs) will be necessary to include in a data integration step to provide a new, more thorough view on the participant's behavior.

Cognitive Models

Cognitive models to predict the scanpath of a participant and the efficiency of wayfinding tasks would be beneficial for the design of metro maps in Scenario 1. Although different strategies for solving the task could be identified, a generalized model was not included in the results of the study. The study was one of the first in this domain where it was important to identify general strategies. For a comprehensive model, additional data for different levels of expertise might be necessary. Here, map designers and map readers are two different target groups that potentially focus on different aspects of the map and viewing tasks might differ significantly between such groups. An implicit model of strategies was applied for the manual annotation of paths, imprecise measures of line tracing. Future models could also consider psycho-physical measures, for example, just noticeable differences to be able to separate close-by metro lines. In the wild, saliency models will also play an important role for the orientation while searching for start and goal locations.

The design of the study in Scenario 2 was based on some assumptions made from theory and observations in pilot experiments. Netzel et al. provided a handcrafted model (Fig. 10.6) on the different strategies during the reading process

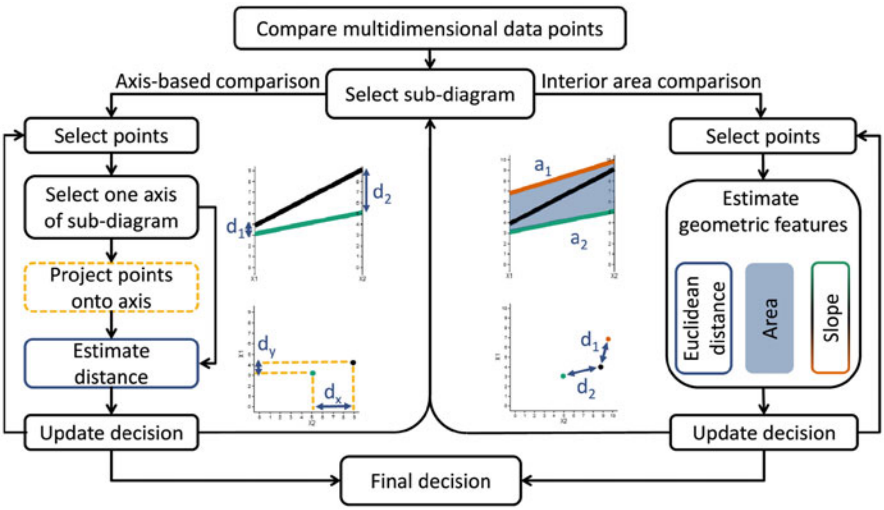

Fig. 10.6 Strategy model for the visual comparison of multidimensional data points with parallel coordinate plots. Netzel et al. [34] identified two strategies, i.e., axis-based and interior area comparison, and comprised them in a handcrafted behavior model. Figure reprinted by permission of Elsevier from Netzel et al. [34]. Copyright 2017 Zhejiang University and Zhejiang University Press. Licensed under the CC BY-NC-ND 4.0 license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

of the visualization. This model was guided by the hypotheses of the study. In future research, such models could be generated more systematically, informed by theoretical perceptual or cognitive models from psychology.

10.6 Call for Actions

Based on our previous observations, we have identified the following interesting points for future development and calls for actions.

Translational Research
Many early guidelines in visualization were informed by perceptual and cognitive science, like efficient visual encoding, Gestalt laws [24, 43], or feature integration theory [16, 39]. However, there is lack of guidelines that inform design decisions for visual analytics systems [37] since current cognitive models are good at explaining cognitive processes on well-defined tasks and simple visual stimuli but are less applicable to the aforementioned scenarios that have become prevalent in today's systems [17]. This line of research offers great potential for translational studies since psychology and visualization research would equally benefit from such results. Distributed cognition could be a promising approach toward translational studies of that kind since it provides a more holistic view of the way humans reason and think. It acknowledges the fact that humans live in materialistic and social environments, and thus, it emphasizes the importance of contextual factors in human cognition [21].

Best Practices
This book chapter only provided a high-level conceptual view on evaluation strategies. So far, our envisioned evaluation strategies have not yet been implemented in real-world empirical studies. Many challenges are left unanswered, such as how to practically design, conduct, and evaluate data-rich empirical studies. It is particularly important to provide researchers with a tool set to perform sophisticated data analysis with minimal effort. There is also need for the whole community of researchers to agree upon a proper way to report results of such studies.

Interdisciplinary Research Venues
Psychologists' core topics are often disconnected from topics relevant for visualization research. Yet, there are some successful examples of combining communities, for example, at the *Symposium on Eye Tracking Research and Applications (ETRA)*. Such events provide great opportunities for interdisciplinary discourse and establishing collaborations. However, publication strategies and research topics might significantly differ between communities. Hence, a fusion of expertise just by project collaborations might cover some research questions, but from a long-term perspective, other solutions are necessary. A key question, of course, is *How can we integrate the expertise from both research fields in a common research endeavor?* We think that activities such as this workshop or our own experience

with the ETVIS workshop[1] and joint research centers (like SFB-TRR 161[2]) are a good way to go but are alone not sufficient and need further action. Building a research area of visualization psychology could be a viable means, for example, by establishing publication and other presentation opportunities that work for visualization researchers, psychologists, and social scientists alike, by setting up a canon of teaching new students, and by lobbying for funding possibilities for such interdisciplinary work.

Psychology Education
Although many design principles are based on perceptual and cognitive theories, in-depth psychological background knowledge is often not part of the education for visualization. Researchers starting with eye tracking studies are confronted with learning eye tracking methodology, which is, starting with proper calibration to a comprehensive analysis of the data, a complex field on its own. As a consequence, deeper knowledge of a whole new research field, i.e., psychology, is hard to achieve within the short time span of an average PhD student's career.

Acknowledgments This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project ID 251654672 – TRR 161 (project B01) and under Germany's Excellence Strategy—EXC 2120/1 – 390831618.

References

1. J. R. Anderson, M. Matessa, and C. Lebiere. ACT-R: A theory of higher level cognition and its relation to visual attention. *Human–Computer Interaction*, 12(4):439–462, 1997.
2. R. Arias-Hernandez, L. T. Kaastra, T. M. Green, and B. Fisher. Pair analytics: Capturing reasoning processes in collaborative visual analytics. In *2011 44th Hawaii International Conference on System Sciences*, pages 1–10, 2011.
3. T. Blascheck, F. Beck, S. Baltes, T. Ertl, and D. Weiskopf. Visual analysis and coding of data-rich user behavior. In *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 141–150, 2016.
4. T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-the-art of visualization for eye tracking data. In *EuroVis (STARs)*, 2014.
5. T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. Visualization of eye tracking data: A taxonomy and survey. *Comput. Graph. Forum*, 36(8):260–284, 2017.
6. M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliva. Beyond memorability: Visualization recognition and recall. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):519–528, 2016.
7. M. Burch, R. Woods, R. Netzel, and D. Weiskopf. The challenges of designing metro maps. In N. Magnenat-Thalmann, P. Richard, L. Linsen, A. C. Telea, S. Battiato, F. H. Imai, and J. Braz, editors, *Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*, pages 197–204. SciTePress, 2016.

[1] https://www.etvis-workshop.org.

[2] https://www.sfbtrr161.de/.

8. V. J. Caracelli and J. C. Greene. Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 15(2):195–207, 1993.
9. S. K. Card, T. P. Moran, and A. Newell. The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, 23(7):396–410, 1980.
10. S. Carpendale. Evaluating information visualizations. In A. Kerren, J. T. Stasko, J. Fekete, and C. North, editors, *Information Visualization-Human-Centered Issues and Perspectives*, volume 4950 of *Lecture Notes in Computer Science*, pages 19–45. Springer, 2008.
11. A. T. Duchowski. *Eye Tracking Methodology - Theory and Practice, Third Edition*. Springer, 2017.
12. S. Elling, L. Lentz, and M. de Jong. Retrospective think-aloud method: Using eye movements as an extra cue for participants' verbalizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 1161–1170, New York, NY, USA, 2011. Association for Computing Machinery.
13. G. Ellis and A. Dix. An explorative analysis of user evaluation studies in information visualisation. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, page 1–7, New York, NY, USA, 2006. Association for Computing Machinery.
14. B. Fisher, T. M. Green, and R. Arias-Hernández. Visual analytics as a translational cognitive science. *Topics in Cognitive Science*, 3(3):609–625, 2011.
15. J. H. Goldberg and J. I. Helfman. Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization*, 10(3):182–195, 2011.
16. C. G. Healey and J. T. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, 2012.
17. M. Hegarty. The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science*, 3(3):446–474, 2011.
18. M. Hegarty, M. S. Canham, and S. I. Fabrikant. Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1):37, 2010.
19. C. Heine. Towards modeling visualization processes as dynamic Bayesian networks. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1000–1010, 2021.
20. W. Huang, P. Eades, and Seok-Hee Hong. A graph reading behavior: Geodesic-path tendency. In *2009 IEEE Pacific Visualization Symposium*, pages 137–144, 2009.
21. E. Hutchins. Distributed cognition. *International Encyclopedia of the Social and Behavioral Sciences. Elsevier Science*, 138, 2000.
22. L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, 2001.
23. R. B. Johnson, A. J. Onwuegbuzie, and L. A. Turner. Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2):112–133, 2007.
24. K. Koffka. *Principles of Gestalt Psychology*, volume 44. Routledge, 2013.
25. A. Kumar, D. Mohanty, K. Kurzhals, F. Beck, D. Weiskopf, and K. Mueller. Demo of the EyeSAC system for visual synchronization, cleaning, and annotation of eye movement data. In *ACM Symposium on Eye Tracking Research and Applications*, ETRA '20 Adjunct, New York, NY, USA, 2020. Association for Computing Machinery.
26. K. Kurzhals, M. Burch, and D. Weiskopf. What we see and what we get from visualization: Eye tracking beyond gaze distributions and scanpaths, 2020.
27. K. Kurzhals, B. Fisher, M. Burch, and D. Weiskopf. Eye tracking evaluation of visual analytics. *Information Visualization*, 15(4):340–358, 2016.
28. C. Körner. Eye movements reveal distinct search and reasoning processes in comprehension of complex graphs. *Applied Cognitive Psychology*, 25(6):893–905, 2011.
29. C. Körner, M. Höfler, B. Tröbinger, and I. D. Gilchrist. Eye movements indicate the temporal organisation of information processing in graph comprehension. *Applied Cognitive Psychology*, 28(3):360–373, 2014.

30. H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.
31. L. A. McNamara and W. A. Stubblefield. Workshop proposal: Visualization and the context of work-qualitative research methods for design deployment evaluation. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2011.
32. E. Moseholm and M. D. Fetters. Conceptual models to guide integration during analysis in convergent mixed methods studies. *Methodological Innovations*, 10(2):2059799117703118, 2017.
33. R. Netzel, B. Ohlhausen, K. Kurzhals, R. Woods, M. Burch, and D. Weiskopf. User performance and reading strategies for metro maps: An eye tracking study. *Spatial Cognition & Computation*, 17(1-2):39–64, 2017.
34. R. Netzel, J. Vuong, U. Engelke, S. I. O'Donoghue, D. Weiskopf, and J. Heinrich. Comparative eye-tracking evaluation of scatterplots and parallel coordinates. *Visual Informatics*, 1(2):118–131, 2017.
35. D. Peebles. A cognitive architecture-based model of graph comprehension. In *11th International Conference on Cognitive Modeling, Berlin*, pages 37–42, 2012.
36. M. Raschke, T. Blascheck, and T. Ertl. Cognitive ergonomics in visualization. In A. Ebert, G. C. van der Veer, G. Domik, N. D. Gershon, and I. Scheler, editors, *Building Bridges: HCI, Visualization, and Non-formal Modeling*, pages 80–94, Berlin, Heidelberg, 2014. Springer.
37. J. Scholtz, C. Plaisant, M. Whiting, and G. Grinstein. Evaluation of visual analytics environments: The road to the visual analytics science and technology challenge evaluation methodology. *Information Visualization*, 13(4):326–335, 2014.
38. B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, page 1–7, New York, NY, USA, 2006. Association for Computing Machinery.
39. A. Treisman. Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, 31(2):156–177, 1985.
40. E. R. A. Valiati, C. M. D. S. Freitas, and M. S. Pimenta. Using multi-dimensional in-depth long-term case studies for information visualization evaluation. In *Proceedings of the 2008 Workshop on BEyond Time and Errors: Novel EvaLuation Methods for Information Visualization*, BELIV '08, New York, NY, USA, 2008. Association for Computing Machinery.
41. S. Vogl. Integrating and consolidating data in mixed methods data analysis: Examples from focus group data with children. *Journal of Mixed Methods Research*, 13(4):536–554, 2019.
42. D. Weiskopf. *Vis4Vis: Visualization for (Empirical) Visualization Research*, pages 209–224. Springer International Publishing, Cham, 2020.
43. M. Wertheimer. Laws of organization in perceptual forms. *A source book of Gestalt Psychology*, 1, 1923.

Part III
Visualization Psychology from an Experimental Perspective

Strong theory can shape the ways we build visualizations. However, theories make assumptions about the state of the world, such as all people sharing the same cognitive traits or a common goal, that may fail in practice. Visualization use is complex: people and data exhibit unique characteristics, and what people look to learn from data often changes dynamically over the course of an analysis. Given this complexity, how do we know if a theory holds for a given data set, analysis goal, or visual design? Visualization's complexity provides researchers with a vast space of opportunities to test theories using a range of experimental techniques. The final part of this book focuses on how we apply innovative experimental methods to better understand visualization psychology.

Visualization research has historically been grounded in design and engineering rather than experimental science, focusing on novel techniques and algorithms for surfacing invisible patterns in data and borrowing inspiration from psychology to fill in key gaps in our knowledge. However, a growing body of work inspired by psychological methods employs empirical techniques to better characterize how people use visualizations. Such controlled studies enable researchers to investigate how well theoretical assertions translate into the complex space of visualization use. Visualization can draw on methods from psychology to better understand the cognitive implications of specific phenomena and lend credence to (or challenge) long-held design heuristics. Psychology can draw on a range of interface designs, including visual representations and interactions, to probe theoretical questions about how people interpret visual information.

The following five chapters cover a range of subjects from specific perceptual characteristics associated with simple visualizations such as scatterplots to higher-level cognitive processes involved in translating data into knowledge. We open with an experiment examining the role of *task*—the knowledge someone seeks to gain from a visualization—in understanding what information people attend to in visualizations (Chap. 11). The subsequent chapter extends this idea to investigate how changing properties of data may interact with task to shift this attention (Chap. 12). The focus then moves to individual differences reflected by specific personality traits, namely *conscientiousness*—an individual's tendency to be diligent and organized in their work (Chap. 13). The book concludes with two chapters examining higher-level cognitive processes in visualization. Chapter 14 explores how the ways we communicate uncertain information change our decision-making abilities. We close with an investigation of how individual observations come together to build knowledge through a process called *sensemaking* (Chap. 15).

The work in this section emphasizes the importance of coupling theory and practice through empirical methods. Looking forward, the discussions and approaches introduced in this chapter exemplify how theories and methods from psychology can help us understand how people use visualizations. The outcomes of such experiments can help build and refine theories of how people work with visual information and craft new guidelines for precise and effective data communication that consider a myriad of factors when working with visualizations.

Part III Editor: Danielle Albers Szafir
University of North Carolina-Chapel Hill, USA

Chapter 11
Task Matters When Scanning Data Visualizations

Laura E. Matzen, Kristin M. Divis, Deborah A. Cronin, and Michael J. Haass

Abstract One of the major challenges for evaluating the effectiveness of data visualizations and visual analytics tools arises from the fact that different users may be using these tools for different tasks. In this chapter, we present a simple example of how different tasks lead to different patterns of attention to the same underlying data visualizations. The experiment used eye tracking to record where people looked in scatterplot visualizations when given different tasks. We argue that the general approach used in this experiment could be applied systematically to task and feature taxonomies that have been developed by visualization researchers. Using eye tracking to study the impact of common tasks on how humans attend to common types of visualizations will support a deeper understanding of visualization cognition and the development of more robust methods for evaluating the effectiveness of visualizations.

11.1 Introduction

What makes a data visualization effective? Evaluating visualizations can be very challenging and is the subject of much research and debate [6, 20, 29, 33, 39]. Members of the visualization research community have called for evaluation approaches that assess how well visualizations support their viewers' cognitive needs [5, 11, 26, 41]. From this perspective, an effective visualization successfully exploits its viewers' cognitive processes to draw their attention to relevant information, minimize their attention to irrelevant information, and increase the likelihood of correct interpretation. In order to meet those requirements, visualization designers

L. E. Matzen (✉) · K. M. Divis · M. J. Haass
Sandia National Laboratories, Albuquerque, NM, USA
e-mail: lematze@sandia.gov; kmdivis@sandia.gov

D. A. Cronin
University of California, Davis, CA, USA

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_11

need to be able to account for the experience, expectations, and biases of the viewer in addition to the low-level, perceptual properties of the data visualization.

There is a growing body of research on how the perceptual aspects of visualizations influence viewers' cognitive processes. For example, researchers have demonstrated that increasing the visual saliency of task-relevant information can improve task performance [10, 14, 15, 18, 28, 30, 38] and that changing the visual representation of a dataset can change how viewers interpret it [8] and their biases in interpretation [31]. However, there has been relatively little research on how different tasks impact viewers' attention to different aspects of visualizations. Visual saliency is driven by low-level visual features, such as color and contrast, which are easily measured. There are multiple models that use these features to predict which parts of an image (e.g., [13, 19]) or a data visualization [25] will draw viewers' attention. However, these models capture only the stimulus-driven aspects of human attention. The viewer's task drives their visual attention from the top down and can override these stimulus-driven effects [7, 15, 21, 27].

Prior eye-tracking research has shown that changing a person's task changes where they look in a natural scene (e.g., [16]). It stands to reason that changing a person's task would change where they look in data visualizations as well. However, the visual properties of many common types of data visualizations are quite different from those of natural scenes [24, 25]. In visualizations, visual cues such as color and shape are used deliberately, to convey specific information to the viewer. In a well-designed visualization, those same features might help to support multiple tasks, such as identifying trends, making comparisons, or looking up specific values. Visualizations often incorporate text, such as titles and axis labels, which draw the viewer's attention even when they do not have high visual saliency [25]. These text cues may be crucial for understanding the visualization, so viewers may read them regardless of their task. Thus, we might expect that changing the task might have little impact on patterns of attention to data visualizations.

It is important to understand whether or not this is the case, because it has important implications for evaluations of data visualizations. Saliency maps are useful for evaluating whether the most important features of a visualization are also the most visually salient. However, different aspects of the visualization may be important for different tasks. If those task differences impact which parts of the visualization the viewers pay attention to, saliency-based evaluations become less meaningful.

Since data visualizations are often designed with specific tasks in mind (i.e., identifying trends or clusters, comparing values, or identifying outliers), it may be possible to incorporate some task-based features into visual saliency models for common types of data visualizations (reference to Chap. 12). But first we need to understand how much task influences patterns of eye movements for data visualizations.

In this chapter, we present a simple experiment to illustrate this point. In this experiment, participants viewed scatterplots and were tasked with describing the trend for half of the stimuli and describing the outliers for the other half (with the stimulus-task groupings counterbalanced across participants). We hypothesized that the two different tasks would lead to two different patterns of eye movements in response to the same physical stimuli, with participants allocating their attention to features of the plots in response to the demands of the task.

11.2 An Experiment on the Impact of Task

Thirty participants (7 males; mean age = 29.57, stdev = 13.79) were recruited from the University of Illinois community and were compensated $20 for their time. All participants were tested for color vision deficiencies (24-plate Ishihara Test; Ishihara, 1972) and near vision acuity prior to completing the study. This experiment was part of a larger project, and these same participants also completed the cluster comparison experiment reported in Chap. 12 of this book [9].

11.2.1 Materials

The participants saw 32 scatterplots consisting of four unique plots for each of eight types of trends: positive linear, negative linear, flat, sinusoidal (cyclical), positive logarithmic, negative logarithmic, positive quadratic, and negative quadratic. All of the stimuli were created in R Software [37] from simulated data, using the standard plotting function to create simple scatterplots. Each stimulus consisted of 100 data points (open circles) plotted in black on a white background. Each graph was labeled with a title and axis labels. The simulated data were drawn from Gaussian distributions, and the data points representing the trend were constrained to fall within two vertical standard deviations of the trend function. Half of the plots of each type had two outliers and half had four. The outliers were at least four standard deviations away from the trend function. Each image was 1000 pixels in height. Examples of the stimuli are shown in Fig. 11.1.

The task (describing the trend or describing outliers), data pattern (positive or negative linear, sinusoidal, positive or negative logarithmic, flat, and positive or negative quadratic), and the number of outliers (2 or 4) were manipulated within subjects.

11.2.2 Procedure

The participants completed the experiment in a dimly lit sound attenuating booth, seated at a nominal viewing distance of 0.8 meters from a computer monitor (0.932 $\times$ 0.523 meters; 1920 $\times$ 1080 pixels). Their eye movements were recorded with a Smart Eye Pro eye tracker. Prior to completing each task, participants underwent the standard Smart Eye camera setup procedure and 9-point calibration.

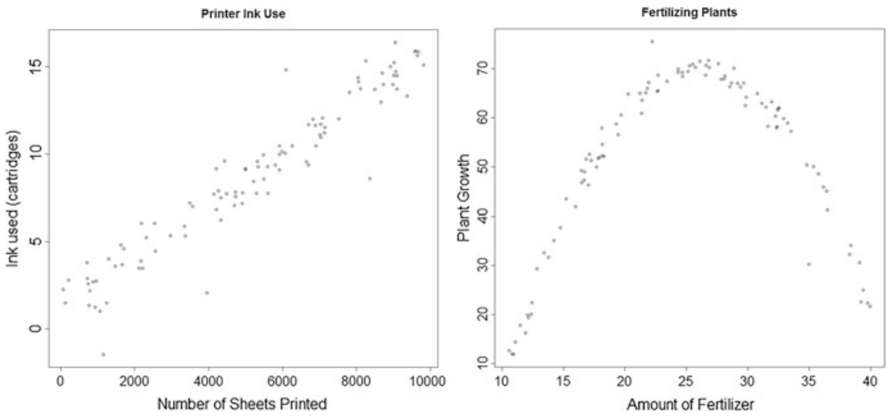

Fig. 11.1 Two representative scatterplot stimuli. The example on the left shows a positive linear trend with four outliers. The example on the right shows a quadratic trend with two outliers

Each trial began with a fixation cross that was presented in the center of the screen for 1000 milliseconds (ms). Then one of the scatterplot stimuli was presented in the center of the screen on a white background. There was a 500 ms interstimulus interval between trials.

Participants viewed each stimulus for up to 10 seconds, but they could end the trial sooner than 10 seconds if they were ready to make their response. When participants were ready to respond (or when the 10-second time limit had been reached), they advanced to a blank white screen. Then they verbally described the trend or the outliers in the scatterplot (depending on condition) to the experimenter, who wrote down their responses and asked for further clarification if necessary. The participants' verbal responses were also captured via audio recordings. The participants were not given feedback about their responses.

The task was divided into three sections: a practice session and two blocks of stimuli. During the practice session, participants worked through two example stimuli and were given the opportunity to ask the experimenter for further clarification. In the first block of experimental trials, half of the participants described the trend in the data and half of the participants described the outliers. The participants then switched tasks for the second block of stimuli. Each block contained 16 scatterplots, with two of each of the eight types of trends, one of which had two outliers and one of which had four.

11.3 Results

Two raters independently scored each participant's description of each scatterplot to ensure that the participants produced reasonable responses for each task. The responses indicated that the participants understood both tasks and appropriately focused their responses on the trend or on the outliers depending on the task condition. The participants accurately described the trend for 87% of the trials. When describing the outliers, participants missed one or more of the outliers on 200 out of 398 trials (50.3%). There were only 20 trials in which participants falsely identified an extra outlier (5.0%).

Fixations were calculated using Smart Eye's default algorithm, where any sample for which the velocity over the preceding 200 ms is less than 15°/s is deemed a fixation. The first fixation in each trial was excluded from the analysis, as was any fixation with a duration less than 100 ms. Each stimulus was divided into the following regions of interest (ROIs): Outliers, Trend, Title, Axes, Axis Labels, and Other. The "Other" ROI corresponded to the white space inside of the scatterplot that did not contain any data points. An example of the ROIs for one stimulus is shown in Fig. 11.2. The ROIs were defined by drawing a polygon around each element of the graph, conforming closely to the edges of the object or text. When assigning participants' fixations to the ROIs, a buffer equivalent to one degree of visual angle was added to the coordinates of each ROI. If the center of a fixation fell within the ROI or within the buffer around the ROI, it was assigned to that ROI. Any fixations

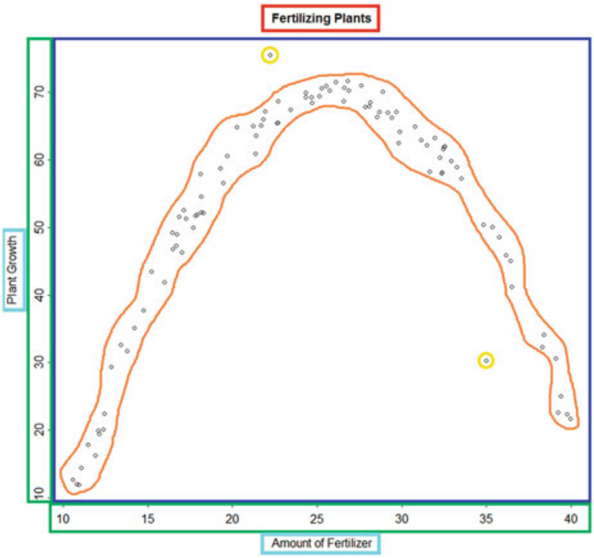

Fig. 11.2 An example of the regions of interest used in the eye-tracking analysis, including the Title (red), Axes (green), Axis Labels (turquoise), Trend (orange), Outliers (yellow), and Other (blue)

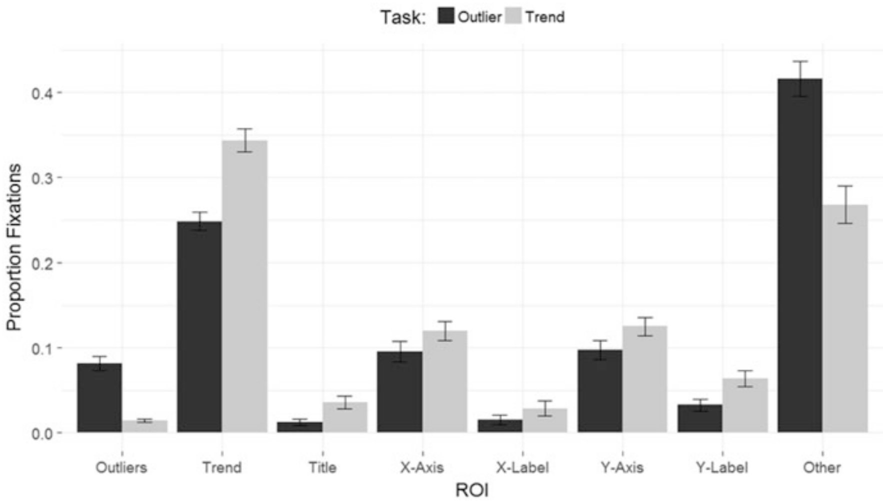

Fig. 11.3 Average proportion of fixations to each region of interest for the outlier and trend tasks

within the white space inside of the scatterplot that could not be assigned to the Outlier or Trend ROIs were assigned to "Other."

In this experiment, we were interested in how visual attention, as reflected by patterns of eye movements, changed in response to different tasks. A mixed effects model (fit with the lme4 package in R software; [2]) with a fixed effect for task and random intercepts for participant and stimulus (using Satterthwaite approximation for degrees of freedom; see [23]) revealed that overall, participants had more fixations per trial and shorter fixation durations in the outlier task than in the trend task (all t's > 10.00, p's < 0.001). This pattern of shorter, more numerous fixations in the outlier task condition is consistent with a visual search process [34].

Task also influenced which regions of the graph participants fixated most frequently. The proportion of fixations to each type of ROI was calculated for each participant and stimulus. The results are shown in Fig. 11.3. The participants' task had a substantial impact on where they allocated their attention within the scatterplots. A mixed effects model was used to predict the proportion of fixations as a function of the fixed effects of task and ROI, with random intercepts for subject and stimulus. For the trend description task, there were significantly higher proportions of fixations to the Trend ROI as well as the Title, Axis, and Axis Label ROIs. For the outlier description task, the proportion of fixations was significantly higher for the Outlier ROI and the Other ROI (all t-statistics > 2.00 and p-values < 0.05). The high proportion of fixations to the Other ROI was likely due to participants searching the graphs for outliers as well as the relatively small size of the Outlier ROIs.

In addition to examining the proportion of fixations in each ROI, we also investigated how attention to the ROIs unfolded over time under the different task conditions. Figure 11.4 shows the probability of visiting each ROI over the time

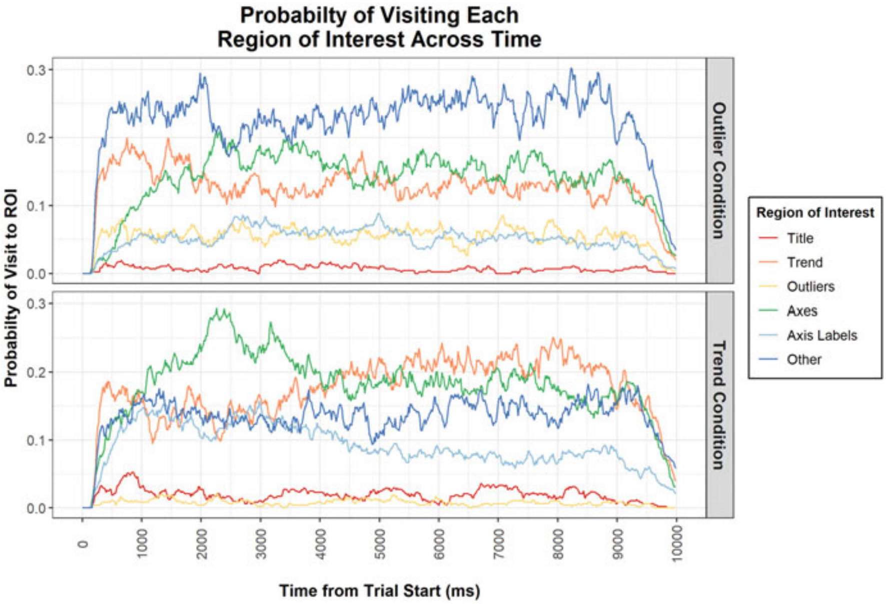

Fig. 11.4 The probability of visiting each region of interest over the course of the trial for the two task conditions

course of the trial for each task. The two tasks produced very different patterns of fixations throughout the trial. In the trend task, participants' attention went first to the Trend ROI, then to the axes. Later during the trial, there was a high probability of the participants visiting both the trend and the axes. The probability of the participants visiting the Outlier ROIs was low throughout the time course of the trial. In the outlier task, the probability of participants visiting the Outlier ROIs was higher throughout the trial, as was the probability of visiting the Other ROI, which has the highest probability of any region throughout most of the trial. As described above, this is likely due to participants searching for the outliers. The participants also had a relatively high probability of looking at the trend and the axes in this task condition, but that probability was lower than in the trend description condition, particularly later in the trial.

11.4 Discussion

In this task, we observed differences in patterns of eye movements when participants were given different tasks using the same data visualizations. The participants performed the two different tasks successfully, although some outliers were overlooked. The eye-tracking data indicated that there were differences in the overall allocation of attention to different elements within the graphs, in addition to different patterns of attention over the course of a trial. The trend and axes received a relatively high proportion of the participants' fixations, regardless of condition, but the fixations on the outliers were dramatically influenced by the participants' task.

We also found task-driven differences in the number and duration of fixations: participants tended to make many short fixations in the outlier condition while making fewer, longer fixations in the trend condition. The eye movement patterns in the outlier task are consistent with participants engaging in visual search, a task that typically encourages more eye movements and short fixations [7, 27]. The long-duration fixations characterizing the trend condition may have better allowed participants to extract the "gist" of the scatterplot (e.g., via ensemble encoding; see [35]).

The classic experiment by Yarbus (1967) demonstrated that a person's task and goals impacted their eye movements, and numerous subsequent studies have found similar effects for natural scenes [3, 12, 17]. Our experiment demonstrates that the same is true for data visualizations. While this is a very straightforward example, it shows how research along these lines could have implications for visualization design and evaluation. In this case, if the designer knew that the outliers might be important to users' tasks, they could choose visual representations that make the outliers more salient and easier to locate. For more complex visualizations, assessing changes in fixation patterns or scan paths for different tasks could reveal patterns that might not be easy to predict. This kind of research could also reveal cases where visual-spatial and cognitive biases [32, 40] are likely to impact viewers' interpretations.

The visualization community has developed numerous taxonomies that break down common visualization types and link them to common tasks, cf. [1, 4, 22, 29, 36]. These taxonomies could serve as an entry point for visual cognition researchers and as a framework for systematic experimentation. For example, [22] provide a taxonomy of objects and tasks that are common in graph visualization. Researchers could use eye tracking to test how different tasks change patterns of fixations to the same underlying graph objects. This research could support the development of new visualization methods for that domain. It could also support the development of widely applicable evaluation methods that take both bottom-up and top-down features into account to determine whether a graph visualization effectively meets the cognitive needs of its intended users.

Acknowledgments This work was supported by the Laboratory Directed Research and Development (LDRD) Program at Sandia National Laboratories. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This chapter describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. SAND2020-7549 C.

References

1. R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *Proc. IEEE Symposium on Information Visualization (INFOVIS)*, pages 111–117, 2005.
2. D. Bates, M. Maechler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.
3. A. Borji and L. Itti. Defending Yarbus: Eye movements reveal observers' task. *Journal of Vision*, 14(3):29–29, 2014.
4. M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.
5. S. Card, J. D. Mackinlay, and B. Schneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco, 1999.
6. S. Carpendale. Evaluating information visualizations. In *Information visualization*, pages 19–45. Springer, Berlin, 2008.
7. M. Castelhano, M. Mack, and J. Henderson. Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):id.6, 2009.
8. M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, 2014.
9. K. M. Divis, L. E. Matzen, M. J. Haass, and D. A. Cronin. Perceptual biases in scatterplot interpretation. In D. Albers Szafir, R. Borgo, M. Chen, D. Edwards, B. Fisher, and L. Padilla, editors, *Visualization Psychology*. Springer, 2022.
10. M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri. Perception of average value in multiclass scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 12(19):2316–2325, 2013.
11. T. M. Green, W. Ribarsky, and B. Fisher. Building and applying a human cognition model for visual analytics. *Information Visualization*, 8:1–13, 2009.
12. A. Haji-Abolhassani and J. J. Clark. An inverse Yarbus process: Predicting observers' task from eye movement patterns. *Vision Research*, 103:127–142, 2014.
13. J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *Advances in Neural Information Processing Systems*, 19, 2006.
14. C. G. Healey and J. T. Enns. On the use of perceptual cues and data mining for effective visualization of scientific datasets. In *Proc. Graphics Interface*, pages 177–184, 1998.
15. M. Hegarty, M. S. Canham, and S. I. Fabrikant. Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 36(1):37–53, 2010.
16. J. M. Henderson, J. R. Brockmole, M. S. Castelhano, and M. Mack. Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray, and R. L. Hill, editors, *Eye Movements: A Window on Mind and Brain*, pages 537–562. Elsevier, 2007.
17. J. M. Henderson, S. V. Shinkareva, J. Wang, S. G. Luke, and J. Olejarczyk. Predicting cognitive state from eye movements. *PloS One*, 8(5):e64937, 2013.
18. V. Interrante. Harnessing natural textures for multivariate visualization. *IEEE Computer Graphics and Applications*, 20(6):6–11, 2000.
19. L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2:194–203, 2001.
20. H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.
21. M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision Research*, 41:3559–3565, 2001.

22. B. Lee, C. Plaisant, C. S. Parr, J. D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proc. AVI workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*, pages 1–5, 2006.
23. S. G. Luke. Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4):1494–1502, 2017.
24. L. E. Matzen, M. J. Haass, K. M. Divis, and M. C. Stites. Patterns of attention: How data visualizations are read. In *International Conference on Augmented Cognition*, pages 176–191. Springer, 2017.
25. L. E. Matzen, M. J. Haass, K. M. Divis, Z. Wang, and A. T. Wilson. Data visualization saliency model: A tool for evaluating abstract data visualizations. *Transactions on Visualization and Computer Graphics*, 24(1):563–573, 2018.
26. L. Micallef, G. Palmas, A. Oulasvirta, and T. Weinkauf. Towards perceptual optimization of the visual design of scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 23(6), 2017.
27. M. Mills, A. Hollingworth, S. Van der Stigchel, L. Hoffman, and M. D. Dodd. Examining the influence of task set on eye movements and fixations. *Journal of Vision*, 11(8), 2011.
28. M. N. Milosavljevic, V. Koch, C., and R. A. Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology*, 22(1):67–74, 2012.
29. T. Munzner. *Visualization Analysis and Design*. CRC Press, Boca Raton, FL, 2014.
30. C. Nothelfer, M. Gleicher, and S. Franconeri. *Redundant encoding strengthens segmentation and grouping in visual displays of data*. Journal of Experimental Psychology: Human Perception and Performance, 2017.
31. L. Padilla, S. H. Creem-Regehr, and W. Thompson. The powerful influence of marks: Visual and knowledge-driven processing in hurricane track displays. *Journal of Experimental Psychology: Applied*, 26(1):1–15, 2020.
32. L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, 3(1):29., 2018.
33. C. Plaisant. The challenge of information visualization evaluation. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 109–116. ACM, 2004.
34. K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 1998.
35. R. A. Rensink. The nature of correlation perception in scatterplots. *Psychonomic Bulletin & Review*, 24(3):776–797, 2017.
36. B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings IEEE Symposium on Visual Languages*, pages 336–343. IEEE, 1996.
37. R. D. C. Team.
38. E. E. Veas, E. Mendeaz, S. K. Feiner, and D. Schmalstieg. Directing attention and influencing memory with visual saliency modulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computer Systems*, pages 1471–1480, 2011.
39. E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, 2018.
40. E. Wall, L. M. Blaha, C. L. Paul, K. Cook, and A. Endert. Four perspectives on human bias in visual analytics. In *Cognitive Biases in Visualizations*, pages 29–42. Springer, Cham., 2018.
41. J. S. Yi, Y. A. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: how do people gain insights using information visualization? In *Proceedings of the 2008 Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*, page 4. ACM, 2008.

Chapter 12
Perceptual Biases in Scatterplot Interpretation

Kristin M. Divis, Laura E. Matzen, Michael J. Haass, and Deborah A. Cronin

Abstract The scientific community heavily relies on data visualizations for communication, precipitating the need to better understand what makes for a "good" data visualization (e.g., informing data visualization evaluation or saliency tools). In addition to the underlying mathematical or bottom-up properties of a visualization, designers must also account for the influence from top-down factors such as the viewer's goal or perceptual biases. In the current study, we asked participants to compare two clusters of data points in a scatterplot (similar to a multidimensional data reduction comparison task). We manipulated both the underlying mathematical properties of the data set and the decision-making task. We found evidence for visual–spatial biases and differences in overt attention patterns (eye movements), even when the compared clusters were mathematically equivalent. These results demonstrate how task and perceptual biases may impact viewers' understanding of relationships between variables in a multidimensional space, possibly leading to error or systematic biases in analysts' interpretation of the plotted data.

12.1 Introduction

Data visualizations serve an important role in scientific inquiry and communication. They are widely relied upon throughout the scientific community—from interpretation of findings in academic journals to high consequence decision-making in domains such as disaster response and national security. A good data visualization can allow its viewer to quickly identify important trends, interesting groups, or outliers in a large data set or to rapidly grasp the take-home message of an entire study. But what makes a data visualization "good"? Evaluating

K. M. Divis (✉) · L. E. Matzen · M. J. Haass
Sandia National Laboratories, Albuquerque, NM, USA
e-mail: kmdivis@sandia.gov; lematze@sandia.gov

D. A. Cronin
University of California, Davis, CA, USA

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_12

visualizations can be very challenging and is the subject of much research and debate [8, 34, 53, 58, 65, 69]. Members of the visualization research community are calling for evaluation of visualizations by examining the extent to which they support their viewers' cognitive needs [7, 21, 48, 76]. From this perspective, a "good" visualization successfully exploits its viewers' cognitive processes to draw their attention to relevant information, minimize their attention to irrelevant information, and increase the likelihood of correct interpretation. In order to meet those requirements, visualization designers need to be able to account for the experience, expectations, and biases of the viewer in addition to the low-level, perceptual properties of the data visualization. However, there is not yet a coherent theory of visualization cognition for designers to draw upon (cf. [55]). While it is clear that many factors influence viewers' interpretation of visualizations, whether by improving viewers' ability to understand the underlying data (cf. [70]) or by inducing cognitive biases [57], there are many aspects of visualization cognition that have not yet been explored by researchers. Gaining a deeper understanding of how factors such as the viewers' task and visual–spatial biases impact their interpretation of visualizations will help to advance visual cognition research as well as have practical applications in designing and evaluating visualizations. In the research reported here,[1] we aim to begin addressing these questions for interpretation of clusters within scatterplots, a type of data visualization widespread in engineering practice and the scientific literature.

12.2 Bottom-Up and Top-Down Attention in Data Visualizations

The patterns of attention that are observed when viewers interpret data visualizations can be understood from the perspective of bottom-up and top-down visual attention. This approach grounds research on visualizations in the larger field of visual cognition, in addition to providing avenues for developing practical, cognition-based tools for evaluating visualizations. Bottom-up visual attention is drawn automatically to salient, low-level visual features (those with physical properties such as color or line orientation that differ from their surroundings); top-down visual attention is driven by the observer's goals, expectations, and prior experience [25, 73–75].

Several research groups have suggested that visual saliency may be a useful tool for evaluating the extent to which visualizations support their users' cognitive processes [30, 33, 45, 61, 68]. Models of bottom-up visual attention based on visual saliency make reasonable predictions of where participants will look in natural scenes (e.g., [28]), man-made scenes (e.g., [3]), and abstract data visualizations

[1] This study was part of a larger project examining human comprehension of data visualizations; see also Chapter 11 in this book [43].

(e.g., [45]). An effective visualization should draw viewers' attention to the most important or task-relevant information. Indeed, increasing the visual saliency of task-relevant information has been shown to affect user performance in day-to-day decision-making tasks [52]; in virtual reality [67]; and, importantly, in data visualization interpretation [20, 23, 24, 26, 54]. Visual saliency can also contribute to cognitive biases when viewers rely too heavily on salient features when interpreting visualizations [57]. Thus, a reasonable first pass at assessing the effectiveness of a visualization is to use saliency models to confirm that the aspects of a visualization that the designer and users deem most important are also the most visually salient.

However, traditional visual saliency models (e.g., [28]) do not include any information about top-down visual attention. Top-down processing has a strong influence over where people choose to look and can override the bottom-up draw of visually salient features [35, 36]. For instance, social cues [5], differing task priorities [9, 13, 19, 24, 36, 51], expertise [38], and prior experiences [11, 41] can all drive top-down attention. Top-down attention is particularly important in understanding how people process data visualizations. Unlike natural scenes, visualizations often require the viewer to have some prior knowledge about how to make sense of a set of abstract data visualizations. They may arbitrarily map meaning onto specific colors and shapes [53], include text and numbers [44], and draw on various conventions related to the use of axes and legends, visual representations of uncertainty, and so forth. The viewers' goals influence where they look in a visualization, and recent studies have demonstrated that participants with different goals looking at the same visualization will inspect it differently [49, 50]. In addition, visualizations are often developed for specific groups of people with expertise in a scientific or technical area, and they may be used differently by people with different levels of expertise [34, 46].

Recent research has used the dual-process account of decision-making to propose a cognition-based framework for investigating how people make decisions using visualizations [56]. Dual-process theories posit that that people make decisions in one of two ways. Type 1 decisions are made rapidly and automatically [17]. Various researchers have described the type of processing that produces Type 1 decisions as the result of intuitive, heuristic processing [10, 16, 31, 32] or associative processing [64]. In contrast, Type 2 decisions are slow, sequential, and effortful, requiring analytic processing that consumes working memory resources [17, 18, 32]. Padilla and colleagues [56] suggest that visual saliency in visualizations supports Type 1 decision-making by drawing viewers' attention to the salient features and by producing visual–spatial biases. The authors define visual–spatial biases as heuristics that are elicited by the visual encoding techniques used in visualizations, and they note that these biases can be beneficial or detrimental, depending on the circumstances. Visual–spatial biases are more complex than visual saliency alone. However, we propose that if these biases support Type 1 processing, they are processed automatically and could be incorporated into visual salience models for use in evaluating the effectiveness of visualizations.

12.3 Expanding the Effectiveness of Saliency Models as a Visualization Evaluation Tool

In addition to developing saliency models that are predictive of bottom-up attention and visual–spatial biases, it may be possible to identify factors that consistently affect top-down attention for viewers who are using visualizations to make decisions. Text typically has low visual saliency from the perspective of traditional saliency models, even though it is very important from a top-down perspective. In our own prior research [45], we demonstrated that incorporating text as a feature in a saliency model dramatically improved the model's performance when predicting where viewers looked in data visualizations. This example illustrates that it is feasible to develop saliency models that incorporate both bottom-up and top-down features, particularly when there are top-down features that draw viewers' attention in consistent ways. For example, if certain types of visual encodings elicit consistent visual–spatial biases, it should be possible to incorporate those biases into saliency models as well. Doing so would produce more effective tools for evaluating visualizations and assessing their impact on decision-making. Since visualizations are typically "born digital" and use common conventions for conveying specific types of information, it is feasible to use their visual features to predict viewers' patterns of attention in ways that may not be possible for other types of visual stimuli, such as natural scenes [22].

12.4 Visual–Spatial Biases with Scatterplots

In this chapter, we focus on scatterplots, which are widely used in science and engineering for analyzing or assessing data sets with two or more dimensions. In the data analysis domain, which is our particular area of interest, scatterplots are often used to help viewers understand correlations, clusters, or outliers [4, 6, 48]. In addition, scatterplots are commonly used to project multidimensional data into a two-dimensional space so that users can better understand the relationships between the variables in the data set. There are many different methods for doing these types of projections and a robust literature focused on evaluating their effectiveness [1, 4, 48, 62, 63, 66]. Much of this research has explored low-level perceptual features that influence participants' ability to distinguish clusters of data, while a small set of studies have characterized common tasks that users perform with dimensionally reduced data and what types of patterns users consider to be important (cf. [4, 6, 29]). However, relatively little work exists that bridges the task level and the perceptual level. In other words, few studies have investigated how a user's task and the perceptual properties of a scatterplot interact to influence the user's patterns of attention.

Prior studies investigating how people perceive scatterplots have indicated that viewers can rapidly extract information from scatterplots, such as trends and correlations [14, 37, 47, 60] or information about the relative properties of different classes of data points [20, 39, 40]. Participants tend to be highly consistent with one another when making judgments about scatterplots (cf. [12, 20, 60]). However, this consistency may be due, at least in part, to pervasive visual–spatial biases. For example, people consistently underestimate the correlation between two variables when interpreting scatterplots [60]. As another example, [20] indicated that participants in their pilot tests used a simple heuristic to decide which class in a multi-class scatterplot had the highest mean value: they simply chose the class that had the highest point. While the researchers did not report investigating this effect in detail, it could indicate that people were using simple heuristics and Type 1 decision-making when interpreting these scatterplots.

Regardless of whether participants were asked to evaluate correlations or clusters, the visual properties of the stimuli can lead to different decisions. In the case of scatterplots depicting trends, manipulating the slope and variance of the data changed participants' estimates of correlation, even for the same value of Pearson's correlation [14, 37]. Similarly, in the case of multi-class scatterplots, Etemadpour, Olk, and Linsen [15] found that factors such as the number of points in a cluster, and the cluster's size, shape, and density all influenced the participants' decisions and patterns of eye movements. In one of their experimental tasks, participants were asked to judge which of two clusters was closest to a reference point. The density and number of points in each cluster were varied. Participants spent more time looking at the sparser cluster and were also significantly more likely to say that it was closer to the reference point than the denser cluster. This pattern held regardless of cluster size. These patterns of results provide further evidence that the visual encoding techniques used in scatterplots can give rise to visual–spatial biases that impact viewers' decision-making.

12.5 Experiment: Interpretation of Clusters in a Scatterplot

In light of this prior work, and of our interest in testing the effects of visual–spatial biases and top-down factors on participants' patterns of attention and decisions regarding visualizations, we designed a set of research experiments investigating interpretation of scatterplots. In the current study, we focused on the interplay between top-down attention and the visual features of clusters in scatterplots (see also Chapter 11 in this book [43] for additional studies in this line of research). We combined a subset of the tasks used by Etemadpour and colleagues [15] and by Gleicher and colleagues [20]. Participants were given two clusters of data points with an intermediate reference point. For half of the stimuli, participants were asked to assign the reference point to one of the two clusters. For the other half, participants were asked to judge which cluster had the higher vertical mean. We manipulated the density and dispersion of the clusters, as well as their relative height and the method used to center the reference point between the two clusters (centering based on mean or standard deviation). We hypothesized that the two different tasks would lead to two different patterns of eye movements in response to the same stimuli. In addition, we hypothesized that the visual properties of the clusters would affect participants' decisions in both tasks. We predicted that varying the density, dispersion, and centering of the clusters would induce the use of visual–spatial heuristics that would affect the participants' assessments of the relative height of the clusters and their relative distance from the reference point.

Note that participants viewed identical sets of stimuli, so there were no differences in perceptual visual saliency that could lead to different patterns of attention across tasks. Thus, any differences in eye movements observed across tasks could only be due to top-down attention and/or visual–spatial biases induced by the visual encoding of the stimuli.

12.6 Experiment: Method

In the current study, participants viewed scatterplots showing two clusters of points and a central reference point. For half of the stimuli, participants were asked to assign the reference point to one of the two clusters. For the other half, participants were asked to judge which of the clusters had a higher vertical mean value. The assignment of stimuli to each task was counterbalanced across participants.

12.6.1 Participants

Thirty participants (7 males; mean age = 29.57, stdev = 13.79) were recruited from students, faculty, and staff in the University of Illinois community and compensated $20 for their time. All participants were tested for color vision deficiencies (24 plate Ishihara Test [27]) and near vision acuity prior to completing the study. This experiment was part of a larger project, and these same participants also completed the study reported in Chapter 11 of this book [43].

12.6.2 Design

The task (reference point assignment or cluster height comparison), reference point centering (standard-deviation-centered or mean-centered), relative cluster height (equal or unequal), cluster density (low or high), and cluster dispersion (low or high) were manipulated within subjects.

12.6.3 Materials

All stimuli were created in R Software [59] from simulated data using the ggplot2 software package [72]. The scatterplots had design characteristics similar to those used in [15]. Each scatterplot consisted of colored circles outlined in black on a white background. The data points formed two clusters, one green and one blue (with random assignment of colors in terms of whether the left cluster was green or blue), with an intermediate reference point that was colored red. No titles, axis labels, or tick marks were provided.

Clusters were manipulated along the dimensions of density (low or high) and dispersion (low or high). Clusters with high density contained more data points per square unit of area than those with low density. Clusters with high dispersion were more spread out (i.e., higher standard deviation) than those with low dispersion. Crossing these two dimensions leads to four types of clusters: Type A: high density and low dispersion (40 data points with a standard deviation of 10 units), Type B: low density and low dispersion (15 data points with a standard deviation of 10 units), Type C: high density and high dispersion (85 data points with a standard deviation of 25 units), and Type D: low density and high dispersion (40 data points with a standard deviation of 25 units). The clusters were created by drawing simulated data from Gaussian distributions with the parameters indicated by each cluster type.

Clusters were paired in all possible combinations (e.g., A-A, C-D, D-C, etc.) to create a total of 80 stimuli. In half of the stimuli, the mean cluster height was the same for each cluster in the pair; in the other half, one cluster was higher than the other. The placement of the red reference point between the two clusters was also varied. For half of the stimuli, the reference point was mean-centered; for the other half, it was standard-deviation-centered. When the reference point was mean-centered, it was exactly halfway between the horizontal and vertical means of the two clusters. When the reference point was standard-deviation-centered, it was exactly four standard deviations along the horizontal axis away from the mean of each cluster (and mean-centered along the vertical axis for clusters with means at the same height or one vertical standard deviation above or below the means of the clusters for clusters at different heights).

12.6.4 Procedure

Participants completed the experiment individually in a dark room, seated at a nominal viewing distance of 0.8 m from a computer monitor (0.932 $\times$ 0.523 m; 1920 $\times$ 1080 pixels). Their eye movements were recorded with a Smart Eye Pro eye tracker. Prior to completing each task, participants underwent the standard Smart Eye camera setup procedure and 9-point calibration. Participants were encouraged to sit still during the tasks and to refrain from leaning forward or backward.

All stimuli were presented in the center of the screen on a white background. Each image was 1000 pixels in height, with a variable width to maintain the aspect ratios of the stimuli. For all tasks, each stimulus was preceded by a fixation cross that was presented in the center of the screen for 1000 milliseconds (ms). There was a 500 ms interstimulus interval between trials. Participants responded by pressing a key on the keyboard to indicate which cluster was higher or which cluster should contain the reference point (depending on the task condition).

The task was divided into three sections: a practice session and two blocks of stimuli. During the practice session, participants worked through two example stimuli and were given the opportunity to ask the experimenter for further clarification. In the first block of experimental stimuli, half of the participants judged the relative height of the two clusters and half of the participants assigned the reference point to one of the two clusters. The participants switched tasks for the second block of stimuli. Each block contained 40 of the 80 scatterplots, with the cluster pairings, reference point centering, and relative cluster height counterbalanced across the two sets.

12.7 Experiment: Results

All statistical tests reported here were held at an $\alpha = 0.05$ level (95% confidence interval, CI). Exact binomial tests analyzed whether the clusters chosen differed significantly from what one would expect based on chance (50%). Except where noted, all analyses were run on stimuli with the same mean cluster height. In all of the stimuli, the reference point was centered perfectly between the two clusters, based on either the mean or the standard deviation of the clusters. Thus, any systematic patterns in the participants' responses were due to visual–spatial biases rather than the mathematical properties of the clusters. If the participants were not biased by the visual encodings of the clusters, we would expect them to select the right and left clusters equally often.

In both tasks, participants showed a slight bias toward choosing the cluster on the right side of the screen (53.6%, CI [50.7%, 56.4%], $p = 0.014$ for the cluster membership task; 53.8%, CI [49.8%, 57.9%], $p = 0.066$ for the cluster height task). However, all conditions were perfectly counterbalanced across the left–right dimension, so this bias does not systematically change the interpretation of the results. All further analyses are collapsed across the left–right dimension.

12.8 Cluster Membership Task: Behavioral Results

12.8.1 Density

Selecting trials in which one cluster had low density and one had high density, we analyzed whether relative density of the clusters influenced participants' decisions in the reference point membership task. Overall, participants consistently indicated that the reference point belonged to the cluster with high density (more data points per square unit). The cluster with high density was chosen 78.8% of the time (CI [73.0%, 83.7%], $p < 0.001$). This pattern held regardless of whether the clusters had low dispersion (high density chosen 81.7%, CI [73.6%, 88.1%], $p < 0.001$) or high dispersion (high density chosen 75.8%, CI [67.2%, 83.2%], $p < 0.001$). It also held regardless of whether the reference point was mean-centered (high density chosen 78.3%, CI [69.9%, 85.3%], $p < 0.001$) or standard-deviation-centered (high density chosen 81.7%, CI [70.8%, 86.0%], $p < 0.001$).

12.8.2 Dispersion

We also examined the influence of low versus high dispersion (how spread out the points were) on participants' preference for reference point cluster membership. When collapsing across density (low vs. high) and centering (mean vs. standard deviation), no significant effects were found (high-dispersion cluster chosen 52.1%, CI [45.6%, 58.6%], $p = 0.561$). However, that null result appears to have been driven by reference point centering technique leading to opposite effects. When the reference point was mean-centered, participants were more likely to indicate the reference point belonged to the cluster with a high dispersion (high-dispersion cluster chosen 91.7%, CI [85.2%, 95.9%], $p < 0.001$). This pattern held regardless of whether the clusters had high density (high-dispersion cluster chosen 83.3%, CI [71.5%, 91.7%], $p < 0.001$) or low density (high-dispersion cluster chosen 100.0%, CI [94.0%, 100.0%], $p < 0.001$). When the reference point was standard-deviation-centered, participants were more likely to indicate the reference point belonged to the cluster with a low dispersion (low-dispersion cluster chosen 87.5%, CI [80.2%, 92.8%], $p < 0.001$). Once again, this pattern held regardless of whether the clusters had high density (low-dispersion cluster chosen 91.7%, CI [81.6%, 97.2%], $p < 0.001$) or low density (low-dispersion cluster chosen 83.3%, CI [71.5%, 91.7%], $p < 0.001$).

Figure 12.1 shows examples of same-height pairings with different reference point centering methods. When the standard deviation of the two types of clusters are different (e.g., Type A and Type C), the reference point centering technique has a profound influence on perceptual grouping.

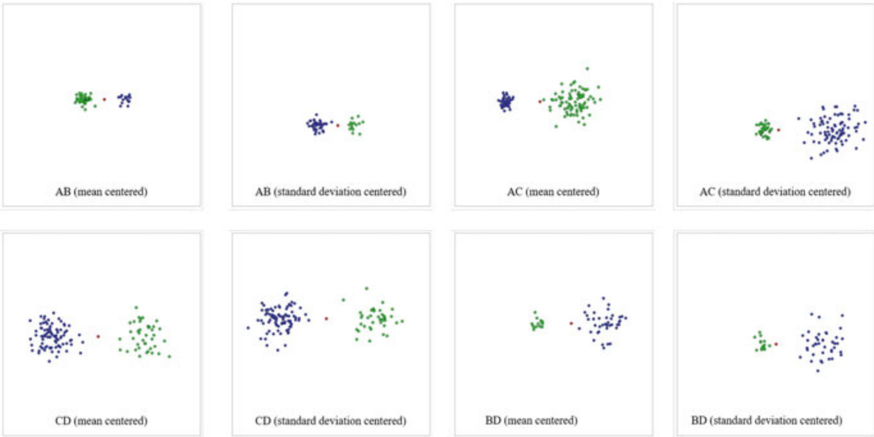

Fig. 12.1 Examples of same-height clusters. Note that Type A and Type B clusters have the same standard deviation, as do Type C and Type D clusters. In these cases, the centering method used to place the reference point does not appear to have a large influence. However, when clusters with different standard deviations are paired, the centering method appears to have a substantial impact on the categorization of the reference point

12.8.3 Nearest Neighbor

The consistency of the participants' responses indicates that their decisions were influenced by visual–spatial biases. Since participants were making rapid, Type 1 decisions in this experiment, they may have employed simple heuristics based on Gestalt principles to make those decisions [15]. In the case of the reference point task, a simple heuristic would be assigning the reference point to the cluster that has the point closest to the reference, which takes advantage of the Gestalt grouping principle of proximity [71]. The perceptual impact of proximity is particularly apparent when the two cluster types have different standard deviations, as shown in Fig. 12.3. Again, recall that the reference point was always perfectly centered based on the mathematical characteristics of the clusters. Yet when clusters of different types are combined, the perceptual impact can be quite striking, with the reference point appearing to be much closer to one cluster than the other.

To investigate whether participants might have used a simple nearest neighbor heuristic to make their decisions, we calculated the distance between the nearest neighbor in each cluster and the reference point. For 91.3% of the stimuli, the cluster with the nearest neighbor was chosen more frequently than the other cluster (CI [82.8%, 96.4%], $p < 0.001$).

12.9 Cluster Height Task: Behavioral Results

For half of the trials, participants were tasked with indicating which of the clusters had the highest average value. In half the stimuli, one of the clusters had a higher (y-axis) mean value; in the other half of the stimuli, the vertical mean value of the clusters was the same. When one of the clusters was higher than the other, the participants chose that cluster 88% of the time. Of more interest were the trials where the clusters had the same mean height—did participants show a consistent bias even when the clusters had identical mean heights? The remainder of this section focuses on the results for the same-height clusters.

12.9.1 Density

When a cluster with low density and a cluster with high density were paired, participants tended to indicate that the cluster with high density had the higher vertical mean (79.2%, CI [70.8%, 86.0%], $p < 0.001$). This pattern was consistent across centering and dispersion manipulations. It held regardless of whether the clusters had low dispersion (high-density cluster chosen 78.3%, CI [65.8%, 87.9%], $p < 0.001$) or high dispersion (high-density cluster chosen 80.0%, CI [67.7%, 89.2%], $p < 0.001$) and whether the reference point was mean-centered (high-density cluster chosen 81.7%, CI [69.6%, 90.5%], $p < 0.001$) or standard-deviation-centered (high-density cluster chosen 76.7%, CI [64.0%, 86.6%], $p < 0.001$).

12.9.2 Dispersion

When a low-dispersion cluster and a high-dispersion cluster were paired, participants tended to choose the cluster with high dispersion as having a higher vertical mean (70.8%, CI [61.8%, 78.8%], $p < 0.001$). It held regardless of whether the clusters had high density (high-dispersion cluster chosen 66.7%, CI [53.3%, 78.3%], $p = 0.013$) or low density (high-dispersion cluster chosen 75.0%, CI [62.1%, 85.3%], $p < 0.001$) and whether the reference point was mean-centered (high-dispersion cluster chosen 66.7%, CI [53.3%, 78.3%], $p = 0.013$) or standard-deviation-centered (high-dispersion cluster chosen 75.0%, CI [62.1%, 85.3%], $p < 0.001$). In this task, the reference point was irrelevant, so the centering method did not influence participants' judgments.

12.9.3 Highest Point

Similar to the nearest neighbor analysis above, the dispersion and density manipulation also influence which cluster tends to have the highest overall point. Participants might simply have chosen the cluster with the highest overall point when deciding which cluster had the highest mean. We found that participants chose the cluster with the highest point 85.0% of the time (CI [70.2%, 94.3%], $p < 0.001$).

12.9.4 Eye Movement Results

Fixations were calculated using Smart Eye's default algorithm, where any sample for which the velocity over the preceding 200 ms was less than 15°/s was deemed a fixation. The first fixation in each trial was excluded from the analysis, as was any fixation with a duration less than 100 ms.

We first examined overall differences in the number of fixations between the two tasks (reference point cluster membership and cluster height, for all stimuli). A mixed effects model (fit with the lme4 package in R software [2]) with a fixed effect for task and random intercepts for participant and stimulus (using Satterthwaite approximation for degrees of freedom; see [42]) revealed that overall, participants had slightly more fixations on average in the cluster height task (mean = 4.75 fixations, stdev = 3.64) relative to the reference point task (mean = 4.59 fixations, stdev = 3.44; $t(1900) = 2.28$, $p = 0.023$).

We also examined the proportion of fixations to each of three region-of-interest (ROI) categories (cluster, reference point, and other—i.e., the background region), as shown in Fig. 12.2. A mixed effects model predicting proportion of fixations from the fixed effects of task (highest cluster vs. reference point membership) and type of ROI along with random intercepts for subject and stimuli (using Satterthwaite approximation for degrees of freedom) revealed significant simple effects of task for each of the ROIs. Relative to the highest cluster task, participants in the reference point membership task tended to have a higher proportion of fixations to both the reference point ROI ($t(5910) = 6.53$, $p < 0.001$) and the "other" ROIs ($t(5910) = 10.18$, $p < 0.001$); in contrast, those in the highest cluster task tended to have a higher proportion of fixations to the cluster ROIs than those in the reference point membership task ($t(5910) = 16.71$, $p < 0.001$).

In addition to examining the proportion of fixations in each ROI, we also investigated how attention to the ROIs unfolded over time under the different task conditions. Figure 12.3 shows the probability of visiting each ROI over the time course of the trial for the reference point and cluster height tasks. For both tasks, participants consistently looked at the right cluster first and then were similarly likely to look at the left or right cluster later in the trial. Participants were more likely to look at the reference point ROI throughout the trial in the reference point task relative to the cluster height task.

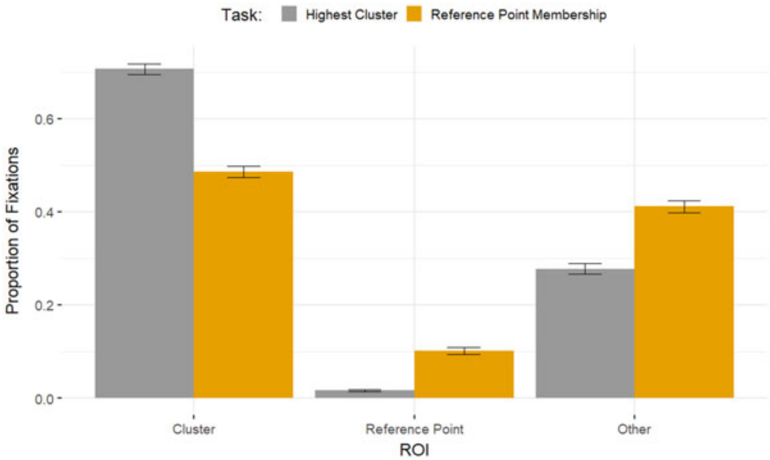

Fig. 12.2 The average proportion of fixations to each region of interest (ROI) based on task. Error bars represent within-subjects standard error of the mean

12.10 Experiment: Discussion

The research presented in this chapter demonstrated that different tasks produced different patterns of eye movements in terms of the number of fixations, the proportion of fixations to each ROI, and the time course of fixations to each ROI. This provides additional evidence that, like natural scenes, data visualizations elicit different patterns of overt attention under different task conditions. In addition, we observed that participants were highly consistent in their judgments of which of two clusters was higher and of which cluster should contain an intermediate reference point. Even when the clusters had the same mean height or a perfectly centered reference point, participants were biased to choose one cluster over the other far more often than we would expect by chance. These results show that visual–spatial biases can have a profound impact on decisions about clusters in scatterplots. These biases may support Type 1 processing and could be incorporated into data visualization saliency tools to evaluate the effectiveness of scatterplot visualizations.

In this experiment, the perceptual features that biased participants toward choosing one cluster over another emerged from the mathematical properties of the clusters. We manipulated the density and dispersion of the clusters as well as the centering method used for the reference point, but we did not manipulate their perceptual properties directly. For example, we did not control which cluster had the highest overall point, like [20], and we did not deliberately position the reference point relative to the clusters to take advantage of Gestalt laws, as did in [15]. Instead, we changed the mathematical properties of the stimuli but allowed the points in the clusters to fall as they may, generating more naturalistic stimuli. We found that simply manipulating the mathematical properties of cluster density, dispersion

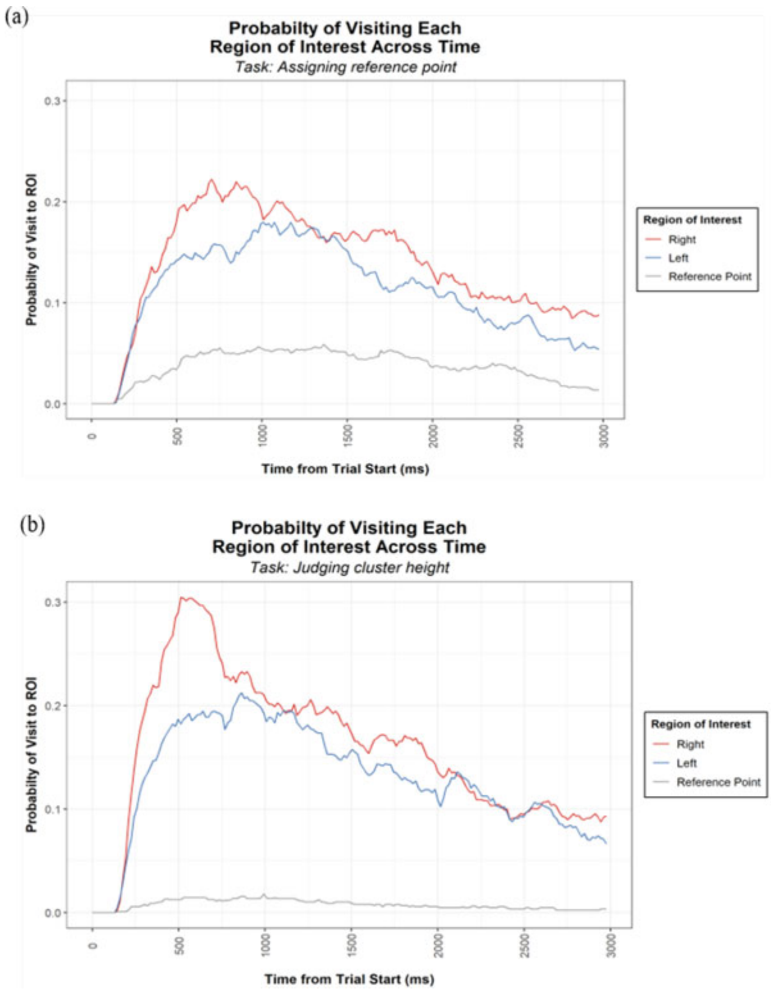

Fig. 12.3 The probability of visiting each ROI over the time course of the trials in the (**a**) reference point and (**b**) cluster height tasks

and reference point centering were sufficient for inducing highly consistent visual–spatial biases in our participants.

Participants tended to show a bias toward choosing the cluster on the right-hand side of the screen, and their patterns of eye movements indicate that they consistently looked at the right-hand cluster first, regardless of task condition. It appears that they were using the right-hand cluster as the anchor point while interpreting the visualization [49, 50]. Similarly, cluster dispersion was a driving factor in the participants' decisions across both tasks. In the cluster height task, clusters with high dispersion were judged to be higher than those with low dispersion.

Of secondary importance was density: clusters with high density were judged to be higher than clusters with low density. In the reference point membership task, the effect of cluster dispersion was modulated by the method used for centering the reference point. When the reference point was mean-centered, participants chose the cluster with high dispersion (higher standard deviation). When the reference point was standard-deviation-centered, participants chose the cluster with low dispersion (lower standard deviation). Cluster density also had a consistent effect on their decisions, although it fell to secondary importance after dispersion. Participants generally chose the cluster with high density (more data per square unit) as opposed to low density. These results contrast those of Etemadpour and colleagues [15], who found that participants tended to group the reference point with the lower density cluster. The current study reveals the critical importance of both cluster dispersion and reference point centering technique on reference point categorization. While prior studies focused on additional elements (e.g., cluster shape), which may have also influenced the results, they did not account for the effect of underlying properties of the data such as dispersion and reference point centering technique. Together, these paint a picture of the multi-faceted factors that can influence interpretation of how a particular data point relates to clusters in the full data set.

Since we manipulated the mathematical properties of the clusters rather than directly manipulating their perceptual properties, it is not clear whether participants based their judgments on the overall characteristics of the clusters (density and dispersion), on simpler heuristics such as which cluster had the highest point or nearest neighbor, or on some combination of the two. In our stimuli, the nearest neighbor and highest point metrics were tightly linked to the density and dispersion manipulations. This is similar to what is likely to be found in real data sets (i.e., higher dispersion clusters are likely to also have the highest points), but it does not allow us to truly tease apart the cause of the participants' visual–spatial biases in the current study. This would be an interesting question to explore in future research, following up on both the present study and Gleicher and colleagues' [20] related observation that participants in their pilot testing tended to use the highest overall point as a heuristic for identifying the class with the highest mean in their intermixed multi-class scatterplots.

Regardless of which of these factors (simple heuristics or underlying mathematical properties of the clusters) was specifically driving the participants' decisions, the fact that their decisions were so consistent has important implications for visualization designers and development of future data visualization saliency tools. Data analysts are not computers—just because there are mathematical equivalencies in a data set does not mean that the viewer who uses the data visualization to make decisions will perceive the same equivalencies. This could be particularly problematic for viewers who are using scatterplots to explore the relationships between variables in a multidimensional space. Our experiment demonstrates that there are perceptual biases that may impact viewers' understanding of these relationships, possibly leading to error or systematic biases. These systematic biases can be accounted for –and potentially predicted– in data visualization saliency tools, helping designers create more effective data visualizations.

Acknowledgments We would like to thank Angela Combs and Andy Wilson (at Sandia National Laboratories) and James Crowell, Camille Goudeseune, Hank Kaczmarski, and Alejandro Lleras (at the University of Illinois at Urbana-Champaign during this study) for their support of this work.

This work was supported by the Laboratory Directed Research and Development (LDRD) Program at Sandia National Laboratories. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This chapter describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

1. G. Albuquerque, M. Eisemann, and M. Magnor. Perception-based visual quality measures. In *Proc. IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 13–20. IEEE, 2011.
2. D. Bates, M. Maechler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.
3. D. Berg and L. Itti. Memory, eye position, and computed saliency. *Journal of Vision*, 8(6):1164, 2008.
4. E. Bertini, A. Tatu, and D. Keim. Quality metrics in high-dimensional data visualization: An overview and systematization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2203–2212, 2011.
5. E. Birmingham, W. F. Bichof, and A. Kingstone. Saliency does not account for fixations to eyes within social scenes. *Vision Research*, 49(24):2992–3000, 2009.
6. M. Brehmer, M. Sedlmair, S. Ingram, and T. Munzner. Visualizing dimensionally-reduced data: Interviews with analysts and a characterization of task sequences. In *Proc. 5th Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pages 1–8. ACM, 2014.
7. S. Card, J. D. Mackinlay, and B. Schneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco, 1999.
8. S. Carpendale. Evaluating information visualizations. In *Information visualization*, pages 19–45. Springer, Berlin, 2008.
9. M. Castelhano, M. Mack, and J. Henderson. Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):1–15, 2009.
10. S. Chaiken, A. Liberman, and A. Eagly. Heuristic and systematic processing within and beyond the persuasion context. In J. S. Veleman and J. A. Bargh, editors, *Unintended Thought*, page 212–252. Guilford, New York, 1989.
11. M. M. Chun. Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4(5):170–178, 2000.
12. W. S. Cleveland, P. Diaconis, and R. McGill. Variables on scatterplots look more highly correlated when the scales are increased. *Science*, 216(4550):1138–1141, 1982.
13. D. A. Cronin, E. H. Hall, J. E. Goold, T. R. Hayes, and J. M. Henderson. Eye movements in real-world scene photographs: General characteristics and effects of viewing task. *Frontiers in Psychology*, 10:2915, 2020.
14. M. E. Doherty, R. B. Anderson, A. M. Angott, and D. S. Klopfer. The perception of scatterplots. *Perception & Psychophysics*, 69(7):1261–1272, 2007.
15. R. Etemadpour, B. Olk, and L. Linsen. Eye-tracking investigation during visual analysis of projected multidimensional data with 2D scatterplots. In *Proc. International Conference on Information Visualization Theory and Applications (IVAPP)*, pages 233–246, 2014.

16. J. S. B. Evans. Heuristic and analytic processes in reasoning. *British Journal of Psychology*, 75(4):451–468, 1984.
17. J. S. B. Evans. In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7(10):454–459, 2003.
18. J. S. B. Evans and K. E. Stanovich. Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3):223–241, 2013.
19. T. Foulsham and G. Underwood. How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception*, 36:1123–1138, 2007.
20. M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri. Perception of average value in multiclass scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 12(19):2316–2325, 2013.
21. T. M. Green, W. Ribarsky, and B. Fisher. Building and applying a human cognition model for visual analytics. *Information Visualization*, 8:1–13, 2009.
22. M. J. Haass, A. T. Wilson, L. E. Matzen, and K. M. Divis. Modeling human comprehension of data visualizations. In *Virtual, Augmented and Mixed Reality: 8th International Conference Proceedings*, volume Lecture Notes in Computer Science 9740, pages 125–134. Springer, 2016.
23. C. Healey and J. Enns. Attention and visual memory in visualization and computer graphics. *IEEE Transactions on Visualization and Computer Graphics*, 18(7):1170–1188, 2011.
24. M. Hegarty, M. S. Canham, and S. I. Fabrikant. Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1):37–53, 2010.
25. J. M. Henderson. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7:498–504, 2003.
26. V. Interrante. Harnessing natural textures for multivariate visualization. *IEEE Computer Graphics and Applications*, 20(6):6–11, 2000.
27. S. Ishihara. *Tests for colour-blindness: 24 plates edition*. Kanehara Shuppan Co., Ltd, Tokyo, 1972.
28. L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2:194–203, 2001.
29. S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):993–1000, 2009.
30. H. Jänicke and M. Chen. A salience-based quality metric for visualization. *Computer Graphics Forum*, 29(3):1183–1192, 2010.
31. D. Kahneman. A perspective on judgment and choice: mapping bounded rationality. *American Psychologist*, 58(9):697–720, 2003.
32. D. Kahneman and S. Frederick. Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases: The Psychology of Intuitive Judgment*, 49(81), 2002.
33. Y. Kim and A. Varshney. Saliency-guided enhancement for volume visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):925–932, 2006.
34. H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.
35. M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision Research*, 41:3559–3565, 2001.
36. M. F. Land, N. Mennie, and J. Rusted. Eye movements and the roles of vision in activities of daily living: Making a cup of tea. *Perception*, 28:1311–1328, 1999.
37. D. M. Lane, C. A. Anderson, and K. L. Kellam. Judging the relatedness of variables: The psychophysics of covariation detection. *Journal of Experimental Psychology: Human Perception and Performance*, 11(5):640–649, 1985.
38. M. Lansdale, G. Underwood, and C. Davies. Something overlooked? How experts in change detection use visual saliency. *Applied Cognitive Psychology*, 24(2):213–225, 2009.

39. J. Li, J. B. Martens, and J. J. van Wijk. Judging correlation from scatterplots and parallel coordinate plots. *Information Visualization*, 9(1):13–30, 2010.
40. J. Li, J. J. van Wijk, and J. B. Martens. A model of symbol lightness discrimination in sparse scatterplots. In *Proc. IEEE Pacific Visualization Symposium (PacificVis)*, pages 105–112, 2010.
41. A. Lleras, R. A. Rensink, and J. T. Enns. Consequences of display changes during interrupted visual search: Rapid resumption is target specific. *Perception and Psychophysics*, 69(6):980–993, 2007.
42. S. G. Luke. Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4):1494–1502, 2017.
43. L. E. Matzen, K. M. Divis, M. J. Haass, and D. Cronin. Task matters when scanning data visualizations. In *Visualization Psychology*, chapter 11. Springer, 2022.
44. L. E. Matzen, M. J. Haass, K. M. Divis, and M. C. Stites. Patterns of attention: How data visualizations are read. In *Proc. International Conference on Augmented Cognition*, pages 176–191, 2017.
45. L. E. Matzen, M. J. Haass, K. M. Divis, Z. Wang, and A. T. Wilson. Data visualization saliency model: A tool for evaluating abstract data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):563–573, 2018.
46. L. E. Matzen, M. J. Haass, J. Tran, and L. A. McNamara. Using eye tracking metrics and visual saliency maps to assess image utility. *Electronic Imaging*, pages 1–8, 2016.
47. J. Meyer and D. Shinar. Estimating correlations from scatterplots. *Human Factors*, 34(3):335–349, 1992.
48. L. Micallef, G. Palmas, A. Oulasvirta, and T. Weinkauf. Towards perceptual optimization of the visual design of scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1588–1599, 2017.
49. A. L. Michal and S. L. Franconeri. Visual routines are associated with specific graph interpretations. *Cognitive Research: Principles and Implications*, 2(1):20, 2017.
50. A. L. Michal, D. Uttal, P. Shah, and S. L. Franconeri. Visual routines for extracting magnitude relations. *Psychonomic Bulletin & Review*, 23(6):1802–1809, 2016.
51. M. Mills, A. Hollingworth, S. Van der Stigchel, L. Hoffman, and M. D. Dodd. Examining the influence of task set on eye movements and fixations. *Journal of Vision*, 11(8):17, 2011.
52. M. Milosavljevic, V. Navalpakkam, C. Koch, and R. A. Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology*, 22(1):67–74, 2012.
53. T. Munzner. *Visualization Analysis and Design*. CRC Press, Boca Raton, FL, 2014.
54. C. Nothelfer, M. Gleicher, and S. Franconeri. Redundant encoding strengthens segmentation and grouping in visual displays of data. *Journal of Experimental Psychology: Human Perception and Performance*, 43(9):1667–1676, 2017.
55. L. Padilla. A case for cognitive models in visualization research. In *Proc. 7th Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 143–151, 2018.
56. L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, 3(1):29, 2018.
57. L. M. Padilla, I. T. Ruginski, and S. H. Creem-Regehr. Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive Research: Principles and Implications*, 2(1):40, 2017.
58. C. Plaisant. *The challenge of information visualization evaluation*. ACM, 2004.
59. R Development Core Team. R: A language and environment for statistical computing [Computer software manual]. http://r-project.org (ISBN 3-900051-07-0), 2008.
60. R. A. Rensink and G. Baldridge. The perception of correlation in scatterplots. *Computer Graphics Forum*, 29(3):1203–1210, 2010.
61. R. Rosenholtz, A. Dorai, and R. Freeman. Do predictions of visual perception aid design? *ACM Transactions on Applied Perception (TAP)*, 8(2):12, 2011.
62. M. Sedlmair, A. Tatu, T. Munzner, and M. Tory. A taxonomy of visual cluster separation factors. *Computer Graphics Forum*, 31(3pt4):1335–1344, 2012.

63. M. Sips, B. Neubert, J. P. Lewis, and P. Hanrahan. Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum*, 28(3):831–838, 2009.
64. S. A. Sloman. The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1):3–22, 1996.
65. J. Stasko. Value-driven evaluation of visualizations. In *Proc. 5th Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, pages 46–53. ACM, 2014.
66. A. Tatu, P. Bak, E. Bertini, D. Keim, and J. Schneidewind. Visual quality metrics and human perception: an initial study on 2D projections of large multidimensional data. In *Proc. International Conference on Advanced Visual Interfaces*, pages 49–56, 2010.
67. E. E. Veas, E. Mendeaz, S. K. Feiner, and D. Schmalstieg. Directing attention and influencing memory with visual saliency modulation. In *Proc. SIGCHI Conference on Human Factors in Computer Systems*, pages 1471–1480, 2011.
68. R. Veras and C. Collins. Saliency deficit and motion outlier detection in animated scatterplots. In *Proc. CHI Conference on Human Factors in Computing Systems*, page 541, 2019.
69. E. Wall, M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko. A heuristic approach to value-driven evaluation of visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):491–500, 2018.
70. C. Ware. *Information visualization: perception for design*. Elsevier, 2012.
71. M. Wertheimer. A brief introduction to gestalt, identifying key theories and principles. *Psychologische Forschung*, 4(1):301–350, 1923.
72. H. Wickham. *ggplot2: Elegant graphics for data analysis*. Springer, New York, 2009.
73. J. M. Wolfe. Guided Search 2.0: A revised model of visual search. *Psychonomic* Bulletin and Review, 1(2):202–238, 1994.
74. J. M. Wolfe, K. R. Cave, and S. L. Franzel. Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance*, 15(3):419–433, 1989.
75. A. L. Yarbus. *Eye Movements and Vision*. Plenum Press, New York, 1967.
76. J. S. Yi, Y. A. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: how do people gain insights using information visualization? In *Proc. Workshop on BEyond time and errors: novel evaLuation methods for Information Visualization*, page 4. ACM, 2008.

Chapter 13
Leveraging Conscientiousness-Based Preferences in Information Visualization Design

Tomás Alves, Bárbara Ramalho, Daniel Gonçalves, Joana Henriques-Calado, and Sandra Gama

Abstract Recent research on information visualization has shown how individual differences such as personality mediate how users interact with visualization systems. Although there is a robust body of research on this relationship, these studies focus on a particular subset of personality constructs. Therefore, there are still personality traits with untapped potential that can provide new findings and inform the design of user-centered visualization systems. This chapter focuses on the conscientiousness personality trait, which measures a person's preference for an organized approach to life over a spontaneous one. In particular, we believe that conscientiousness may regulate how one prefers graphical encodings and organization. We leverage design guidelines based on user preferences and conscientiousness levels to prototype different information visualization systems. We conducted a user testing phase to understand how these prototypes affect user task efficiency, task efficacy, perceived ease of use, perceived usefulness, and preference. Our findings show that conscientiousness levels lead to distinct user preferences, suggesting an interaction effect between conscientiousness and design guidelines in task efficiency. Additionally, individuals with low conscientiousness scores appear to be faster at completing tasks independently of the design guidelines. Moreover, individuals with high and low conscientiousness scores prefer a visualization specifically designed based on their preferences. Finally, the design guidelines lead to different perceived ease-of-use scores. Our study sheds new light on the relevance of personality as an adaptation technique in the design pipeline of visualization systems.

T. Alves (✉) · B. Ramalho · D. Gonçalves · S. Gama
HUMAN-LAB, Graphics and Interaction, INESC-ID and Instituto Superior Técnico—University of Lisbon, Lisboa, Portugal
e-mail: tomas.alves@tecnico.ulisboa.pt; barbara.ramalho@tecnico.ulisboa.pt; daniel.j.goncalves@tecnico.ulisboa.pt; sandra.gama@tecnico.ulisboa.pt

J. Henriques-Calado
CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Lisboa, Portugal

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_13

13.1 Introduction

Recent research in the field of information visualization has leveraged individual differences as an adaptation metric to tackle the limitations of one-size-fits-all approaches [15, 36, 49]. Among the several psychological constructs that differentiate individuals, personality traits and cognitive abilities have shown promising results [41]. Nevertheless, there is a disparity in the number of studies that focus on each of these traits. For instance, we found that the state-of-the-art studies more consistently the Locus of Control (LoC) than the Five-Factor Model (FFM) traits of agreeableness and conscientiousness [41]. Alves et al. [4] conducted a preliminary study to bridge this gap. The authors suggest that personality is a differentiating factor regarding designing information visualization systems and, in particular, results showed promise on the influence of conscientiousness on user preferences for information visualization techniques. In particular, Alves et al. [4] used two approaches to identify personality-based preferences: one based on correlations and the other on clustering. Conscientiousness showed more effects while interacting with the other personality variables in the cluster-based approach than by analysis with correlations. Further, people with high scores tend to prefer line charts with points compared to the remaining population, while individuals in the extremes of the distribution prefer a sunburst in comparison to a tree map. In light of these findings, the present study continues this line of research by studying whether design preferences tailored by conscientiousness scores affect the final evaluation of information visualization systems regarding user preference and performance.

Regarding user preferences, researchers usually developed their experimental apparatus based on what they expect to be relevant for a particular subject rather than focusing on any input from the participants. Consequently, those apparatus may contain biased artifacts from the perceptions designers have regarding user preferences from different personality profiles. Additionally, we found no case in which researchers leverage user preferences as a basis for information systems development and, in particular, how those preferences and personality factors modulate user performance and experience dimensions. This follow-up study continues to investigate conscientiousness since this trait measures the preference for an organized approach to life as opposed to a spontaneous one [28]. Weighing this predisposition, we believe that conscientiousness can, for instance, bias the user preference for graphical layouts, even when the hierarchical and quantitative structures are the same.

Based on these findings, we conducted an in-depth follow-up user study to *understand whether designing visualizations according to the preferences of individuals with different* ***conscientiousness*** *personality profiles is relevant to information visualization*. In particular, we collected design preferences for information visualization techniques and found that conscientiousness affects user preferences. We continued by creating conscientiousness-based design guidelines and developing dashboard prototypes. To validate these guidelines, participants interacted with the different prototypes to understand whether conscientiousness-based design guidelines affect user task efficiency, task efficacy, and perceived ease of use, usefulness, and preference. Results suggest an interaction between conscientiousness and the prototypes' design in task efficiency. Additionally, individuals with low conscientiousness are usually faster, independently of the design guidelines. Both individuals with lower and higher conscientiousness scores prefer a visualization created based on their preferences. Finally, the design guidelines lead to different perceived ease-of-use scores. These findings provide implications for user modeling and adaptive information visualization systems specifically tailored to user preferences. For instance, researchers should not only consider conscientiousness when they define a personality profile to tailor visualizations to user preferences.

This chapter is as follows: In Sect. 13.2, we present the fundamentals of personality psychology before we tackle a selection of studies that have addressed the influence of different personality variables on information visualization (Sect. 13.3). Section 13.5 follows with a description of how we collected personality and preference data and the creation of the design guidelines. Next, Sect. 13.6 covers the validation of these guidelines, including a discussion on the results and limitations of our study. Finally, we conclude our work with a presentation on future directions.

13.2 Fundamentals of Personality Psychology

Personality psychology is a branch of psychology that examines personality and its variation among individuals [46]. Two of the classic definitions of personality belong to Allport [3] and Child [13]. Allport [3] considers personality as a unique psychological system inside individuals. A few years later, Child [13] deems personality as an internal factor that gives consistency over time to the individual's behavior. However, both authors agree that personality is an integrated part of individuals. Recent research in visualization has consistently found that individual differences such as personality can predict goal-setting behaviors as well as how individuals interpret and use information through visualization [41, 50, 68, 70]. This richness prompted us to leverage personality constructs to study their relevance in information visualization design. In particular, we expect to enhance the user profile with personality data and understand how to improve the design of visualizations tailored to specific personality profiles. However, we need a model to define different personality types according to specific metrics to classify and compare people based on their personalities without difficulty.

There is a broad range of theories and models, each with differing perspectives on particular topics when defining personality constructs [16]. Among them, several models have been developed based on different personality theories such as the FFM [17, 44], the HEXACO [37], the Eysenck's Model [10, 21], the Learning Style Inventory (LSI) [32, 33], the Myers-Briggs Type Indicator (MBTI) [47, 48], and the LoC [54, 55]. The basis of these models is personality traits[1] to describe individuals' behaviors and characteristics. In particular, McAdams [43] views traits as underlying biologically determined dispositions. Genetic disparities powerfully drive individual differences, and maturational trends follow a biologically mediated program.

Among the first trait-based personality models, Eysenck [21] created the PEN model, where we measure personality through three dimensions: psychoticism, extraversion, and neuroticism. By the 1980s, however, many researchers began to agree that five broad, roughly independent dimensions best-summarized personality variation. These five dimensions led to the creation of the FFM [44]. This model is a hierarchical organization of personality traits in five dimensions: neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. Each of the dimensions divides into six subdimensions called facets. It is usually referred to as the OCEAN model—the acronym of the five presented dimensions—or simply by Big Five. The HEXACO model [37] shares four traits with the FFM and introduced two new personality traits. In particular, it is composed of honesty–humility (H), emotionality (E), extraversion (X), agreeableness (A), conscientiousness (C), and openness to experience (O).

Although all of the mentioned models consider a trait a continuous scale, some models address personality traits as a dichotomy. The LSI has a set of four personality profiles for practical learning styles—conversing, accommodating, diverging, and assimilating—based on four dichotomies—abstract/concrete, conceptualization/experience, active/reflective, and experimentation/observation. Based on the Jungian theory [31], the MBTI uses a similar approach by assessing personality through four dichotomies: extraversion/introversion, sensing/intuition, thinking/feeling, and judging/perceiving. In contrast, it provides 16 different personality profiles, depicting a higher range than the LSI. Finally, the LoC proposed at first by Rotter [55] treated this personality factor as a dichotomy between internal and external. Nevertheless, soon after Rotter published his paper on the LoC construct, the author decided to reiterate his work and address several problems, limitations, and misuses concerning his original conceptualization [56]. In the light of this, further work by Levenson [39] started to address the LoC with three independent scales: internal, powerful others, and chance. This new approach to LoC allows researchers to process this trait similar to the trait continuum scales of the other models such as the FFM.

Among these personality models, the FFM stands out since research shows it subsumes most known personality traits, and researchers claim that this model represents the "basic structure" underlying the variation in human behavior and preferences [34]. In addition, the FFM is the most widespread and generally accepted model of personality [25, 57], since it provides a classification and a

[1] Allport [2] first defined personality traits as generalized and personalized determining tendencies, consistent and stable modes of an individual's adjustment to his environment. Furthermore, the author built a vast lexical collection of adjectives that could describe these traits.

conceptual framework that unifies much of the research findings in the psychology of individual differences. Nevertheless, given the complexity of personality as a psychological construct with the different dimensions interacting with one another, we decided to focus on a single trait to provide a broader scope of the impact of personality in human–computer interaction. Specifically, we address the personality dimension of conscientiousness, a core component of well-studied personality models such as the FFM and the HEXACO. This trait suggests self-use of socially prescribed restraints that facilitate goal completion, following norms and rules and prioritizing tasks [28]. In particular, it measures the preference for an organized approach to life than a spontaneous one. This dimension is considered to be the least emotionally charged and correlates with positive and negative emotions [27, 60]. On the one hand, people with high values of this trait are more likely to be well organized, reliable, and consistent. They plan, seek achievements, and pursue long-term goals. They also live less emotional lives overall, are more balanced, more predictable, and will encounter fewer emotionally intense situations (fewer extreme lows and fewer extreme highs). On the other hand, individuals with low conscientiousness are generally more easy going, spontaneous, and creative. They tend to be more tolerant and less bound by rules and plans [51].

13.3 Related Work

As we mentioned, recent research leverages how personality predicts goal-setting behaviors [11] and how individuals interpret information [7] to inform visualization design. Two of the most studied personality traits are the LoC [38] and the FFM [17, 25, 57]. The LoC is the most studied personality trait [41]. This trait shows relevant results between people with internal (*Internals*) or external (*Externals*) LoC in search performance across hierarchical [26], time series [59], and item comparison [12] visualization designs, visualization use [68, 70], and behavioral patterns [50]. In particular, while *Externals* are faster and more accurate than *Internals* regarding inferential tasks such as comparing two items [68, 70], *Internals* are significantly faster than *Externals* when performing procedural tasks (search tasks to locate items) [26]. Additionally, *Internals* are usually faster than *Externals* in image-based search tasks [8]. Regarding behavioral patterns, Ottley et al. [50] found that *Externals* adopt a strategy similar to depth-first search in indented trees. In contrast, *Internals*' strategies looked like a breadth-first search. When the visualization was a dendrogram, *Externals* did not follow a specific strategy, while *Internals* were consistent with a mix of breath- and depth-first. Brown et al. [8] found similar results, thus reinforcing how LoC acts as a mediator of search patterns in visual search tasks.

Although not to the same extent, there is also a body of research for the FFM traits. More specifically, neuroticism, extraversion, and openness to experience show measurable effects in the visualization field. Ziemkiewicz and Kozara [69] found that individuals with high openness to experience scores are faster while solving problems related to hierarchical visualizations that include conflicting visual and verbal metaphors. Furthermore, Ziemkiewicz et al. [70] concluded that, while neurotic individuals attained high accuracy on hierarchical search tasks, introverted participants were more accurate in answering the questions. Oscar et al. [49] also explored neuroticism and extraversion. The authors manipulated the visualization's information granularity to approximate the moment an adaptive system presents a visualization to the user. Both traits showed direct effects on task accuracy and completion time. In particular, people with high neuroticism were less likely to be deceived by spurious correlations. Moreover, extroverted participants were less likely to indicate that there was not enough detail available to answer the task. Indeed, these individuals were also less likely to complete a low need for detail task accurately (e.g., "What was your performance on your exercise goal today?") regardless of whether it was a find or compare values task.

Regarding conscientiousness, to the best of our knowledge, only three works have evaluated the relationship between this trait and information visualization. Both Ziemkiewicz and Kosara [69] and Brown et al. [8] measured the conscientiousness level of the participants. However, neither work reported any measurable effect of conscientiousness under the studied conditions. The third study was conducted by Alves et al. [4], where the authors focused on user preferences for information visualization techniques, similarly to Ziemkiewicz et al. [70] and Lallé and Conati [36]. Alves et al. [4] leveraged two distinct approaches to study this relationship: (i) correlation-based analysis—correlations between user preference and an idiom—and (ii) cluster-based analysis—extract preference patterns from each group composed of individuals with common characteristics. Although their preliminary results suggested that personality affects user preferences with both types of analysis, the authors consider the cluster-based approach more appropriate to this problem. In particular, it helps to reduce multiple comparison issues. In this approach, clusters were defined based on all variables from the FFM and the LoC. In particular, conscientiousness levels were significantly different across clusters, and each of those depicted distinct idiom preferences for evolution over time and hierarchy contexts. Based on these results, we decided to follow the previous work [4] and focus on conscientiousness to study in-depth whether this trait can solely model user preferences, task efficiency, efficacy, and perceived ease of use and usefulness.

13.4 Methodology Overview

In this chapter, we investigate *the relevance of using design preferences based on personality profiles in the design pipeline of information visualization systems*. Past studies in visualization leveraged the effect of personality on user preferences [36, 70]. In contrast, conscientiousness has not been studied yet in this context. We decided to address this gap by conducting an in-depth user study on the conscientiousness personality trait based on its role in regulating a preference for organization [28]. In particular, we believe it is interesting to study whether this trait affects user preference for more structural graphic elements and layout designs of information visualization techniques. Further, we want to investigate *how individuals judge their user experience with an information visualization system designed according to their personality-based preferences*.

There are several steps to take to answer these research questions. First, we need to define which visualization techniques we study. Afterward, we ask users to assess their preferences for said techniques and continue with an analysis based on the conscientiousness scores for shared preferences sets, e.g., whether conscientious individuals share a set of design preferences. At this point, we will know how individuals from different conscientiousness scores prefer to see their visualizations. Finally, we prototype visualizations according to the conscientiousness-based preferences and ask individuals to interact with them. The following sections cover the mentioned study phases in depth.

13.5 Assessment of Personality and Design Preferences

As the first step to understanding ***how conscientiousness affects user preferences regarding information visualization techniques***, we decided to include familiar information visualization contexts, as they allow researchers to understand preconceived structures of information. As such, we address: (i) **hierarchy**, one of the most common in research (e.g., [70]); (ii) **evolution over time**, giving the importance of time series data analysis [59]; and (iii) **comparison**, as it is more appropriate to show differences or similarities between values at a fixed granularity [12]. We then chose a set of representative idioms for each context. Additionally, we include a simple and familiar scenario with each context to stimulate users to reflect on the implications of using each idiom rather than the complexity of the data. We also focused on minimizing the number of channels and marks of each graph and keeping them consistent across contexts while keeping the same data within a context.

Regarding hierarchy (Fig. 13.1), items are all related to each other by the principle of containment. We opted for a tree map (Fig. 13.1a), a circular packing diagram (Fig. 13.1b), a sunburst (Fig. 13.1c), and a Sankey diagram (Fig. 13.1d) to display the distribution of food consumed by a household within a month. For evolution over time context (Fig. 13.2), we chose line charts with and without points (Figs. 13.2a and b) and area charts (Fig. 13.2c). The scenario asked the participant to imagine that the data referred to the number of registrants and participants in a marathon held annually in the United States. Finally, we decided to use radar charts (Fig. 13.3a), word clouds (Fig. 13.3b), horizontal and vertical bar charts (Figs. 13.3c and d), and pie charts (Fig. 13.3e) for the comparison context (Fig. 13.3). In particular, the scenario represents the levels of the happiness index among six different countries (France, Italy, Portugal, Spain, Germany, and the United Kingdom).

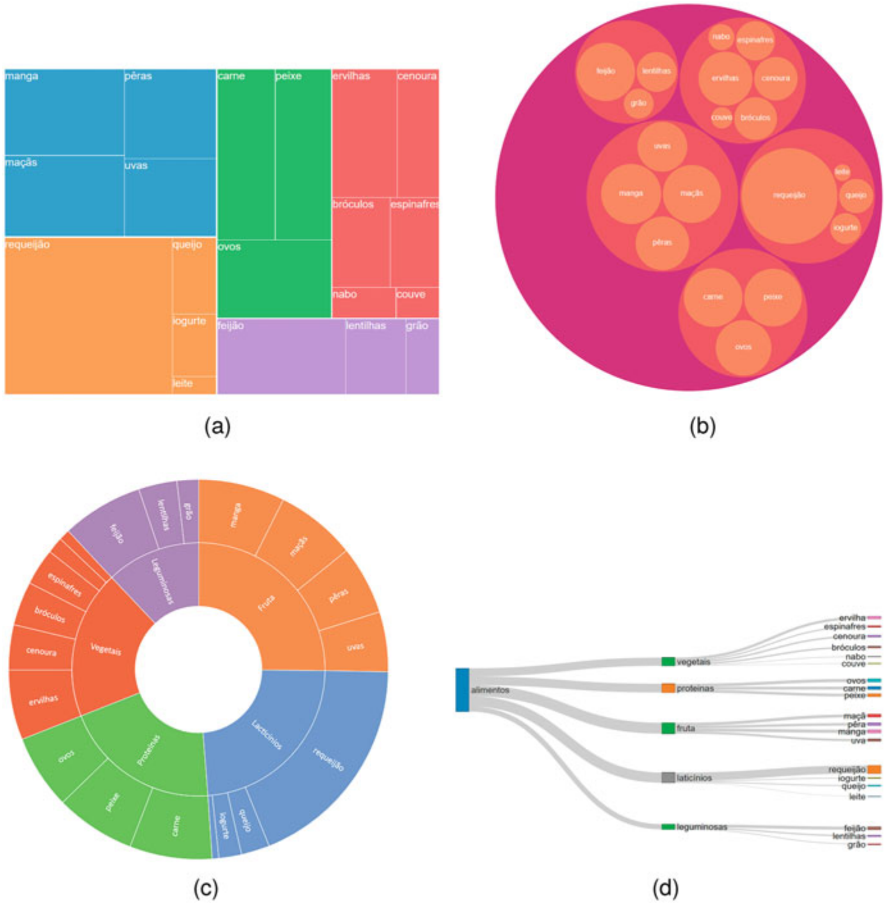

Fig. 13.1 Idioms for the hierarchy context. (**a**) Tree map idiom. (**b**) Circular packing diagram idiom. (**c**) Sunburst idiom. (**d**) Sankey diagram idiom

Besides the idioms, we also want to address the graphical elements and layout encompassing the information visualization techniques. We decided to address both **font style**—the font family used in a dashboard—and **size**. Font style varies between Arial, Calibri, Calibri Light, Times New Roman, and Lucinda Handwriting. Sarsam and Al-Samarraie [58] also studied the relationship between personality and user preferences and found that these fonts produce measurable effects. The font size can be either small (12pt), medium (14pt), or large (16pt). Moreover, we want to understand whether conscientiousness may lead users to prefer different **information button styles**, i.e., how the dashboard represents the help button. We tested the button with only an icon, an icon and text, and only text. Regarding layout, we focused on the **information density** of a dashboard and the **menu bar position**. Information density defines how much information the dashboard should

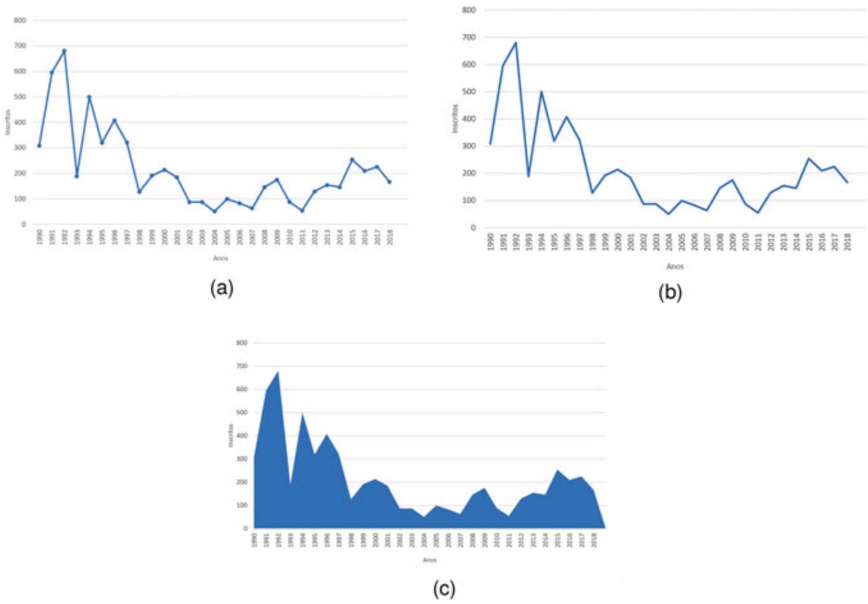

Fig. 13.2 Idioms for the evolution over time context. (**a**) Line chart with points idiom. (**b**) Line chart without points diagram idiom. (**c**) Area chart idiom

represent. In this case, we allow a dashboard to have between two, four, or six sections (Fig. 13.4). Finally, the **menu bar position** can be either at the top, the bottom, the left, or at the right of the screen.

13.5.1 Data Collection

Subjects were recruited through standard convenience sampling procedures such as direct contact and word of mouth. Subjects included any Portuguese interested in participating if at least 18 years old. Our final data set comprises 64 participants (30 males, 34 females) between 18 and 60 years old ($M = 24.27$; $SD = 7.10$). In addition, we asked whether they were using glasses or contact lenses and the apparatus used while filling in the questionnaire. We then verified through one-way ANOVAs that neither of these factors affected how participants responded to each item.

Before the experiment, participants were informed about the experience and provided informed consent. We also informed them that they could quit the experiment at any time. We then collected the conscientiousness value with the Portuguese version of the Revised NEO Personality Inventory (NEO PI-R) [18, 40]. This questionnaire allows researchers to assess the FFM five personality traits and their 30 facets. We calculate the score for each trait by the sum of the Likert

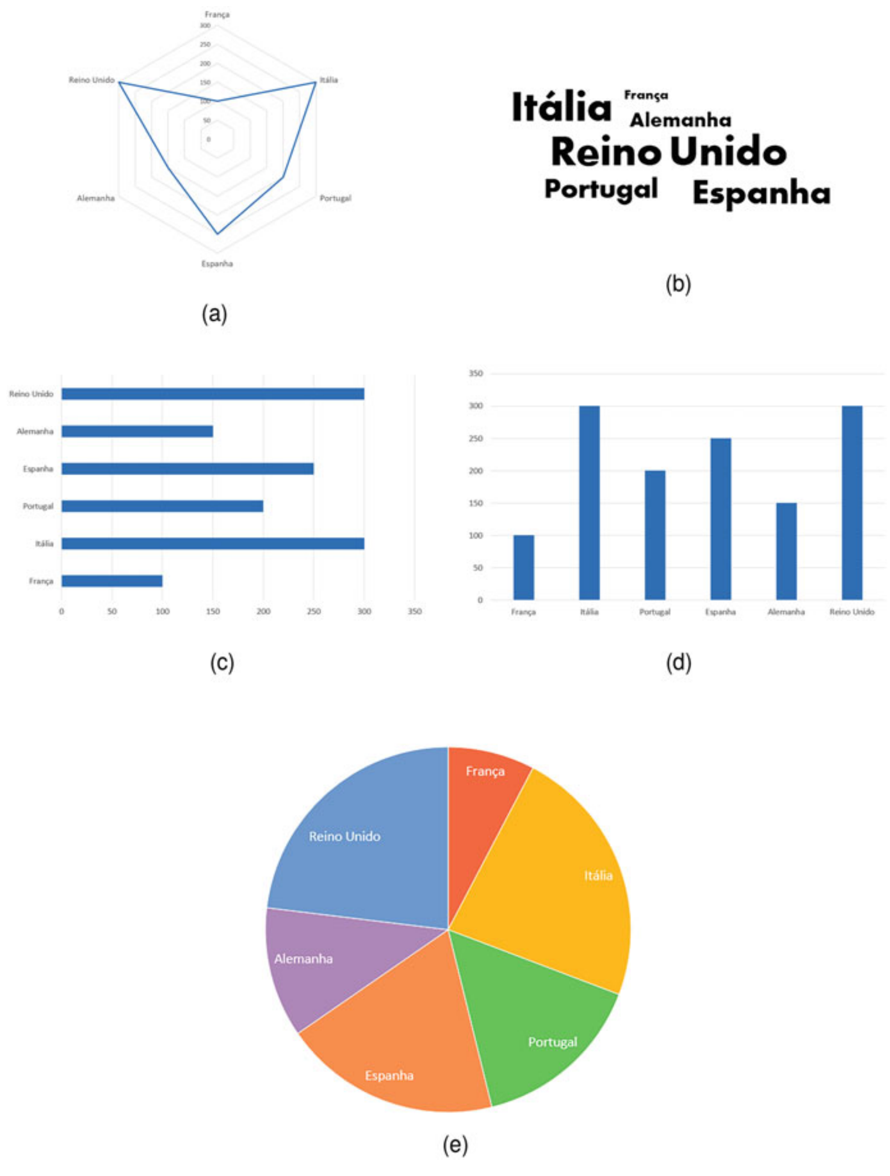

Fig. 13.3 Idioms for the comparison context. (**a**) Radar chart idiom. (**b**) Word cloud idiom. (**c**) Horizontal bar chart idiom. (**d**) Vertical bar chart idiom. (**e**) Pie chart idiom

Scales based on assertions semantically connected to behaviors, e.g., "I have a vivid imagination." In particular, each Likert Scale has five possible alternatives of agreement: *strongly agree*, *agree*, *undecided*, *disagree*, and *strongly disagree*. Overall, the questionnaire has 240 items, including 30 different subscales (one

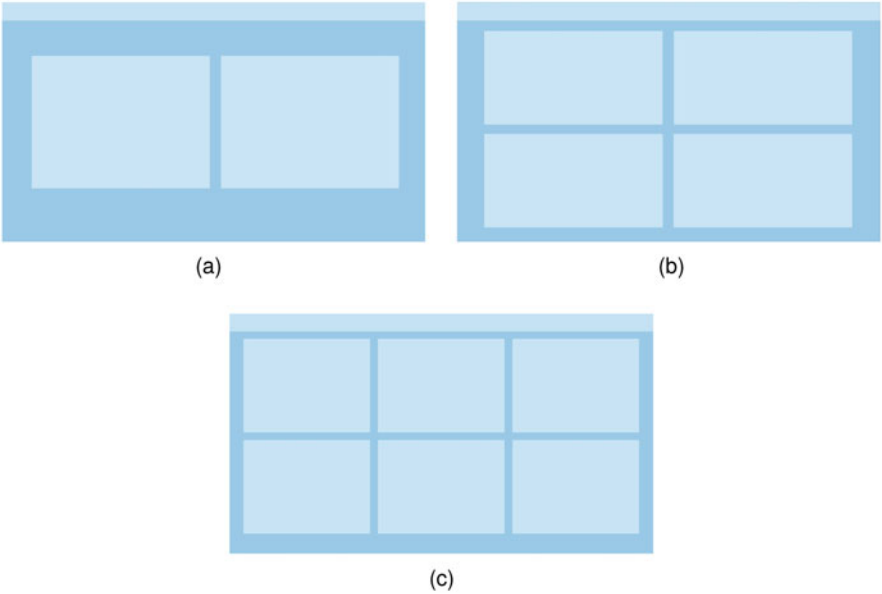

Fig. 13.4 Dashboard examples with different information density. (**a**) Low information density. (**b**) Medium information density. (**c**) High information density

for each facet), with eight items for each subscale. High scores exacerbate the characteristics of the trait and vice versa, i.e., higher scores on the conscientiousness scale mean that individuals have a stronger disposition to be more organized.

Afterward, we presented to the participants an online questionnaire[2] with two parts. The first part addressed the graphical elements and layout features (font family, information density, etc.), and the second part contained a visual example of each idiom (Figs. 13.1, 13.2, and 13.3) grouped by context. We prompt participants to read the description of the feature or the scenario for the respective context and then assess their preference for each style by completing a seven-point Likert scale ranging from *Low Preference* (1) to *High Preference* (7). Each participant saw all possible graphical elements, layout features, and chart types simultaneously. Moreover, we allowed participants to freely change their scores until they were satisfied with all ratings to avoid anchoring bias.

[2] https://web.tecnico.ulisboa.pt/~tomas.alves/publications/Alves2022_Conscientiousness-InfoVis-Preferences.pdf.

13.5.2 *Data Analysis*

Based on the work of Alves et al. [4] and Sarsam and Al-Samarraie [58], we follow a cluster-based approach to study the relationship between conscientiousness and user preferences. In particular, we run a clustering algorithm to group users according to their personality characteristics and find whether participants with similar personality profiles share preferences for specific information visualization techniques.

13.5.2.1 Clustering Personality Variables

Our first step is to understand how many distinguishable groups of conscientiousness scores exist in the sample. We started by applying hierarchical density-based clustering [29, 45] and the elbow criterion [8, 58] to find which was the most appropriate number of clusters to work. We obtained a value of three clusters using the silhouette and Davies–Bouldin index scores analysis [52], as well as Ward's cluster method [20]. Then, we used the k-means clustering algorithm [66] to avoid the noise labels that hierarchical density-based clustering yields. The k-means clustering algorithm divides n observations into k clusters in which each person belongs to the group with the nearest mean. We started by normalizing our data and allowing the algorithm to run 100 iterations with different centroid seeds using Euclidean distance. The final result contained the best output of 100 consecutive runs in terms of inertia. Finally, we need to validate if the clusters have distinct groups of conscientiousness scores. Therefore, we also conducted an ANOVA to validate whether each cluster contained people with statistically significant differences in conscientiousness levels and its facets. We found a significant difference ($p < 0.05$) in between the three clusters regarding each of the personality variables, which shows that all groups have participants who differ among themselves in conscientiousness scores. Additionally, we found that the clusters were significantly different from each other FFM trait (see Fig. 13.5).

Table 13.1 depicts the means and standard deviation values for all personality traits of the FFM and facets of conscientiousness. We can see that the first group ($N = 19$) has participants with the **highest levels of conscientiousness** across groups. It means that Cluster 1, which we will refer to henceforth as *C-High*, includes people who are predictably the most competent, goal- and detail-oriented, and organized. Moreover, these individuals score the highest mean values for the other traits except neuroticism, the lowest mean value. Regarding the second cluster ($N = 27$), it depicts people with **medium values of conscientiousness**, being the *C-Medium* cluster. These people are also the most disagreeable and less open to experiences, as we can observe from the low agreeableness and openness to experience scores. Finally, the third cluster ($N = 18$) includes participants with the **lowest levels of conscientiousness**. We also renamed Cluster 3 to *C-Low*, as

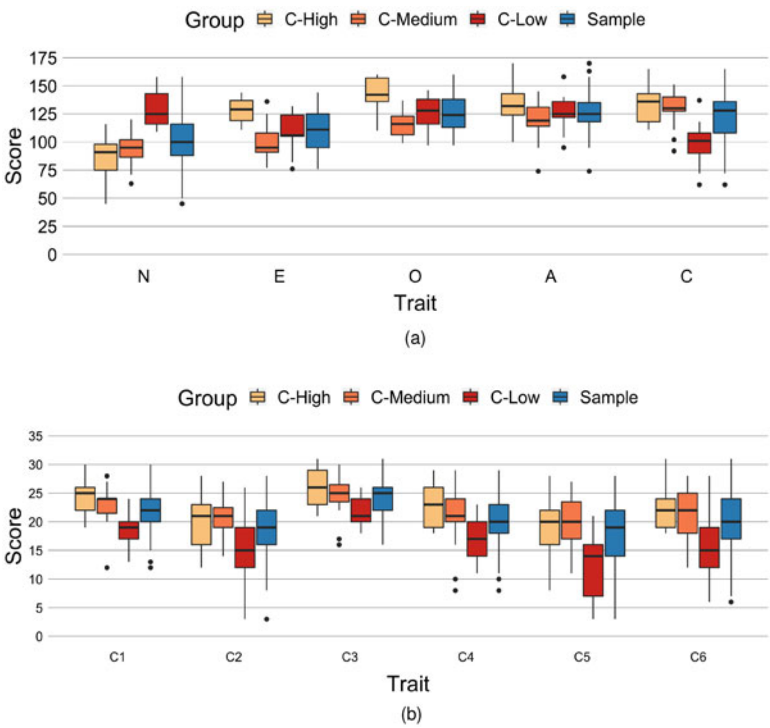

Fig. 13.5 Boxplots of the distribution of traits and conscientiousness' facets between clusters and the sample. (**a**) FFM personality traits. (**b**) Facets of conscientiousness

Table 13.1 Results of the K-means clustering algorithm for each personality trait and conscientiousness' subdimensions

	C-High		C-Medium		C-Low	
Personality variable	M	SD	M	SD	M	SD
Neuroticism[a]	80.79	21.87	90.81	16.68	125.90	16.03
Extraversion[a]	125.74	13.87	101.74	15.43	99.94	21.00
Openness to experience[a]	139.47	14.10	110.52	16.18	128.00	16.37
Agreeableness[a]	134.32	14.33	118.81	16.60	126.39	16.13
Conscientiousness[a]	142.63	18.74	126.00	21.13	98.89	19.20
Competence[a]	24.00	3.37	22.00	3.90	18.50	3.09
Order[a]	22.00	4.98	20.00	5.04	14.50	5.80
Dutifulness[a]	28.00	3.09	25.00	3.92	22.50	2.82
Achievement striving[a]	26.00	4.17	20.00	5.47	16.50	5.54
Self-discipline[a]	22.00	5.62	19.00	4.97	14.00	4.64
Deliberation[a]	23.00	3.86	22.00	4.99	15.00	5.48

[a] This variable is significantly different across clusters at a significance level of 0.05

it includes more impulsive individuals who abide less by the rules and are less perfectionistic. In addition, these individuals are the most introverted and neurotic.

13.5.2.2 Extracting Association Rules

To extract information visualization preferences for the different features and contexts among individuals of those three clusters, we opted for the Apriori algorithm [30], an association rules method to find common patterns. In particular, the Apriori algorithm uses prior knowledge of frequent itemset properties in a data set to create Boolean association rules. Data preprocessing included the creation of an array for each participant containing the styles or idioms that they preferred the most for each feature or context, respectively. In case of a tie between two or more items in their preference ratings, we included all items tied together. Afterward, we divided users by their cluster labels and used the Apriori algorithm in each group. We performed each run with lower bound minimal values of 0.1 for support, 1 for confidence, and 3 for lift. We chose the inputs for us to obtain a good balance between generating a reasonable number of rules that would cover most of our design styles and a robust confidence value. The algorithm yielded 24 distinct rules for C-High, 46 for C-Medium, and 13 for C-Low. An Apriori association rule is represented often as $itemA \rightarrow itemB$, which translates into $itemB$ being frequently present in a set of preferences that also contains $itemA$.

13.5.2.3 Finding Preferences for Clusters

We continued our analysis by choosing which rules to use based on their frequency. We started by choosing the rule with the highest frequency value and then continued by picking rules with a lower frequency that share a design style and do not conflict with a design style previously selected for a feature. In addition, we focused on maximizing the number of design elements that we could derive from the association rules. When a feature did not have a style associated with it at the end of our analysis, we chose the most frequent preferred style for that feature among group participants. Table 13.2 illustrates the final rule sets for each cluster. Based on the final set of rules for each group, we were able to derive which values to apply to the different features and contexts (Table 13.3). Notably, all features and contexts have different styles across the three clusters. Additionally, the C-Medium and C-Low preferences are very similar, differing only on the menu bar positioning and the hierarchical context. Nevertheless, we were not able to derive styles for specific features.

As we mentioned, we address this issue by choosing the most frequent style among cluster participants. Based on the frequencies, we found people in the high cluster commonly preferred the "vertical bar chart" for the comparison context. Additionally, those with low conscientiousness scores commonly chose a "large" font size and put the menu bar on the "top" of the screen. With

Table 13.2 Association rules chosen for each cluster. Each rule is represented by $itemA \rightarrow itemB$, which translates into $itemB$ being frequently present in a set of preferences that also contains $itemA$. Frequency is the number of times the rule was present in the output of the Apriori algorithm. Support refers to items' frequency of occurrence in the data. Confidence indicates the number of times the if-then statements are found true. Lift shows how many times the if-then statement is expected to be found to be true. For instance, the rule highDensity → calibriLight means that participants who prefer high information density also prefer the Calibri Light font and that this pair appears 12% of the time

Rules for the C-High	Frequency	Support	Confidence	Lift
highDensity → calibriLight	8	0.120	1.00	4.67
highDensity → iconText	3	0.115	1.00	5.57
calibriLight → iconText	2	0.115	1.00	6.19
highDensity → sankeyDiagram	2	0.115	1.00	4.77
mediumFont → highDensity	2	0.115	1.00	4.46
barDown → mediumFont	1	0.115	1.00	8.67
calibriLight → sankeyDiagram	1	0.115	1.00	3.71
highDensity → barDown	1	0.115	1.00	4.33
highDensity → mediumFont	1	0.115	1.00	4.33
iconText → barDown	1	0.115	1.00	3.71
linechartPoints → highDensity	1	0.115	1.00	3.71
sankeyDiagram → mediumFont	1	0.115	1.00	5.20
Rules for the C-Medium	Frequency	Support	Confidence	Lift
mediumDensity → largeFont	66	0.200	1.00	5.00
timesNewRoman → mediumDensity	52	0.200	1.00	5.00
linechart → mediumDensity	38	0.200	1.00	5.00
barChartHorizontal → linechart	37	0.200	1.00	5.00
linechart → largeFont	23	0.200	1.00	5.00
barChartHorizontal → mediumDensity	21	0.200	1.00	5.00
linechart → barChartHorizontal	9	0.200	1.00	5.00
timesNewRoman → barLeft	7	0.200	1.00	5.00
linechart → iconOnly	4	0.200	1.00	5.00
mediumDensity → iconOnly	4	0.200	1.00	5.00
timesNewRoman → iconOnly	4	0.200	1.00	5.00
highDensity → barLeft	3	0.200	1.00	5.00
linechart → tree map	3	0.200	1.00	5.00
Rules for the C-Low	Frequency	Support	Confidence	Lift
barChartHorizontal → linechart	90	0.114	1.00	6.21
timesNewRoman → sankeyDiagram	25	0.107	1.00	5.77
linechart → barChartHorizontal	18	0.113	1.00	5.98
barChartHorizontal → mediumDensity	12	0.107	1.00	5.60
linechart → mediumDensity	10	0.107	1.00	6.16
barChartHorizontal → timesNewRoman	8	0.107	1.00	5.95
linechart → iconOnly	1	0.107	1.00	5.60

Table 13.3 Features and styles for each cluster. The percentage represents the amount of times the design style was chosen compared to the other styles for a feature in each cluster. Bold styles were derived from the association rules

Feature	C-High	C-Medium	C-Low
Font family	**Calibri Light** (23.68%)	**Times New Roman** (9.43%)	**Times New Roman** (17.14%)
Font size	**Medium** (36.84%)	**Large** (66.67%)	Large (77.78%)
Info density	**High** (10.52%)	**Medium** (51.85%)	**Medium** (44.44%)
Menu bar	**Bottom** (10.53%)	**Left** (3.7%)	Top (77.78%)
Buttons	**Icon and Text** (31.58%)	**Icon Only** (70.37%)	**Icon Only** (66.67%)
Hierarchy	**Sankey** (26.09%)	**Tree map** (11.11%)	**Sankey** (27.27%)
Evolution	**Line chart w/ points** (72.73%)	**Line chart w/out points** (33.33%)	**Line chart w/out points** (40.00%)
Comparison	Vertical bar chart (37.93%)	**Horizontal bar chart** (39.58%)	**Horizontal bar chart** (33.33%)

the conscientiousness-based preferences defined for each group, we can create information visualization design guidelines for the different features and contexts. In particular, we were able to derive the following guidelines:

- People **high** on conscientiousness prefer visualizations with medium Calibri Light font, high information density, the menu bar at the bottom of the screen, and buttons with icons and text. Their preferred idiom to represent hierarchical information is a Sankey diagram, for evolution over time is a line chart with points and for comparisons a vertical bar chart.
- People with **medium** conscientiousness levels prefer visualizations with large Times New Roman font, medium information density, the menu bar at the left of the screen, and buttons with icons. Their preferred idiom to represent hierarchical information is a tree map, for evolution over time is a line chart without points and for comparisons is a horizontal bar chart.
- People with **low** conscientiousness prefer visualizations with large Times New Roman font, medium information density, the menu bar at the top of the screen, and buttons with icons. Their preferred idiom to represent hierarchical information is a Sankey diagram, for evolution over time is a line chart without points and for comparisons is a horizontal bar chart.

To critically analyze how impactful the clustering algorithm was, we have to analyze how these preferences were rated independently of the cluster. User preferences for the hierarchy context (Fig. 13.6) show an evenly distributed rating across idioms. Interestingly, while the tree map idiom has the poorest ratings ($M = 3.66$; $SD = 0.236$), it was the most preferred for C-Medium. For the remaining clusters, the Sankey idiom was the most preferred one, which means that those personality profiles consistently rated this idiom higher ($M = 4.67$; $SD = 0.251$), compared to the sunburst diagram ($M = 5.34$; $SD = 0.183$) and the circular packing idiom ($M = 4.14$; $SD = 0.228$).

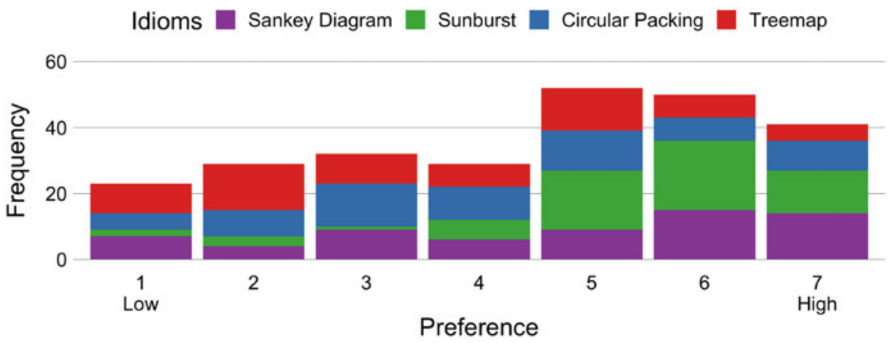

Fig. 13.6 Stacked frequency bar chart for user preferences in the hierarchy context

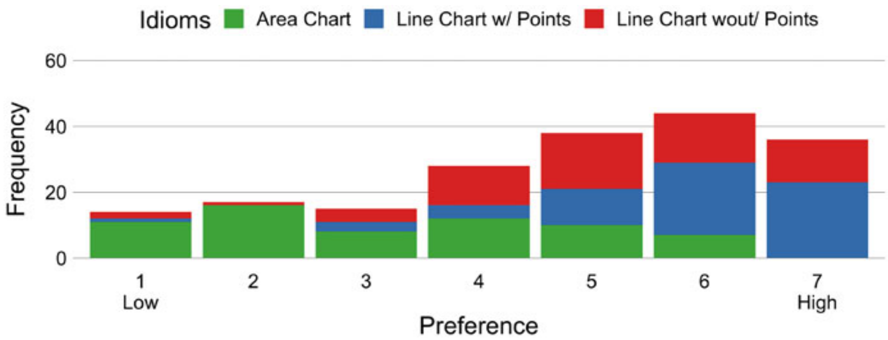

Fig. 13.7 Stacked frequency bar chart for user preferences in the evolution over time context

In order to critically analyze how impactful the clustering algorithm was, we have to analyze how these preferences were rated independently of the cluster. User preferences for the hierarchy context (Fig. 13.6) show an evenly distributed rating across idioms. Interestingly, while the tree map idiom has the poorest ratings ($M = 3.66$; $SD = 0.236$), it was the most preferred for C-Medium. For the remaining clusters, the Sankey idiom was the most preferred one, which means that those personality profiles consistently rated this idiom higher ($M = 4.67$; $SD = 0.251$), compared to the sunburst diagram ($M = 5.34$; $SD = 0.183$) and the circular packing idiom ($M = 4.14$; $SD = 0.228$).

We can also observe that, for the evolution over time context (Fig. 13.7), participants poorly rated the area chart ($M = 3.23$; $SD = 0.206$) in terms of preference compared to line chart types with ($M = 5.84$; $SD = 0.158$) or without points ($M = 5.16$; $SD = 0.183$). The area chart did not appear in the strongest association rules as expected, leaving the differences between clusters for line charts. We found that despite line charts with points being highly rated compared to this idiom without points, both the C-Medium and C-Low clusters preferred the latter.

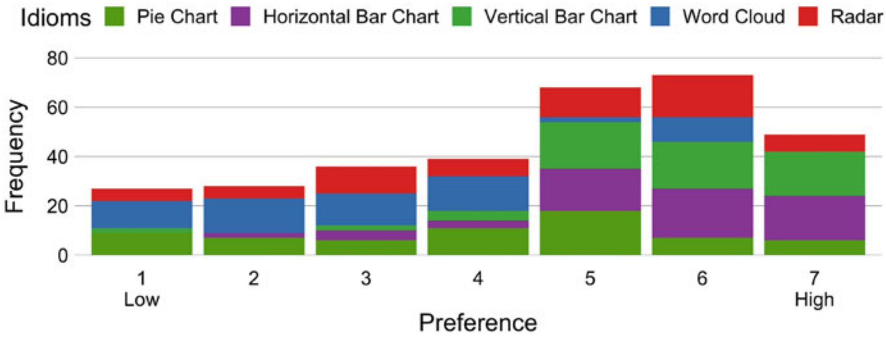

Fig. 13.8 Stacked frequency bar chart for user preferences in the comparison context

Finally, bar charts dominated the comparison context preferences. In fact, looking into how participants rated their preferences (Fig. 13.8), both horizontal ($M = 5.61$; $SD = 0.162$) and vertical ($M = 5.61$; $SD = 0.166$) bar charts covered the best ratings. These preferences are reflected in Table 13.3 with the C-High cluster preferring the vertical version and the other two groups the horizontal one. Regarding the other idioms, both radar ($M = 4.48$; $SD = 0.225$) and pie ($M = 4.05$; $SD = 0.231$) charts had an average rating higher than four, while the word cloud ($M = 3.19$; $SD = 0.204$) had the poorest rates. Only the word cloud, the tree map, and the area chart idioms showed an average rating below the mark of four, showing how poorly preferred these idioms are in the respective contexts. Nevertheless, individuals from C-Medium prefer tree maps compared to the counterparts we studied.

In this light, our results suggest that conscientiousness is a differentiating factor when designing information visualization systems. For instance, we can observe that conscientious individuals prefer line charts with points. In contrast, the remaining individuals do not prefer the chart version with exact points. We believe that this type of effect can provide an understanding of a potentially causal relationship between conscientiousness and user preferences. We believe that the preference for having points and, consequently, the exact position of data stem from the fact that individuals with high conscientiousness are more prone to be organized and thorough. Another example is the information density on the screen. Highly conscientious individuals also showed distinct preferences from the remaining groups. In particular, these individuals tend to prefer a screen with high information density that may derive from their tendency to follow strict goal-oriented strategies and, consequently, favor having the highest amount of information available.

With the design guidelines created, our next step is to conduct a user study to validate them. As we mentioned, our research question is to study whether these design guidelines created from user preferences based on conscientiousness affect user performance and self-assessment metrics such as perceived usefulness, ease of use, and preference. Therefore, the following sections describe our methodology to investigate this research question, ending with a discussion of our findings.

13.6 Evaluation

In this section, we describe our study to understand whether preferences solely based on the conscientiousness trait are indeed relevant to the field of information visualization. In particular, our research question addresses **how individuals judge their user experience with an information visualization system designed according to their personality-based preferences**. We used a mixed design where each participant belongs to one of three possible clusters and they tested all dashboards: 3 clusters $\times$ 3 dashboards.

13.6.1 Visualizations

We developed an information visualization dashboard based on the design guidelines for each group. We named each dashboard based on the group to which they were developed, e.g., we designed V-High for the C-High group. The dashboards (see Fig. 13.9) contain information regarding terrorist attacks in Europe from a public data set. As a means to present a variety of information densities on the dashboards similar to the one that participants assessed (Fig. 13.4), we decided to add one or more idioms to each dashboard so that they had at least a medium density. In this case, the added idiom was a choropleth map, where luminance encodes the number of terrorist attacks per country. In particular, we added a fifth idiom to the V-High dashboard, as this group of people prefers a higher density than the others. This idiom was a bubble chart showing the number of injured and casualties in terrorist attacks over time. We specifically chose these charts since they do not cover any of the target contexts. Moreover, we believe that adding another instance of a chart from the contexts could interfere with the context representation. To have each context only contribute to the dashboard with a single chart, we opted for two charts that are ignored in the tasks that users need to complete and, consequently, minimize their interference with the experience.

13.6.2 Tasks

To prompt user interaction, participants were asked a set of nine questions, three per dashboard. These questions were divided into two types of information-seeking behaviors [1]—factual and interpretive—which led the participant to search for information in the content of the interface and reflect on what they learned. We

Fig. 13.9 Visualizations created based on the conscientiousness-based design guidelines. (**a**) V-High, the dashboard for the C-High cluster. (**b**) V-Medium, the dashboard for the C-Medium cluster. (**c**) V-Low, the dashboard for the C-Low cluster

opted to not consider exploratory tasks and focus on questions that had a single correct measure to be able to track when the task ended. In factual questions, the user seeks a specific piece of data. They were the following:

F1 *What are the three main targets for terrorist attacks in Germany?*
F2 *What are the main targets for terrorist attacks in Ireland?*
F3 *What are the main targets for terrorist attacks in the Netherlands?*
F4 *What is the number of assassinations that happened in Greece between* 1990 *and* 2018?
F5 *What is the most common terrorist attack type that happened in France between* 1990 *and* 2004?
F6 *What is the number of terrorist attacks against infrastructures that happened in Sweden from* 1990 *and* 2004?

Regarding the interpretive tasks, this type of information-seeking task requires users to actively create possible scenarios to interpret information about its amount or quality [35]. In our study, we created the following interpretive questions:

I1 *You are planning your next trip to Europe, and you want to know how has been the evolution of terrorism cases in Italy. Do you consider that terrorism has been increasing or decreasing? In what year has there been more attacks and what was the number of attacks in that year?*
I2 *You are about to go on a business trip to the United Kingdom. Considering the evolution of terrorist attacks in this country, do you believe it to be safe? In what year has there been more attacks and what was the number of attacks in that year?*
I3 *You are writing an article on the evolution of terrorism in Spain. Considering the evolution of the attacks, do you consider them to be increasing or decreasing? In what year was registered the highest number of attacks and what was the number of attacks in that year?*

While the first three factual questions (F1 to F3) address the hierarchy context idiom, the remaining three (F4 to F6) focus on the comparison idiom. The interpretive questions (I1 to I3) focus on the evolution over time context idioms.

13.6.3 Measures

User Performance We measure the **accuracy** and **response time** for each task. Accuracy refers to whether the participant answered correctly or incorrectly to the task, and we evaluate the response time in seconds.

User Assessment Perceived **usefulness** and **ease of use** were assessed with the Technology Acceptance Model 3 (TAM3) [63]. We found no validated Portuguese translation, since most studies (e.g., [23, 61, 64]) solely translate the original scale [63] and adapt to their context by replacing "the system" with their specific product. Finally, we collected user preference for each dashboard with a seven-point Likert scale ranging from *low preference* (1) to *high preference* (7).

Personality Similar to the previous study, we collected the **conscientiousness** scores with the Portuguese version of the NEO PI-R [18, 40].

Demographics We recorded for each participant their gender, age, self-reported visual acuity, and whether they were color-blind. Moreover, we collected visualization literacy by showing each participant an instance of each chart with an exemplary domain. These examples contained a different domain from the one we used for the experiment to reduce learning bias. Then, we asked participants to assess their familiarity with that visual representation on a five-point Likert scale ranging from *not familiar* (1) to *very familiar* (7).

13.6.4 Expected Findings

User Performance Regarding user efficiency, previous studies [8, 26] showed that the personality trait LoC affected the time users took to complete a search task. We believe that conscientiousness may also affect the time users take to perform tasks since it prompts one to follow norms and rules while prioritizing tasks [28]. In particular, we want to study whether users complete tasks faster while using an information visualization system according to their design preferences based on conscientiousness:

H1 *Users complete tasks faster when they interact with a visualization designed according to their preferences by conscientiousness level.*

Another interesting metric to leverage in user efficiency is the number of errors. Research has shown that users make fewer mistakes while interacting with an interface designed based on their preferences [8, 9, 26]. Moreover, Ziemkiewicz et al. [70] showed that neuroticism and extraversion affect user efficacy. In this light, we believe that conscientiousness may show a similar effect based on its impact on how one approaches life in an orderly way [28]. Thus, having users interact with an information visualization designed based on their preferences by conscientiousness level may affect their mistakes:

H2 *Users make fewer mistakes when they interact with a visualization designed according to their preferences by conscientiousness level.*

User Assessment We want to study how conscientiousness models the perception of user experience dimensions. Based on previous research [36], we believe that conscientiousness may regulate perceived usefulness and ease of use. In particular, we expect users will rate an information visualization system higher in both perceived usefulness and ease of use when designed based on their conscientiousness level:

H3 *Users rate a visualization designed according to their preferences by conscientiousness level higher on scores.*

Finally, user interfaces designed according to personality affect user preference [58]. Therefore, we predict that users will prefer a system designed for their conscientiousness level:

H4 *Users prefer a dashboard designed according to their preferences by conscientiousness level.*

13.6.5 Procedure

Again, we recruited subjects through standard convenience sampling procedures such as direct contact and word of mouth. Our final data set comprises 45 participants (25 males, 20 females) between 19 and 60 years old ($M = 24.9$; $SD = 8.1$). 18 testers (40%) did not participate in the first study where we collected design preferences. In this case, we invited participants to complete the NEO PI-R [18] questionnaire on an online platform. Participants were then informed about the experience and provided informed consent. We also informed them that they could quit the experiment anytime without prejudice or consequence.

We conducted the user testing sessions via an online videoconference platform, where the researcher recorded a screen-shared browser and gave remote control to the user. Participants started by filling in the demographic questionnaire. Next, participants were randomly assigned an order through which they would interact with the three dashboards (V-High, V-Medium, and V-Low). As we mentioned, each dashboard was designed based on the guidelines for the respective conscientiousness cluster (see Sect. 13.5.2.3) and contained the same data. While interacting with a dashboard, we asked participants two factual questions (one for each context), followed by an interpretative question. We also randomized the order of the questions across dashboards. Moreover, we asked each question only once in the experiment. This approach guarantees that participants interacted with each target context idiom on the dashboards and minimize learning bias. After answering the questions for a dashboard, we invited participants to fill in the TAM3 to assess their perceived usefulness and ease of use and the Likert scale to measure their preference. Similar to the previous data collection, participants were allowed to freely change their scores until they were satisfied with all ratings to avoid anchoring bias.

13.6.6 Data Analysis

Regarding the independent variables, the group acts as a between-subject variable and has three possible values: C-High, C-Medium, C-Low. The dashboard is a within-subject variable and, as we mentioned, we renamed the dashboards designed for the C-High, C-Medium, and C-Low clusters as V-High, V-Medium, and V-Low, respectively. All evaluation sessions were video recorded, which we analyzed to collect task performance metrics after the evaluation phase. We measured task performance through the sum of the time participants took to complete all tasks and the sum of the number of wrong answers to those tasks. For the self-assessment scales, we calculated the responses to TAM3 through the sum of the answers to the perceived usefulness and ease-of-use items. Finally, we examined the user preference scores for each participant. We assigned the preferred dashboard(s) according to the highest scores, e.g., if two dashboards have the same score and it is the highest score, we count both as preferred.

After fitting the personality data to the clustering method, we obtained 13, 13, and 19 participants for the C-High, C-Medium, and C-Low clusters, respectively. However, we were not able to record the task completion time of two participants. Therefore, we remove them from the analysis due to constraints in the experiment. Both participants belonged to the C-High, which makes the final sample of this cluster 17. We ran a two-way mixed ANOVA with cluster (3 levels) and dashboard (3 levels) as factors. We tested for sphericity (Mauchly's test) and used the Greenhouse–Geisser correction when the assumption was not met. Finally, we examined user preference by counting per participant the dashboard that scored higher between the three Likert scales. We ran a chi-square test of independence for $r \times c$ contingency tables.

13.7 Results

In this section, we present results related to the effect of conscientiousness on task performance and self-assessment metrics for each visualization system. Data are presented as mean $\pm$ standard deviation unless otherwise stated.

13.7.1 Performance Metrics

Task Completion Time We study task performance through two metrics: task completion time and the number of errors participants commit while performing them. When analyzing task time completion, Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(2) = 4.064$, $p = 0.131$. There was no statistically significant interaction between the

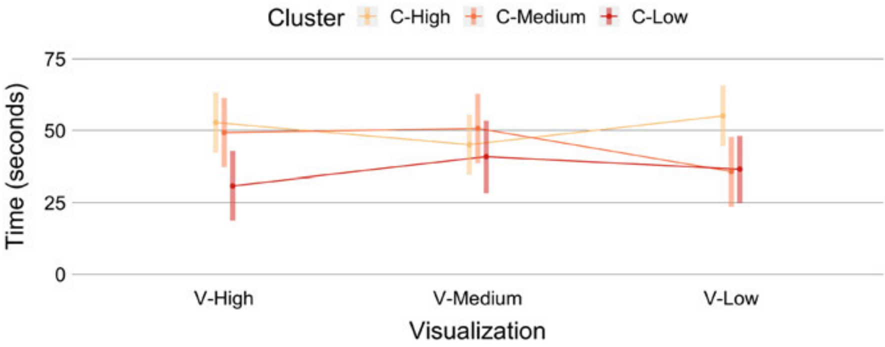

Fig. 13.10 Estimated marginal means graph for the time users took to complete tasks in each dashboard

cluster and the design guidelines on task completion time, $F(4, 80) = 2.234$, $p = 0.073$, partial $\eta^2 = 0.100$. Additionally, we found no significant main effects; the conscientiousness profile did not show significant differences, $F(2, 40) = 1.287$, $p = 0.287$, partial $\eta^2 = 0.06$, as well as the dashboard, $F(2, 84) = 0.280$, $p = 0.756$, partial $\eta^2 = 0.007$. Therefore, we cannot accept **H1**; in addition to conscientiousness not affecting task completion time, participants were not faster while interacting with a dashboard designed for their conscientiousness-based preferences.

By taking a closer look at the distributions, we can observe in Fig. 13.10 that people from the C-Low cluster were usually faster completing the tasks independently of the dashboard (107.744 ± 21.193 seconds; SE) compared to the C-High (152.843 ± 18.532; SE) and C-Medium (135.564 ± 21.193; SE) clusters. They were only slower (109.000 ± 44.488) than people from the C-Medium cluster in their V-Low dashboard (106.920 ± 88.590). Another interesting result is that only people from the C-Medium group took more time (152.077 ± 68.746) in the dashboard designed for them (V-Medium) compared to V-High (147.692 ± 104.687) and V-Low (106.923 ± 88.590). In contrast, individuals from the C-High group took 165.235 ± 113.682 seconds in V-Low compared to the 135.059 ± 86.734 seconds in V-Medium and the 158.235 ± 136.776 seconds in V-High. Finally, the C-Low group took 109.000 ± 44.488 seconds in V-Low compared to the 92.385 ± 26.554 seconds in V-High and the 121.846 ± 44.562 seconds in V-Medium.

We decided to continue our analysis by tackling task completion time per task. As we can observe in Fig. 13.11, it appears that individuals usually take a similar amount of time independently of the visualization. However, we can observe that three tasks suggest an effect by the dashboard. Regarding the factual tasks, we can observe that one dashboard made participants have a higher task completion time. Participants were slower completing F1 in the V-Low dashboard (55.22 ± 56.25 seconds) compared to V-Medium (35.31 ± 19.81) and V-High (29.83 ± 31.31). As we mentioned, F1 was associated with the hierarchy context. However, both V-High and V-Low had a Sankey diagram in that context. In this case, we do not believe that

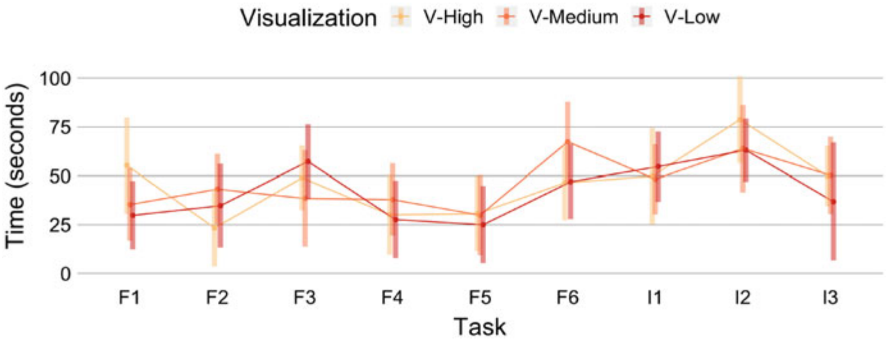

Fig. 13.11 Estimated marginal means graph for the time users took to each task across the dashboards

the dashboard affected task completion time. In F6, individuals were slower in the V-Medium (67.39 $\pm$ 52.74), followed by V-High (47.00 $\pm$ 29.98) and then V-Low (46.40 $\pm$ 41.48). F6 was a factual task dedicated to the comparison context. Again, the slowest task completion time was on a dashboard (V-Medium) that shared the idiom with another dashboard (V-Low). Consequently, we believe that the design guidelines did not affect the time users took to answer the factual tasks.

Concerning the interpretive tasks, a similar trend was present in I2 and I3. In I2, V-Low led participants to spend more time completing the task (78.73 $\pm$ 82.91), while V-Medium (63.91 $\pm$ 21.95) and V-High (63.10 $\pm$ 56.34) showed similar task completion times. Finally, I3 appears to be the only task where a specific dashboard benefited the participants. In particular, V-High showed lower task completion times (36.83 $\pm$ 9.45) compared to V-Medium (50.21 $\pm$ 45.28) and V-Low (49.87 $\pm$ 42.36). In the I2 case, we find a similar trend as we described in the factual tasks. However, I3 appears to be the only case where a specific plot led participants to be faster in completing the task. In particular, the points in the line chart appear to have played a role in an effect on the task completion time of I3. Interestingly, I1 and I2 also leveraged the line chart yet there was no suggestion of the line chart with points playing a major role. We assume that this artifact can be neglected and, consequently, the dashboard did not affect the time users took to answer the questions.

Task Accuracy Regarding the number of errors that participants made, Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(2) = 2.877$, $p = 0.237$. Similar to the previous analysis, we found no statistically significant interaction between the cluster and the design guidelines on the number of errors users committed while performing the tasks, $F(4, 80) = 0.620$, $p = 0.649$, partial $\eta^2 = 0.030$. Moreover, the cluster variable did not show significant differences, $F(2, 40) = 0.316$, $p = 0.731$, partial $\eta^2 = 0.016$, as well as the dashboard, $F(2, 80) = 0.057$, $p = 0.945$, partial $\eta^2 = 0.001$. Again, we cannot accept **H2**, since the number of errors was not affected by either independent

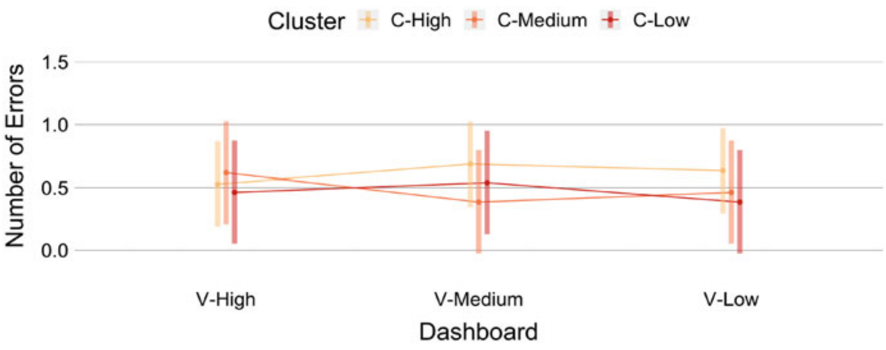

Fig. 13.12 Estimated marginal means graph for the number of errors users committed while interacting with each dashboard

variable. However, we can observe (Fig. 13.12) that individuals from each cluster made less mistakes using the dashboard designed for them. In particular, the C-Low made fewer mistakes when they performed the tasks using their dashboard (0.385 ± 0.650) than with V-High (0.462 ± 0.877) or V-Medium (0.538 ± 0.660). Following the same trend, individuals from C-Medium also made fewer mistakes in V-Medium (0.385 ± 0.650), followed by V-Low (0.462 ± 0.660), and then V-High (0.615 ± 0.768). Finally, C-High participants made fewer errors in the dashboard that was designed based on their preferences (0.471 ± 0.717) than in V-Medium (0.706 ± 0.772) or V-Low (0.647 ± 0.862).

13.7.2 Self-assessment Metrics

We measured perceived usefulness, ease of use, and user preference as self-assessment metrics.

Usefulness For the perceived usefulness, Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(2) = 1.516$, $p = 0.469$. Nevertheless, there was no statistically significant interaction between the cluster and the design guidelines on perceived usefulness, $F(4, 80) = 1.001$, $p = 0.412$, partial $\eta^2 = 0.048$. Additionally, there were no main effects for the independent variables, with the cluster variable not showing significant differences, $F(2, 40) = 0.571$, $p = 0.570$, partial $\eta^2 = 0.028$, as well as the dashboard, $F(2, 80) = 1.975$, $p = 0.145$, partial $\eta^2 = 0.047$. A closer inspection at Fig. 13.13 shows that C-High individuals provide a higher usefulness score to their dashboard (V-High) than to the others (V-High: 24.35 ± 4.429; V-Medium: 22.00 ± 5.196; V-Low: 23.88 ± 3.160). In contrast, C-Medium attributed their lowest usefulness score to V-Medium (23.46 ± 4.352), followed by V-High (24.85 ± 3.158) and then V-Low (25.31 ± 3.772). Regarding C-Low, people from

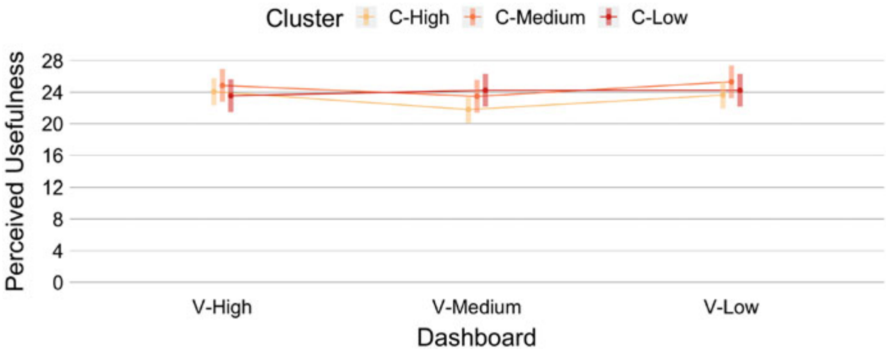

Fig. 13.13 Estimated marginal means graph for the perceived usefulness of each dashboard

this group attributed the highest scores to V-Medium and V-Low, followed by V-High (V-High: 23.54 ± 2.727; V-Medium: 24.23 ± 2.976; V-Low: 24.23 ± 3.219). Moreover, conscientiousness did not seem to have an effect on how participants rated usefulness independently of the dashboard (C-Low: 24.000 ± 0.798; C-Medium: 24.538 ± 0.798; C-High: 23.412 ± 0.698; SE).

Ease of Use Similarly, Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction regarding ease of use, $\chi^2(2) = 0.892$, $p = 0.640$. Although we found no significant interaction effect between the cluster and the design guidelines on perceived ease of use, $F(4, 80) = 0.845$, $p = 0.501$, partial $\eta^2 = 0.041$, there was a statistically significant main effect of the design guidelines, $F(2, 80) = 4.070$, $p = 0.021$, partial $\eta^2 = 0.092$. Additionally, results hint toward a main effect of conscientiousness, $F(2, 40) = 2.266$, $p = 0.117$, partial $\eta^2 = 0.102$, which may suggest that the way individuals perceived ease of use may be regulated by this personality trait. Scores regarding perceived ease of use followed the trends of the perceived usefulness (Fig. 13.14); nevertheless, in this case, results suggest that conscientiousness had a nonsignificant

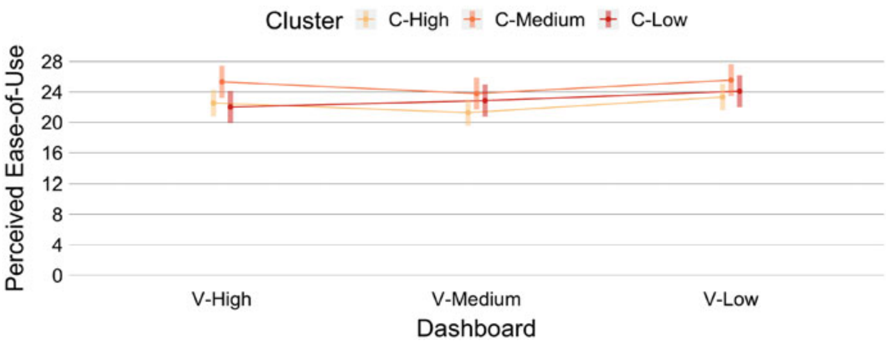

Fig. 13.14 Estimated marginal means graph for the perceived ease of use of each dashboard

Table 13.4 Count of user preferences per dashboards

	C-High	C-Medium	C-Low
V-High	10	3	3
V-Medium	3	4	2
V-Low	6	7	8

effect on how participants rated usefulness independently of the dashboard (C-High: 22.647 ± 0.722; C-Medium: 24.872 ± 0.825; C-Low: 22.974 ± 0.825; SE) with a medium effect. Another interesting result was the significant effect of the design guidelines on the perceived ease of use with a medium-size effect; V-Low shows the highest results (24.421 ± 0.563; SE) compared to V-High (23.397 ± 0.557; SE) and V-Medium (22.676 ± 0.616; SE). In particular, we can observe that the people from the C-High and C-Low attributed similar scores to the V-High (C-High: 22.882 ± 4.241; C-Medium: 25.308 ± 2.428; C-Low: 22.000 ± 3.719) and V-Low (C-High: 23.647 ± 3.639; C-Medium: 25.538 ± 4.075; C-Low: 24.077 ± 3.226). In contrast, the groups differed when they assessed V-Medium (C-High: 21.412 ± 4.459; C-Medium: 23.769 ± 3.586; C-Low: 22.846 ± 3.760). However, we cannot accept **H3**.

User Preference Finally, we investigated whether conscientiousness affected user preference. After counting the preference of each user, we obtained the preference matrix depicted in Table 13.4. A chi-square test of independence was conducted between the dashboard and the cluster. Five expected cell frequencies were lower than five. There was no statistically significant association between the dashboard and the cluster, $\chi^2(4) = 5.406$, $p = 0.248$. The association was moderate [14], Cramer's V $= 0.242$. Since there was not a statistically significant association between the two variables, we cannot reject the null hypothesis and cannot accept the alternative hypothesis. We found **H4** to be inconclusive since we can observe that only the C-High and C-Low clusters preferred the dashboard that was designed for those conscientiousness profiles. Therefore, we cannot accept it.

13.7.3 Discussion

Our objective was to study whether design guidelines from preferences based on the conscientiousness trait are relevant in information visualization. In particular, we extend prior work by Alves et al. [4]. Weighting our findings, we are not able to reach any strong conclusions that highlight that this personality trait is relevant to the design of visualization systems. On the one hand, we found that different conscientiousness scores led individuals to have distinct preferences for information visualization techniques. Additionally, we found that design guidelines formulated based on these preferences lead users to report disparate ease-of-use scores. Based on the statistical tests, results suggest a trend toward conscientiousness affecting how users assess the perceived ease of use of an apparatus. Additionally, when individuals interacted with the dashboard, a nonsignificant trend also shows that the design guidelines affected task completion time. Further, participants made fewer mistakes using the dashboard designed according to their preferences. Finally, people with lower or higher conscientiousness scores preferred the dashboard that was designed according to their preferences (V-Low and V-High, respectively). In addition, individuals with low scores were usually faster than their counterparts in completing the tasks. These findings shed new light on the possibility that **conscientiousness may have a relevant role in information system adaptability based on personality traits**.

On the other hand, we have not accepted our hypotheses, and our findings are in line with the few works that exist in the state-of-the-art [8, 69]. Indeed, we found no significant interactions between the conscientiousness level of an individual and the designed guidelines we followed to prototype the dashboards on task performance and self-assessment metrics. Several possible reasons may explain our findings. The first and most direct one is that design preferences based on conscientiousness levels are not relevant to information visualization systems adaptation, as they do not significantly affect user performance or self-assessment metrics. Another reason might be that the tasks were not sufficiently demanding to trigger a conscientiousness-based response.

Regarding personality, it may also be the case that personality interactions had a more prominent effect than the conscientiousness trait by itself, thus prompting actions and self-assessment by the users that cannot be explained under these study conditions. Finally, the familiarity with the experimental context may have had an effect. Although we did not collect familiarity with the theme of terrorism in Europe or information systems, we did ask users to rate their familiarity with each idiom to understand whether it would have an impact on their interactions. In this case, we measured user familiarity with a seven-point Likert scale. Figure 13.15 shows the distribution of familiarity for the relevant idioms that we used in the factual and interpretative tasks. The most known idioms are the vertical (6.91 $\pm$ 0.358)

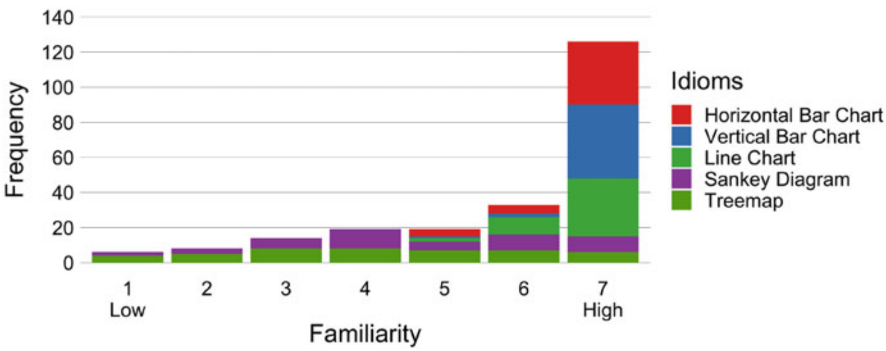

Fig. 13.15 Stacked frequency bar chart depicting the distribution of familiarity for the different idioms

and horizontal (6.71 ± 0.626) bar charts, followed closely by the line chart (6.69 ± 0.557). It means that all idioms from the comparison and evolution over time contexts were well-known to the participants.

For the hierarchy context, both the Sankey and the tree map showed a positive familiarity with ratings of 4.71 ± 1.72 and 4.2 ± 1.85, respectively. Although we believe this is the least likely reason to shape our findings, given how familiar users were with the idioms, we understood from the recordings that the bubble chart from V-High confused some participants, especially those with high conscientiousness. We added this element to increase the information density of the dashboard. Although it may have led conscientious individuals to spend time analyzing the irrelevant bubble chart and, consequently, spending more time performing the tasks, V-High made them perform fewer mistakes. By having a better accuracy, we believe that it led people with high conscientiousness values to rate V-High with the highest perceived usefulness scores. In contrast, the lack of participation of the bubble chart in the tasks may have also led these individuals to not rate V-High with the highest ease-of-use scores. We were also able to understand from the recordings that users commented how the menu bar on the top of the dashboard felt "more familiar" than on the bottom or left side. Again, this factor may have affected how users rated their perceived ease of use and usefulness and user preference. Finally, we also consider that the idiom pool we offered users in the phase of the study was sufficiently complete with the most used idioms, thus allowing flexibility to rate preferences.

Interestingly, the results of both studies appear to be in conflict. Our studies have shown that, although users with different conscientiousness scores have distinct design preferences for visualization features, designing visualization systems based on those design preferences does not have a significant effect on user performance or experience. However, we believe researchers can leverage our design preference findings with other approaches besides the clustering and apriori algorithms. Indeed, all previous studies showed that when users needed to assess their preference for design elements individually, there are differences in user preferences following distinct personality profiles. We hypothesize that the lack of significant effects on user experience and performance of the visualization systems designed according to the personality-based user preferences may be based on the context and the act of interacting with a system rather than ranking design preferences in a form. As such, future studies should aim at refining user preferences by including user interaction in the collection phase. By exposing participants to interaction, users may winnow and adjust their design preferences based on the context and the tasks they have to perform. These refined user preferences may then be examined to understand whether they are triggered by personality dynamics. In this case, we suggest that designers should aim at performing changes that have multiple graphical elements varying in style rather than performing small changes to provide more consistency to the users according to their personality profile.

13.7.3.1 Research Implications

Based on our results and observations, we were able to devise a set of implications that can be useful for information visualization designers targeting personality-based adaptive systems. Designers should take into account that personality—more specifically, conscientiousness—does affect user preference regarding information visualization techniques. Nevertheless, while we found that conscientiousness modeled information visualization preferences, we were not able to find strong results regarding their impact on performance metrics or self-assessment metrics. Therefore, our findings suggest that design preferences based on conscientiousness scores are not a strong metric to leverage in information systems adaptation regarding those factors. Instead, designers should take into account how conscientiousness appears to task performance, for instance, independently of user preferences. Prior work suggests that conscientiousness may affect user performance since it regulates the propensity to facilitate goal completion and prioritize tasks [28]. We expected that high scores in conscientiousness would make individuals more efficient and accurate and vice versa. However, we found contrasting results hinting that individuals with lower conscientiousness scores completed the tasks faster, and it did not affect the number of errors they committed. As we have seen, individuals with lower conscientiousness are usually faster and more correct in completing search tasks than their counterparts.

Another interesting aspect to focus on is the assessment of the visualization trustworthiness [67]. Dinesen et al. [19] found that highly conscientious individuals tend to be careful and seek to retain control over a situation. We believe that it may lead to these individuals trusting less in a visualization. In addition, the assessment of information visualization systems may be affected as these individuals are more sensitive to no-functionality artifacts, thus prompting that every visualization element should have a function and a reason to exist. Finally, we also found that users are generally acquainted with most of the idioms that we studied, which provides designers with a larger design space to take into account in their systems.

Regarding our research, the previous studies showed that taking into account personality factors in advance of the design of graphical features in visualization systems does not affect how individuals perceived those apparatus. Nevertheless, personality variables still hinted at their regulatory nature in human–computer interaction. In particular, this psychological construct suggested an effect on task completion time and on how individuals decided which were their design preferences. As such, we decided as the next step of our research to study how personality affects decision-making in the context of information visualization, focusing on their impact on the priming effect by cognitive biases as well as how personality affects the way users trust and which insights they can obtain from visualizations.

13.7.3.2 Limitations and Future Work

Some important factors may provide additional information regarding the lack of significance observed in some of our results. First, we have a modest sample size to test the three personality clusters. A larger number of participants would allow conclusions with a stronger impact. Future work includes recruiting more participants for follow-up studies and analyzing whether a larger sample size affects user preference, performance, and self-assessment metrics. We have to highlight that a larger sampling could change the final clusters of the personality profiling, thus adding a variation that may lead to different preferences from each cluster. Nevertheless, our methodology based on state-of-the-art research [4, 58] is sound to differentiate people based on personality factors, since each cluster was independent of the others for all main personality traits.

Conscientiousness should also be further explored independently of user preferences to leverage how this trait leads users to follow rules to achieve their goals. There is still room to explore how conscientiousness models user performance and preferences for different information visualization techniques. In particular, there is still a large set of personality and emotional intelligence traits that remain underexplored in the domain of information visualization. Among these examples, we find the FFM facets of self-discipline, aesthetics, and deliberation.

Second, although we chose the most common idioms for the contexts, more idioms could be shown to participants who may have revealed other preferences. In this line, the scenario and the complexity of the data sets used to illustrate the different contexts may have affected how people perceived the idioms. Additionally, not asking users to perform any task rather than rating their preference for the aesthetics of an idiom may not impact visual task analysis. Therefore, task types and contexts should be further explored as they may lead to distinct interactions of users given their differences. Finally, our experiment included only dashboards tuned to individual conscientiousness levels. The lack of a control dashboard may have increased the strength of the overlap degree between the groups and, consequently, affect the significance of our results. Future work should also consider adding a control condition to understand whether individuals prefer a sample dashboard or one that is designed by their preferences.

Nowadays, Personality Psychology researchers use the traditional method of validated psychological questionnaires to assess personality variables. The static and long nature of these instruments restricts practical applications since it takes a lot of time to answer the questionnaires. In addition, users may not provide truthful responses by responding in a way they think is best for a given situation or due to privacy issues [24]. These limitations led researchers to explore other ways to assess personality traits [42, 53], such as eye tracking [6], social media [5, 22], and electroencephalography (EEG) [62, 65]. Although these methods are still understudied, they allow a passive collection of personality variables leveraged by ubiquitous computing to provide adaptive information visualization systems with user context. Moreover, this assessment needs to be only performed once, since personality is often stable and these data can be saved to be used afterward.

To complement, personality-based preferences can also work the other way around by allowing designers to extract personality profiles based on user preferences, thus strengthening current personality assessment techniques. For instance, Ottley et al. [50] and Brown et al. [8] showed that mouse data predict personality factors. Since conscientiousness drives individuals to follow norms and rules while prioritizing tasks [28], we believe that the conscientiousness trait may manifest its effect through the interaction data that users show while interacting with a dashboard. In particular, we expect that individuals with high conscientiousness scores will follow an optimized search strategy more and, consequently, perform fewer interaction events such as hovers, and vice versa. Therefore, we expect that the system can predict the conscientiousness score based on the number of interactions.

13.8 Conclusions

The objective of our study was to assess how conscientiousness models user preferences regarding information visualization techniques and whether these preferences are relevant for information visualization systems design. We presented the creation of a set of design guidelines based on the relationship between conscientiousness and user preference and then designed a set of dashboards based on those guidelines. We continued by reporting a user study where we evaluate whether the dashboards lead users with different conscientiousness levels to change their task performance as well as their self-assessment of usefulness, ease of use, and preference. Our findings show that conscientiousness leads to distinct user preferences while suggesting that there is an interaction effect between conscientiousness and design guidelines in task efficiency. Additionally, individuals with low conscientiousness scores are usually faster independently of the design guidelines. Individuals also performed fewer mistakes while interacting with a dashboard designed for their conscientiousness levels. Moreover, people with lower or higher levels of this trait prefer a visualization designed specifically for them. These participants also perceive the dashboard designed for them with higher usefulness. Finally, the design guidelines lead to different perceived ease-of-use scores. Nevertheless, we were not able to find any significant interaction effects between the conscientiousness level of an individual and the design guidelines. In light of this, the present study provides important implications that may be used in the design pipeline as guidelines to customize information visualization systems.

Acknowledgments This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references UIDB/50021/2020 and SFRH/BD/144798/2019, which is a doctoral grant awarded to the first author.

References

1. H. Al-Samarraie, A. Eldenfria, and H. Dawoud. The impact of personality traits on users' information-seeking behavior. *Information Processing & Management*, 18:17–18, 10 2016.
2. G. W. Allport. Concepts of trait and personality. *Psychological Bulletin*, 24(5):284, 1927.
3. G. W. Allport. Pattern and growth in personality. *The 1937 Allport and Stagner Texts in Personality Psychology*, 1961.
4. T. Alves, B. Ramalho, J. Henriques-Calado, D. Gonçalves, and S. Gama. Exploring how personality models information visualization preferences. *arXiv preprint arXiv:2008.13133*, 2020.
5. D. Azucar, D. Marengo, and M. Settanni. Predicting the big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, 124:150–159, 2018.
6. S. Berkovsky, R. Taib, I. Koprinska, E. Wang, Y. Zeng, J. Li, and S. Kleitman. Detecting personality traits using eye-tracking data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, page 221. ACM, 2019.
7. N. Bolger and E. A. Schilling. Personality and the problems of everyday life: The role of neuroticism in exposure and reactivity to daily stressors. *Journal of Personality*, 59(3):355–386, 1991.
8. E. T. Brown, A. Ottley, H. Zhao, Q. Lin, R. Souvenir, A. Endert, and R. Chang. Finding Waldo: Learning about users from their interactions. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1663–1672, 2014.
9. S. Butt and J. G. Phillips. Personality and self-reported mobile phone use. *Computers in Human Behavior*, 24(2):346–360, 2008.
10. B. J. Carducci. *The psychology of personality: Viewpoints, research, and applications*. John Wiley & Sons, 2009.
11. C. S. Carver and M. F. Scheier. *On the self-regulation of behavior*. Cambridge University Press, 2001.
12. D. Cashman, Y. Wu, R. Chang, and A. Ottley. Inferential tasks as a data-rich evaluation method for visualization. In *EVIVA-ML: IEEE VIS Workshop on EValuation of Interactive VisuAl Machine Learning systems*, volume 7, 2019.
13. I. L. Child. Personality in culture. *Handbook of Personality Theory and Research*, pages 82–145, 1968.
14. J. Cohen. *Statistical power analysis for the behavioral sciences*. Academic Press, 2013.
15. C. Conati, G. Carenini, D. Toker, and S. Lallé. Towards user-adaptive information visualization. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
16. P. J. Corr and G. Matthews. *The Cambridge handbook of personality psychology*. Cambridge University Press Cambridge, 2009.
17. P. Costa and R. R. McCrae. The revised NEO personality inventory (NEO-PI-R). *The SAGE Handbook of Personality Theory and Assessment*, 2:179–198, 01 2008.
18. P. T. Costa Jr and R. R. McCrae. *The Revised NEO Personality Inventory (NEO-PI-R)*. Sage Publications, Inc, 2008.
19. P. T. Dinesen, A. S. Nørgaard, and R. Klemmensen. The civic personality: Personality and democratic citizenship. *Political Studies*, 62:134–152, 2014.
20. X. Du, F. Shao, S. Wu, H. Zhang, and S. Xu. Water quality assessment with hierarchical cluster analysis based on Mahalanobis distance. *Environmental Monitoring and Assessment*, 189(7):335, 2017.
21. H. J. Eysenck. The biological basis of personality. *Nature*, 199:1031–4, 10 1963.
22. G. Farnadi, G. Sitaraman, S. Sushmita, F. Celli, M. Kosinski, D. Stillwell, S. Davalos, M.-F. Moens, and M. De Cock. Computational personality recognition in social media. *User Modeling and User-Adapted Interaction*, 26(2-3):109–142, 2016.

23. D. B. d. S. Ferreira. *A influência do modelo de aceitação da tecnologia na utilização da rede social Facebook na perspetiva do perfil pessoal*. PhD thesis, Instituto Politécnico do Porto. Instituto Superior de Contabilidade e Administração do Porto, 2014.
24. A. Friedman, B. P. Knijnenburg, K. Vanhecke, L. Martens, and S. Berkovsky. Privacy aspects of recommender systems. In *Recommender Systems Handbook*, pages 649–688. Springer, 2015.
25. L. Goldberg. The structure of phenotypic personality traits. *American Psychologist*, 48:26–34, 02 1993.
26. T. M. Green and B. Fisher. Towards the personal equation of interaction: The impact of personality factors on visual analytics interface interaction. In *2010 IEEE Symposium on Visual Analytics Science and Technology*, pages 203–210. IEEE, 2010.
27. J. J. Gross and O. P. John. Mapping the domain of expressivity: Multimethod evidence for a hierarchical model. *Journal of Personality and Social Psychology*, 74:170–91, 02 1998.
28. S. Halko and J. A. Kientz. Personality and persuasive technology: an exploratory study on health-promoting mobile applications. In *International Conference on Persuasive Technology*, pages 150–161. Springer, 06 2010.
29. J. Han, J. Pei, and M. Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.
30. M. Ilayaraja and T. Meyyappan. Mining medical data to identify frequent diseases using apriori algorithm. In *2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering*, pages 194–199. IEEE, 2013.
31. C. G. Jung. *Psychological types*. Routledge, 2014.
32. D. Kolb and A. Kolb. *The Kolb Learning Style Inventory 4.0: Guide to Theory, Psychometrics, Research & Applications*. Experience Based Learning Systems, 01 2013.
33. D. A. Kolb. *Experiential learning: Experience as the source of learning and development*. FT Press, 2014.
34. M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel. Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, 95(3):357–380, 2014.
35. C. C. Kuhlthau. A principle of uncertainty for information seeking. *Journal of Documentation*, 49(4):339–355, 1993.
36. S. Lallé and C. Conati. The role of user differences in customization: a case study in personalization for infovis-based content. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*, pages 329–339, 2019.
37. K. Lee and M. C. Ashton. Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2):329–358, 2004.
38. H. M. Lefcourt. *Locus of control: Current trends in theory and research*. Psychology Press, 2014.
39. H. Levenson. *Reliability and Validity of the I, P, and C Scales-A Multidimensional View of Locus of Control*. ERIC, 1973.
40. M. Lima and A. Simões. NEO-PI-R manual professional. *Lisboa: CEGOC*, 2000.
41. Z. Liu, R. J. Crouser, and A. Ottley. Survey on individual differences in visualization. *arXiv preprint arXiv:2002.07950*, 2020.
42. N. Majumder, S. Poria, A. Gelbukh, and E. Cambria. Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems*, 32(2):74–79, 2017.
43. D. P. McAdams. Three lines of personality development. *European Psychologist*, 2015.
44. R. R. McCrae and O. P. John. An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2):175–215, 1992.
45. L. McInnes, J. Healy, and S. Astels. HDBSCAN: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205, 2017.
46. H. A. Murray. *Explorations in personality: A clinical and experimental study of fifty men of college age*. Oxford Univ. Press, 1938.
47. I. B. Myers, M. H. McCaulley, N. L. Quenk, and A. L. Hammer. *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator*, volume 3. Consulting Psychologists Press Palo Alto, CA, 1998.

48. T. P. O'brien, L. Bernold, and D. Akroyd. Myers-Briggs type indicator and academic achievement in engineering education. *International Journal of Engineering Education*, 14, 01 1998.
49. N. Oscar, S. Mejía, R. Metoyer, and K. Hooker. Towards personalized visualization: Information granularity, situation, and personality. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, pages 811–819, 2017.
50. A. Ottley, H. Yang, and R. Chang. Personality as a predictor of user strategy: How locus of control affects search strategies on tree visualizations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3251–3254, 2015.
51. D. Ozer and V. Benet. Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, 57:401–21, 02 2006.
52. S. Petrovic. A comparison between the silhouette index and the Davies-Bouldin index in labelling IDS clusters. In *Proceedings of the 11th Nordic Workshop of Secure IT Systems*, pages 53–64, 2006.
53. L. Qiu, H. Lin, J. Ramsay, and F. Yang. You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*, 46(6):710–718, 2012.
54. J. B. Rotter. *Social learning and clinical psychology*. Prentice-Hall, Inc, 1954.
55. J. B. Rotter. General expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80:1–28, 02 1966.
56. J. B. Rotter. Some problems and misconceptions related to the construct of internal versus external control of reinforcement. *Journal of Consulting and Clinical Psychology*, 43(1):56, 1975.
57. M. T. Russell, R. B. Cattell, A. Cattell, H. E. Cattell, and D. L. Karol. *16PF fifth edition administrator's manual*. Institute for Personality and Ability Testing, Incorporated, 1994.
58. S. M. Sarsam and H. Al-Samarraie. A first look at the effectiveness of personality dimensions in promoting users' satisfaction with the system. *SAGE Open*, 8(2):2158244018769125, 2018.
59. J. Sheidin, J. Lanir, C. Conati, D. Toker, and T. Kuflik. The effect of user characteristics in time series visualizations. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, pages 380–389, 2020.
60. M. N. Shiota, D. Keltner, and O. P. John. Positive emotion dispositions differentially associated with big five personality and attachment style. *The Journal of Positive Psychology*, 1(2):61–71, 2006.
61. P. M. d. Silva, G. A. Dias, and J. R. d. Almeida. *Modelo de aceitação de tecnologia (TAM) aplicado ao Sistema de Informação da Biblioteca Virtual em Saúde (BVS) nas Escolas de Medicina da Região Metropolitana do Recife*. PhD thesis, Universidade Federal da Paraíba, Brasil, 2013.
62. R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe. Ascertain: Emotion and personality recognition using commercial sensors. *IEEE Transactions on Affective Computing*, 9(2):147–160, 2016.
63. V. Venkatesh and H. Bala. Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, 39(2):273–315, 2008.
64. M. A. d. S. Vilar. *Modelo de Aceitação da Tecnologia adaptado às compras online*. PhD thesis, [sn], 2013.
65. J. Wache, R. Subramanian, M. K. Abadi, R.-L. Vieriu, N. Sebe, and S. Winkler. Implicit user-centric personality recognition based on physiological responses to emotional videos. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 239–246. ACM, 2015.
66. J. Wang, J. Wang, Q. Ke, G. Zeng, and S. Li. Fast approximate k-means via cluster closures. In *Multimedia Data Mining and Analytics*, pages 373–395. Springer, 2015.
67. C. Xiong, L. Padilla, K. Grayson, and S. Franconeri. Examining the components of trust in map-based visualizations. In R. Kosara, K. Lawonn, L. Linsen, and N. Smit, editors, *EuroVis Workshop on Trustworthy Visualization (TrustVis)*. The Eurographics Association, 2019.

68. C. Ziemkiewicz, R. J. Crouser, A. R. Yauilla, S. L. Su, W. Ribarsky, and R. Chang. How locus of control influences compatibility with visualization style. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 81–90. IEEE, 2011.
69. C. Ziemkiewicz and R. Kosara. Preconceptions and individual differences in understanding visual metaphors. *Computer Graphics Forum*, 28(3):911–918, 2009.
70. C. Ziemkiewicz, A. Ottley, R. J. Crouser, A. R. Yauilla, S. L. Su, W. Ribarsky, and R. Chang. How visualization layout relates to locus of control and other personality factors. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1109–1121, 2012.

Chapter 14
Visualizing Uncertainty in Different Domains: Commonalities and Potential Impacts on Human Decision-Making

Laura E. Matzen, Alisa Rogers, and Breannan Howell

Abstract Visualizing uncertainty is a difficult but important task. Many techniques for visualizing uncertainty are designed for a specific domain, such as cartography or scientific visualization, and the effectiveness of these techniques is tested within that domain (when it is tested at all). This makes it difficult to generalize the findings to other tasks and domains. Recent work in visualization psychology has begun to focus on this problem from the perspective of how different visualization techniques impact human cognitive processes, including perception, memory, and decision-making. Taking this perspective allows researchers to develop theories that can generalize across domains. This is a rich area for research, but given the large number of papers about uncertainty visualization, it can be difficult to know where to begin. The goal of this chapter is to provide a broad overview of what kinds of uncertainty visualization techniques have been developed in different domains, which ones have been evaluated with respect to their impact on human cognition, and where important gaps remain.

14.1 Introduction

Data visualizations are useful tools for helping people to explore, understand, and make decisions based on data. To support high-quality decision-making, it is often necessary for data visualizations to include some representation of the uncertainty in the data or in the analysis pipeline that produced the visualization. However, designing effective visualizations that include uncertainty information is extremely challenging. While there have been numerous studies that have developed and tested different methods for incorporating information about uncertainty into visualizations, to date there are few coherent guidelines regarding the best choices

L. E. Matzen (✉) · A. Rogers · B. Howell
Sandia National Laboratories, Albuquerque, NM, USA
e-mail: lematze@sandia.gov; anroger@sandia.gov; bchowel@sandia.gov

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_14

for communicating uncertainty without overwhelming the viewer or exacerbating cognitive biases.

Recent work has attempted to address these gaps by linking visualization comprehension in general (e.g., [102, 118]) and uncertainty visualizations specifically [116, 117] to theoretical frameworks in cognitive science. This work has identified perceptual and cognitive factors that lead to pervasive biases in the comprehension of visualizations. For example, people tend to treat a region that is surrounded by a visual boundary as being categorically different from regions outside of that boundary. This can have substantial impacts on their decision-making, even when the location of the visual boundary is somewhat arbitrary [117] or an artifact of the mapping between uncertainty information and visual attributes [126].

The emerging field of visualization psychology has identified a number of common visualizations that consistently produce biases in interpretation and decision-making. These biases can stem from intuitions about color maps [138], misunderstandings of probability [72], misinterpretation of common visual cues such as error bars [101], and interpretations of visual boundaries [117]. Much of the work that has identified these common biases has focused on visualizations of statistical variability, such as bar graphs, and spatial visualizations, such as hurricane forecasts. However, the same types of biases are highly likely to appear in other domains, since they are caused by human perceptual and cognitive processes.

Visualizations of uncertain information are especially prone to errors or biases in interpretation (cf. [116]). Humans struggle to deal with uncertainty under the best of circumstances, and it can be difficult to characterize the uncertainty in a dataset, much less visualize it in a way that viewers will find easy to understand. However, people make high-consequence decisions based on uncertain information all the time, and visualizations often play a crucial role in the decision-making process.

To give a concrete example, a real-world need for uncertainty visualizations comes from the emergency preparedness domain. To prepare for natural or man-made disasters, scientists develop mathematical models and hazard maps that can aid in evacuations. The maps are intended to help local emergency response teams make decisions about which areas to evacuate, yet there is uncertainty inherent in these visualizations because the risks cannot be modeled or predicted with perfect accuracy. There are tradeoffs between risks caused by the disaster itself and the risks that are inherent in large-scale evacuations. Evacuating an area that is too large would unnecessarily put more people at risk from other hazards, such as car crashes. Given these tradeoffs, what are the best ways to present information about uncertainty and risk so that these decision-makers can make the best possible decisions under stressful circumstances? To date, there is little research that would support a principled answer to this question.

Research on visualization psychology has begun to develop frameworks that can help visualization designers to take human cognition into account. However, there is much more work to be done in this domain. There are numerous methods for visualizing uncertainty, and relatively few have been evaluated in controlled experiments that can indicate why one method supported better decision-making performance than another [64, 66]. The uncertainty visualization literature spans numerous domains, including cartography, scientific visualization, medicine, and natural resource management. While there have been numerous reviews and taxonomies of uncertainty visualization techniques written within the visualization research community, they tend to focus on the data being visualized rather than the *people* who will be using those visualizations. The lack of consensus about when and how to visualize information about uncertainty is a challenge for visualization designers [65]. This can also make it daunting for cognition researchers to identify the many techniques that are ripe for study from the perspective of their impact on human cognition and decision-making.

The goal of this chapter is to compile the prior research on uncertainty visualizations in a way that will inspire new research in the field of visualization psychology. We provide a broad survey of the different methods of uncertainty visualization that are commonly used in different domains, including statistical graphs, spatial and temporal visualizations, three-dimensional visualizations, and visualizations designed to support risk assessments and decision-making. Our focus is on identifying techniques that have been used in multiple domains and, when possible, what evaluations of those techniques can tell us about their effectiveness in terms of supporting the viewer's decision-making process.

We begin by providing an overview of the research on how people deal with uncertainty in general and then discuss important considerations for designing visualizations that incorporate uncertainty. We then review common methods of visualizing uncertainty for different types of datasets, highlighting those that have been empirically tested with respect to their impact on human decision-making. We identify common visual mappings and discuss the types of perceptual and cognitive biases that can be produced by each, linking these techniques back to the emerging research on visualization cognition.

We approach the literature on uncertainty visualization from the perspective of human cognition because humans are the consumers of data visualizations, and they are using them to accomplish specific goals. By focusing on those goals and the types of uncertainty that are important in those situations, visualization designers can narrow down the range of visual metaphors that are appropriate to their specific application and weigh their pros and cons from the perspective of human cognition. In addition, cognition researchers can identify gaps in the existing literature where new research in visualization psychology can be brought to bear.

14.2 Uncertainty and Human Decision-Making

Several studies have demonstrated that information about uncertainty, whether verbal or visual, can impact people's decision-making (e.g., [1, 5, 27, 33, 74, 89]). Making decisions under uncertainty is inherently difficult because even a "good" decision, based on the best available information, can lead to the wrong outcome [66]. In decision-making contexts, uncertainty often takes the form of different states or different outcomes that could occur with different probabilities. Numerous studies have demonstrated that people struggle to understand probabilities in general (cf. [46, 83]). When confronted with information that has uncertain accuracy, people often reject it completely, ignoring it during their decision-making process (cf. [59]). Alternatively, they may accept it completely, ignoring the probabilistic nature of the uncertainty [59] or misinterpreting uncertain information as if it were deterministic [72].

Foundational work by Tversky and Kahneman showed that when people make decisions in the face of uncertainty, they tend to use heuristics to reduce the complexity of the task. While these heuristics are generally very useful, they can lead to biases [145] and make it difficult for people to assess their own ability to manage uncertainty [114]. Three key biases described by Tversky and Kahneman are the representativeness heuristic, the availability heuristic, and the anchoring heuristic. Each of these heuristics has the potential to impact how people interpret uncertain information.

The representativeness heuristic arises when people judge a particular item or event to be part of a category due to its similarity to other items or events in that category. People tend to give more weight to superficial similarities or stereotypes while giving less weight to sample sizes and the baseline probability of encountering the item or event in question [40, 51, 75, 145]. In the case of visualizations, visual similarities can invoke the representativeness heuristic. For example, visual boundaries can lead people to treat things inside of the boundary as being categorically different from things outside of the boundary, even if that is not actually the case [118].

The availability heuristic often comes into play when people assess the probability of an event or the frequency of a class. People tend to give higher estimates of frequency or probability in cases where they have personal experience with a particular outcome, cases where they can easily generate or imagine a larger number of examples, and cases with vivid or recent examples. While these patterns can indicate that there is a higher prevalence for a particular type of event, they can also produce substantial biases in reasoning [59, 75]. There is some evidence indicating that visual representations of information can also make that information more "available," which could potentially lead to biases in reasoning and risk assessment. For example, comparisons of visual, verbal, and numerical representations of uncertainty have found that visual representations can lead people to treat the information as being more specific or more certain, changing the threshold at which they decide to act on the uncertain information [6, 72].

Anchoring bias occurs when people form an initial impression and make adjustments based on subsequent information. In many cases, those adjustments are too small, leading to a final estimate that is biased toward the person's initial estimate [76, 109]. This can occur even if the information that supported the initial estimate is later discredited. The impact of discredited information can persist, biasing a person's ultimate judgment [59]. In other instances, people choose to discount new evidence that contradicts their preferred hypothesis [59, 144]. In the case of visualizations, people may anchor their reasoning to one aspect of the visualization, such as the average value for a dataset, producing biases in judgment [112].

In addition to cognitive biases, humans are subject to perceptual biases. Our top-down expectations about what we are going to see can impact and even override the bottom-up perceptual processing that is driven by the physical characteristics of a stimulus [23, 142, 149]. In other words, we tend to see what we expect to see [59]. In the context of data visualizations, there is also evidence that we also expect other people to see the same things that we see, even when that is not necessarily the case [151].

Perceptual biases also occur in other sensory modalities, such as touch. People are subject to illusions such as the size–weight illusion, where the smaller of two objects with the same mass is perceived to be heavier [135]. In our daily lives, we tend to compensate for these perceptual illusions because we can get real-time feedback. For example, even though people perceive smaller objects to be heavier no matter how many times they lift them, their motor system will adjust to the actual weight of an object after the first lift [135]. When dealing with data visualizations, there is often no feedback to correct for perceptual biases. However, recent research has shown that interactive visualizations that allow users to remove information that is salient but distracting can help to reduce biases in decision-making [28].

Several recent studies have identified perceptual biases that arise in the context of interpreting uncertainty visualizations (see [117], for a review). Even simple, widely used representations such as error bars can produce perceptual biases that lead to misinterpretations. These biases persist despite repeated exposure and domain expertise [24, 101, 112]. Other aspects of visualizations, such as visual boundaries, presentation of low-relevance items, occlusion, distortion, or over-simplification, can lead to perceptual and cognitive biases, exacerbating the problem of decision-making in the face of uncertainty [14, 35, 117]. It may be feasible to reduce these biases in some cases, by choosing better visual representations of uncertainty or by introducing features such as interactivity. However, only a handful of studies have attempted to reduce biases through the manipulation of visual features or other aspects of visualization design, and those approaches have had mixed success [28].

14.2.1 How Do Different Representations of Uncertainty Impact Decision-Making?

Given the various pitfalls related to decision-making under uncertainty, what kinds of representations are best for supporting unbiased (or at least less biased) decision-making? A number of studies have compared verbal, numerical, and/or visual representations of uncertainty in order to address this question.

Decision-making performance can be measured in terms of the speed of the decision, its accuracy, or both. It can also be assessed in terms of how well participants understand the risk involved in different decisions. Decision-making and human understanding of risk have received a great deal of attention from psychology researchers, so there are a relatively large number of studies that make direct comparisons between different visual, verbal, and numerical indicators of state uncertainty.

Verbal representations of uncertainty can include descriptions of the uncertainty itself (i.e., "high," "medium," or "low") or terms that describe the likelihood of a particular event, such as "very unlikely," "rather likely," or "almost certain" [6]. Numerical representations of uncertainty can include probabilities (0.15), percentages (30%), ranges of probabilities or percentages (50–65%), odds ratios ("4.32 times more likely"), or statistics about a distribution, such as mean and confidence intervals [6, 48, 72]. Visual representations of probabilistic information include pictographs [86], plots of probability distributions [66], and histograms [48, 66]. Many decision-making tasks also involve state uncertainty, situations in which the current or future status of some phenomenon is unknown. State uncertainty can be visualized in a variety of ways, as detailed in the sections that follow. These visual cues are often combined with verbal or numerical cues, particularly for tasks that involve assessments of risk [18, 43].

Several studies have demonstrated that the way in which uncertain information is presented, be it numerically, verbally, or visually, has an impact on human decision-making performance. These studies have used a number of different tasks, including target detection tasks, choosing stocks, and making decisions based on weather forecasts [43, 82, 110]. The results indicate that visual representations of probability distributions can improve people's understanding and decision-making relative to verbal or numeric information [48]. Visualizations of uncertainty can be particularly helpful when a task is difficult [18, 82] or when uncertainty is high [72, 74]. Some studies have also provided evidence that visualizations of uncertainty can help to mitigate biases in decision-making [73]. Combining information from multiple sources helps participants to understand uncertainty in the data and improves their trust in predictions based on that data [49, 72, 78].

A common finding across these studies is that when the task is relatively easy, such as when participants have unlimited time to make their decision, the way in which the uncertainty is represented does not typically have much impact on participants' performance (cf. [6, 18, 82]). However, when the task difficulty is increased by adding time pressure or by adding more noise to the data, participants typically perform better with visual representations of uncertainty. For example, in target detection/identification tasks, participants tend to make decisions faster [43] and more accurately [82] when using visual representations instead of verbal or numerical representations. In a wildfire evacuation scenario [18], participants who had ample time to decide whether or not to evacuate performed best with text-based information, such as "Your house is located in the >80 to 100% burn zone." However, when time pressure was applied, making the task more difficult, maps with visual representations of uncertainty outperformed the text condition. Similarly, Kirschenbaum and Arruda [82] found that participants performed equally well with verbal and visual representations of uncertainty when their target detection task was relatively easy. When more noise was added to the data, making the task more difficult, participants performed better when using visual representations of uncertainty.

At the extreme end of the scale, when tasks become extremely difficult, all representations of uncertainty can become equally unhelpful. In the wildfire evacuation scenario used by Cheong and colleagues [18], adding a secondary task made participants' performance to drop to chance, regardless of what uncertainty information they had or how it was represented. When the participants had high cognitive load, they were not able to make effective use of any of the representations of uncertainty.

Although people often perform better when uncertainty information is provided than when it is not, many prefer point estimates and ignore uncertainty information when given a choice [78, 119]. They may not use all of the information that is available to them, particularly when that information is uncertain or probabilistic [26, 49]. Instead, they may be drawn to a single, salient piece of information at the expense of exploring all of the information [26], or they may develop a routine in which they use a few favored sources of information while ignoring others [71]. Finally, people who are using visualizations of uncertain information may fall prey to deterministic construal errors, in which they assume that the information they are seeing is deterministic rather than probabilistic. For example, Savelli and Joslyn [137] found that some participants interpreted a visual representation of the uncertainty in a weather forecast as if it were a deterministic forecast showing the high and low temperature. This type of error did not occur when participants saw only text, with no visualization of the uncertainty [72].

Experience with a particular domain or task also plays a role in whether and how people use information about uncertainty to make decisions. For example, Cliburn and colleagues [21] found that less experienced participants preferred a less complex representation of uncertainty, while more experienced participants preferred a representation that was complex and somewhat difficult to understand but conveyed more detailed information. Repeated exposure to a decision-making task with uncertain information can help participants to learn what kinds of information are most important to the task [110]. When studies use tasks that are highly familiar to the participants, such as asking experienced surgeons to complete a common procedure, participants may ignore visualizations of uncertainty, opting instead to look at the displays with which they are already familiar [139]. Aside from experience, individual differences in cognitive processing also play an important role in how people comprehend visualizations of uncertainty. For example, eye-tracking data have shown that people with low and high numeracy (ability to understand numerical information) used very different processing strategies when shown pictographs that conveyed numerical information about risk [86].

In summary, numerous studies have demonstrated that humans struggle to deal with uncertain information and that the way in which that information is presented to people (whether numerically, verbally, or visually) can impact their decision-making and task performance. Visual representations of uncertainty can help people to understand probabilistic or uncertain information, but the effectiveness of those visualizations often depends on the nature and difficulty of the task as well as the experience level of the participants. Even when useful information about uncertainty is available, people may not take advantage of it.

Given that effective visualizations can help people to comprehend uncertainty and improve their decisions, how do we design visualizations that are effective for a particular task, data type, or user population? In the next section, we present several of the challenges that arise for designing and evaluating uncertainty visualizations.

14.3 Why Is Visualization of Uncertainty Difficult?

Uncertainty visualizations can be useful tools, but they also present a number of challenges for visualization designers and viewers alike.

Designing effective data visualizations is a non-trivial task, and adding information about uncertainty to the mix can dramatically increase the level of difficulty [65]. As more information is added to a visualization, it is easy for the visualization to become cluttered or confusing [8, 50, 66, 124]. Designers are faced with many choices about how to represent uncertainty [50, 121, 124]. It can be represented directly by mapping it to unused visual variables (e.g., color, size, transparency), adding graphical elements (e.g., error bars, glyphs, labels, overlays), altering graphical elements (e.g., scaling, warping, distortion), creating multiple visualizations to enable side-by-side comparisons, adding dynamic features (motion, animation, or interactivity that reveals information about uncertainty), or adding non-visual features, such as sound. Uncertainty can also be used to modify the visualization in less direct ways, such as filtering the data or changing the way in which the data are processed, weighted, or modeled [50]. Both approaches have their drawbacks. Direct representations of uncertainty may not scale well to large datasets [50], may obscure important information [8], or can potentially lead to visualizations that are cluttered and confusing to the viewer [124]. Indirect representations may hide information that the viewer needs, lead to misinterpretations of the information, or lead to overconfidence or other types of cognitive biases if the viewer does not fully understand how uncertainty impacts the visualization.

Visualizations of uncertainty are also very difficult to evaluate [64, 66, 80, 127]. A recent review paper by Hullman and colleagues (2018) characterized common methods for evaluating uncertainty visualizations. They found that most papers on uncertainty visualization did not evaluate the visualization's effectiveness. For example, in one collection of 241 publications on uncertainty visualizations [124], they found only 48 papers that contained evaluative user studies. Among papers that did include user studies, few used realistic tasks that would translate to real-world decision-making, likely because it is very difficult to design realistic evaluation tasks with that are controlled enough to support clear conclusions [66, 80, 127]. In addition, few evaluations compared uncertainty visualizations to those that had no representation of uncertainty, making it difficult to evaluate whether adding uncertainty information was helpful to participants. Similarly, few studies tried to explore the *reasons* for the effects that they found (i.e., why and how participants made their decisions). Thus, even for studies that conducted empirical tests of their uncertainty visualization techniques, the challenges of designing evaluation studies mean that we are still left with relatively little insight into how the studied techniques would impact human decision-making in the real world.

14.3.1 We Do Not Really Know What Uncertainty Is

As complex as it is to select an appropriate and useful method for visualizing uncertainty, there is also a deeper, fundamental problem: there is no clear definition of uncertainty (cf. [98, 121, 140]). In theory, there are two fundamental types of uncertainty: aleatoric uncertainty, which is due to randomness and cannot be resolved, and epistemic uncertainty, which is due to a lack of knowledge and could be resolved with access to additional information [8, 124]. In practice, the term "uncertainty" is often used as a catch-all for all sorts of problems, many of which represent a mixture of aleatoric and epistemic uncertainty. This can include problems with data acquisition, such as missing data, erroneous data, imprecise measurements, noise, contradictory information, and outdated or untrustworthy sources [8, 50, 69, 99, 146]. It can also include problems that appear along the data processing pipeline, such as simplifying abstractions, resampling or interpolation, incorrect models or model parameters, or human error [8, 50, 69, 99, 103, 121, 150]. In addition to introducing uncertainty, these issues can create visualization mirages where a visualization appears to support an interpretation that is not actually accurate [105]. Finally, uncertainty can include issues that are introduced by the visualization process itself, such as the choice of rendering algorithms [121], and by perceptual or cognitive difficulties encountered by people viewing the visualization [8, 12, 69].

The wide array of issues that fall under the heading of "uncertainty" in the realm of data visualization creates challenges. In practice, uncertainty is often treated as an attribute of the data [99]. Visualizations of uncertainty typically depict the presence of uncertainty and sometimes the amount or location of the uncertainty, but other aspects of uncertainty are rarely visualized, and different sources or causes of uncertainty are rarely distinguished from one another. For example, in the public health domain, data are collected and aggregated from numerous disparate sources, which leads to errors and discrepancies. Public health experts are knowledgeable about the causes and impact of this implicit measurement error, but information about the inherent uncertainty in the data is not explicitly measured or recorded. This represents a challenge for data visualization, since visualizations of data with implicit error will not reflect the domain experts' knowledge of the situation. However, it also presents opportunities for using visualizations to expose and analyze implicit error [103].

In summary, any given dataset and analysis pipeline will have different types of uncertainties, and two different visualizations of uncertainty, even in the same domain, may be representing very different types of information. Visualization designers must make choices about what to represent, but it is often unclear which types of uncertainty are meaningful or helpful to the people who are using the visualizations.

14.3.2 Why Should We Bother?

Given all of these difficulties, how should we proceed? As discussed in Sect. 14.2, uncertainty visualizations can be helpful for conveying risk [18, 86], making decisions [43, 82, 110], and avoiding cognitive biases. But are they worth the effort if they are so difficult to design and to evaluate? Should we bother to visualize uncertainty at all?

In many cases, the answer is still "yes." Comprehension of uncertainty is often necessary for establishing scientific or analytical understanding [117]. For this reason, various organizations have stressed the importance of conveying uncertainty [68, 111, 143]. Many visualization designers work in domains where visualizing uncertainty is crucially important, and they must figure out how to do it in the most effective way possible.

Fortunately, there is growing body of studies that have assessed the impact of different representations of uncertainty on human comprehension, task performance, or decision-making. These studies are often specific to a particular task or data type (such as geospatial data or flow modeling), but it is likely that there are commonalities in how different representations of uncertainty impact human cognition across a variety of different tasks and domains. Visualization psychology has emerged as a field of research that looks for systematic relationships between cues that are used in visualizations and patterns of human cognition and decision-making. This research is helping to connect the large body of research on human cognition to the research on visualization techniques, enabling the development of principled approaches to designing visualizations that will best support cognition [45, 117, 118].

In this chapter, we aim to aid both visualization practitioners and visualization psychology researchers by surveying approaches to uncertainty visualization from multiple domains. In Sect. 14.4, we discuss the types of questions that can inform choices about the design of a visualization that includes information about uncertainty. In Sect. 14.5, we present a high-level overview of different techniques that are commonly used to convey information about uncertainty. In Sect. 14.6, we dig into more detail to examine how different approaches to uncertainty visualization have been applied within specific domains. Wherever possible, we focus on methods that have been evaluated and shown to benefit users' performance or cognitive processing. Our goal is to identify the advantages and disadvantages of different types of visual mappings, based on prior empirical comparisons and theories of visualization psychology.

Methods for quantifying uncertainty are outside of the scope of this review. While this is an interesting and challenging problem in and of itself, our focus is on presenting an overview of the current landscape of uncertainty visualization methods. Once the relevant uncertainty in a dataset, analysis pipeline, or rendering approach has been characterized, what are the most effective methods for conveying that information to the viewer? For overviews of methods for quantifying or characterizing uncertainty, we direct the reader to prior review papers on that topic [8, 50, 140, 144].

14.4 Design Considerations for Uncertainty Visualizations

The starting point when designing a visualization that includes information about uncertainty should be the needs of the intended users. A very useful approach for thinking about the users' needs is the typology of visualization tasks developed by Brehmer and Munzer [13] (see also [108]). This typology considers why people are performing a task, how they are performing that task, and what kinds of information are involved. A similar approach could be used for thinking about the design of uncertainty visualizations: *why* do users need information about uncertainty and *why* are specific types of uncertainty important? *How* are users interacting with the data visualization and *how* is uncertainty likely to impact those interactions? *What* type of data or information is being used and *what* kinds of visual representations are appropriate? The answers to these questions can vary widely across domains, and they can help to constrain the range of possible uncertainty visualization techniques that should be considered.

14.4.1 Why Do Users Need Information About Uncertainty?

The key question for uncertainty visualizations is what types of uncertainty are important for the intended users and use cases. As discussed above, there are numerous sources of uncertainty in any data collection, analysis, and visualization pipeline. However, some of these sources of uncertainty may be critically important to the user, while others may be irrelevant. Since uncertainty visualization is difficult to begin with, it is important to focus only on the types of uncertainty that will add value for the user, such as those that should have a high weight in the user's decision-making [51]. In some cases, representing the *existence* of uncertainty is sufficient. This is relatively easy and can be accomplished with many different methods [50]. In other cases, it is important to convey the *amount* of uncertainty, which requires a quantitative approach, or the *nature* of the uncertainty, which could be quantitative or qualitative. The specific details of these situations constrain the types of visual metaphors that can be used to convey the relevant information. Finally, in many situations, there are multiple types of uncertainty that are important to the user [98, 140, 144]. It may be necessary to prioritize certain types of uncertainty over others or to create multiple visualizations of a single dataset to address this challenge. A common approach is to simplify the uncertainty information into a more manageable value, such as a standard deviation, but this can lead to misrepresentations of the data [124].

14.4.2 How Will Uncertainty Impact Users' Interactions with the Data Visualization?

Here we must consider the cognitive impacts of uncertainty and how uncertainty visualizations might alter users' approach to their tasks. Skeels and colleagues [140] found that people confronted with uncertainty either sought ways to increase their certainty or reached a point at which they had to live with the uncertainty and make a decision despite it. In other cases, people confronted with uncertain outputs will mentally substitute that information with other information that is easier to understand [117]. These findings highlight the importance of considering how users will respond to information about uncertainty. Ideally, they will use it to understand the limitations of the information they have and then seek out additional information to inform a more nuanced decision. However, if we do not consider the difficulty that people, including domain experts, have with comprehending uncertainty, visualizations of uncertainty might confuse and frustrate viewers, contribute to cognitive biases, lead to poor decisions, or cause them to throw up their hands and give up altogether.

14.4.3 What Kinds of Visual Representations are Appropriate?

In many cases, the nature of the data and the uncertainty being visualized constrain the number of viable visual metaphors. Many of the existing taxonomies of uncertainty visualization map characteristics of the data, such as its dimensionality, to appropriate visualization techniques. For example, Pang and colleagues [121] developed a classification of visualization methods with five characteristics:

1. The value of a datum and the uncertainty of that value (scalar, vector, tensor, multivariate)
2. The location of a datum and its positional uncertainty (spatial, temporal, etc.)
3. The extent of datum location and value (discrete or continuous)
4. The extent of the visualization (discrete or continuous)
5. Axes mapping (experiential or abstract)

By contrasting any of these characteristics against one another, it is possible to map out the approaches to uncertainty visualization that provide the best fit for that set of characteristics. For example, datasets with scalar values and a discrete visualization extent are good candidates for techniques such as error bars, while scalar data with a continuous visualization extent are better suited to techniques such as contour lines or color maps.

Potter and colleagues [125] simplified this approach by developing a taxonomy that contrasts the dimensionality of the data (1D, 2D, 3D, ND) with the dimensionality of the uncertainty (scalar, vector, or tensor). They describe the uncertainty visualization approaches that have been used for various combinations of these dimensions and identify combinations where techniques have yet to be developed.

In addition to considering the dimensionality of the data and of the uncertainty, it is important to consider the types of visual metaphors that are commonly used in a particular domain. Some domains lend themselves naturally to certain types of representations, such as the use of cartographic representations for geographic information systems (GIS) data [98]. Similarly, isosurfaces representing uncertainty are widely used in 3D scientific visualizations [69, 124]. For these domains, these techniques are experiential visualization methods, which use representations that link to a viewer's experience with the visualized phenomenon [121]. Other domains may lend themselves better to abstract mappings, such as projecting multidimensional data into a 2D scatterplot or using other types of aggregation and clustering methods.

It is important to note that some types of visual variables work better for specific types of uncertainty than others. There are fewer techniques that can be applied to 3D data than there are for 1D or 2D data [124]. Yet even for 1D or 2D data, not all representations are created equal. MacEachren [98] points out that certain types of visual representations are better suited for ordered or numerical data, while others are better suited for representing nominal or categorical differences. In the case of categorical data, the number of categories that the user must track is also important. It is difficult for viewers to track more than a handful of textures or colors [30, 98]. In addition, some representations of uncertainty, such as glyphs added to a display, may work well for small datasets but will not scale to larger datasets [121]. Finally, some visual representations of uncertainty are more intuitive to viewers than others for certain types of information. For example, the visual metaphors that are most intuitive for representing spatial uncertainty differ from those that are most intuitive for representing temporal uncertainty [100].

14.5 Common Methods for Visualizing Uncertainty

As discussed in Sect. 14.5, uncertainty can be represented explicitly, by mapping it to visual features in the visualization, or it can be represented implicitly, by changing the way the data are processed, filtered, or weighted. Although at least one study has indicated that people can successfully identify the presence of uncertainty when it is visualized implicitly rather than explicitly [27], the vast majority of the research on visualizing uncertainty has focused on explicit representations.

Explicit representations of uncertainty can include both intrinsic methods, where the appearance of an object is changed, and extrinsic methods, where symbols such as glyphs or error bars are added to an image [63]. A third approach is to create multiple visualizations (or animations) to depict uncertainty via comparisons. In this section, we compare and contrast these common approaches. The examples that follow do not present an exhaustive list of methods for representing uncertainty, but rather aim to illustrate the advantages and disadvantages of using intrinsic representations, extrinsic representations, and multiple visualizations to convey information about uncertainty.

14.5.1 Intrinsic Representations of Uncertainty: Modifying Visual Attributes

Any graphical variable (color, size, shape, etc.) that is not being used to represent other attributes of a dataset can be used to represent uncertainty about the data. In practice, the most common approach for intrinsic representations of uncertainty is to use fog as a visual metaphor [98, 100, 117]. Visualization designers manipulate color and/or texture to make more certain data appear crisp while less certain data appear blurry or faint (Fig. 14.1). While this analogy is rarely stated explicitly, it shows up in many different domains, as illustrated in the sections below. The fog metaphor also exploits the human visual system's tendency to perceive things with sharper edges and more saturated colors as being closer, while objects with fuzzier edges or muted colors are perceived as being farther away [145]. This means that the uncertain information is perceived as if it is farther away from the viewer and therefore less "solid" or less important.

Fig. 14.1 An example of saturation and edge fuzziness being manipulated using the fog metaphor for uncertainty

Visual attributes that are used in service of the fog metaphor include blur, reduced saturation, reduced sharpness, transparency, and dashed or irregular outlines. Fogginess is a natural and easily understandable representation of uncertainty, so these techniques are often effective, as demonstrated by the papers reviewed in this chapter. However, this approach is not without its drawbacks. In some cases, viewers interpret these types of visual manipulations as stylistic rather than mapping them to uncertainty [12]. The fog metaphor can also be interpreted in conflicting ways. Some researchers have used increased *opacity* to represent increased certainty [34], while others have used increased *transparency* to represent increased certainty [98]. Given that even a naturalistic metaphor can be applied in different ways, viewers must be informed of the meaning of these cues in order to interpret them correctly.

Color coding is another visual manipulation that is widely used to encode information about uncertainty. Color can be manipulated in a variety of ways by changing hue, value, saturation, or transparency (Fig. 14.2). A common approach is to map uncertainty to saturation (i.e., data with higher certainty are represented

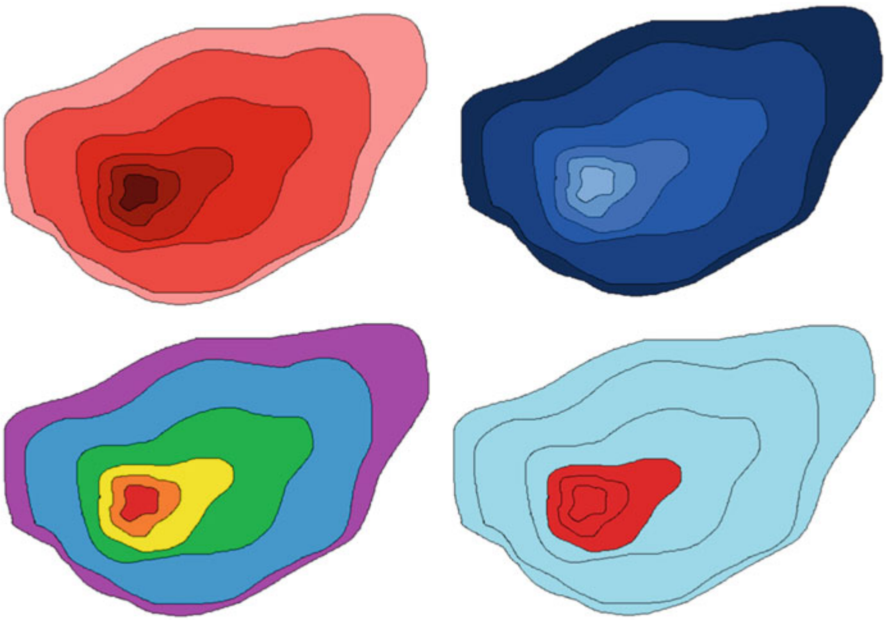

Fig. 14.2 Examples of color being used to represent the uncertainty in some dataset. In the two examples on top, saturation is used to represent uncertainty, but in opposite ways. In the top left, more saturated colors represent higher certainty, while in the top right, more saturated colors represent higher uncertainty. In the bottom-left example, each level of uncertainty is assigned to a different hue. In the bottom-right example, levels of uncertainty lower than some threshold are assigned to one hue, and levels of uncertainty higher than that threshold are assigned to another. This creates strong perceptual boundaries that may lead participants to treat things inside the boundary as being categorically different from things outside of the boundary, even if that is not the case

with more saturated colors). Several researchers have argued that this is an intuitive mapping, but some experimental results suggest otherwise. Across several studies, participants reported a preference for visualizations in which uncertainty was mapped to color saturation. However, those same participants performed worse on decision-making tasks when uncertainty was mapped to color saturation than when other representations of uncertainty were used [34, 99, 113]. This indicates that color saturation was not a particularly helpful index of uncertainty even though the participants felt that they had an intuitive understanding of it.

However, it is possible to use color to draw viewers' attention to information about uncertainty. Some recent work has developed color palettes that make it difficult for viewers to determine the precise value of uncertain data points [25]. The researchers demonstrated that these color palettes led viewers to give more consideration to the uncertainty of the data when making decisions. In this case, it is not necessarily the color saturation that is drawing the participants' attention to the uncertainty, but rather the difficulty of identifying the values of specific data points.

Other issues with color mappings can arise when people interpret them in unexpected ways. Recent work has shown that people exhibit a bias toward expecting darker colors to represent larger quantities. This bias runs counter to many common color maps, which typically use brighter values to represent higher values [138]. This could lead to misinterpretations in the context of uncertainty visualization, where a mapping between hue and uncertainty may seem intuitive to the visualization designer, but may be interpreted in a different way by viewers.

Finally, mapping uncertainty to a small set of colors can produce undesirable perceptual effects, such as creating visually salient boundaries that emphasize one particular change in probability while obscuring other, equally important changes [126]. Thus, while mapping uncertainty to manipulations of color appeals to many visualization designers and viewers, the research to date indicates that these mappings may produce perceptual biases that are not yet well-understood.

For any intrinsic representation of uncertainty, it is important to note that some visual metaphors may be better suited to specific domains than others. For example, using blur to depict uncertainty about the position of a moving vehicle maps well to human perception of motion [132], but this technique cannot be applied to particle movement datasets, which are often so large that blur would be imperceptible [56]. To give another example, size is a graphical element that can mean different things in different domains. In scientific and geospatial visualizations, data with more uncertainty often map to larger graphical elements. In the case of 3D flow visualizations, this might take the form of a larger envelope of uncertainty [17], and in the case of land use maps, this might take the form of smaller squares in areas where the data are more precise and larger squares for data that are less precise [34]. However, in more abstract visualizations, larger elements might represent information that has higher certainty, while less certain information is minimized.

14.5.2 Extrinsic Representations of Uncertainty: Adding Graphical Elements

Perhaps the best-known way to depict uncertainty in data visualizations is through the addition of error bars. These can be used in one-dimensional representations, such as statistical graphs [121], and they can also be extended to two- or three-dimensional datasets. For example, in a 2D dataset, error bars can be depicted as an ellipse or cone around the point of interest [70]. In 3D datasets, error is often represented by a transparent isosurface or surrounding a more solid shape (cf. [17]). Error bars have the advantage of being widely used, but they are often misunderstood, even by domain experts [3, 24, 101, 112]. The use of error bars can also imply that the data have a normal distribution even when that is not the case [124]. Alternative approaches that address this concern include plotting the probability distribution function (e.g., [78]) or providing a summary of it with a box plot [104] or a five-number summary [128]. Distributional visualizations such as quintile dot plots, which give information about the shape of a distribution while also allowing viewers to compute the probability of different outcomes, have been shown to be particularly effective for helping people to understand risk [42, 78, 116]. Distributional visualizations expose aspects of the data that would be hidden by simpler representations, such as bar charts with error bars.

Labels, numbers, and glyphs can also be added directly to a visualization to represent uncertainty. Numbers or labels can directly convey information such as probability [6, 43], while glyphs can convey uncertainty through their shape, color, or sharpness [6, 70, 121]. While this approach may be highly effective for small datasets, it may not scale well to larger datasets, where the added information may obscure other parts of the visualization, create clutter, or lead to unintended perceptual effects [21, 121]. Similarly, overlays can be used to represent uncertainty over large sections of a visualization. These can be particularly useful for spatial data [80, 81, 98], but they can also obscure features of interest in the data.

Uncertainty information can be added in the form of non-visual information, such as auditory or tactile signals [2]. However, there is relatively little research on multi-modal representations of uncertainty, and the utility and effectiveness of this approach are unclear.

14.5.3 Creating Multiple Visualizations

Uncertainty can also be conveyed through the use of multiple visualizations. Common approaches include presenting the data and the uncertainty side-by-side, allowing users to toggle between different representations, using flickering or blinking in which the uncertainty visualization appears as a transient overlay on top of the data, or creating animations of the data. Side-by-side comparisons are relatively common for geospatial data [36, 80]. The drawback of this type of display is that subtle differences can be difficult to detect. Flickering or blinking displays will draw viewers' attention to differences via illusory motion, but users may find this to be distracting or annoying [39, 107]. Allowing users to toggle between the visualization of the data and visualizations of uncertainty can provide a compromise between these two approaches, but users may not understand how to toggle between the displays or may simply choose not to, ignoring the uncertainty information [39]. Finally, animations can provide a comprehensive picture of a dataset and its associated uncertainty, but they can place a high burden on the viewer's cognitive resources, particularly if the viewer must remember and integrate information from different parts of the animation [94].

Another approach to conveying uncertainty is the use of hypothetical outcome plots (HOPs). HOPs use animation to show different samples from a distribution in a sequence of frames [77]. Research on HOPs has shown that they allow viewers to estimate the probabilities of different events, often leading to better estimates than other types of visualizations [62, 67, 77]. Although this approach seems promising, HOPs may introduce sampling error if viewers only see a subset of samples from the distribution (or if they stop watching before the animation is complete), and they may struggle to integrate some types of information across the different frames [67, 152]. For example, in a network visualization task, participants had a hard time identifying distinct communities in the network when they had to integrate that information across multiple realizations of a network [152]. HOPs have only been tested for a relatively small number of tasks to date, so there is not yet a clear consensus on the types of tasks that are well-supported by HOPs [67].

14.5.4 Summary

For each of the techniques outlined in this section, there is an orthogonal question regarding how much the viewer's attention should be drawn to the uncertainty in the data as opposed to the data itself. This brings us back to the questions about *why* users need information about uncertainty and *how* it will impact their task. Do they simply need to be aware of the existence of uncertainty? Or do they need to know the type or amount of uncertainty? Is information about uncertainty likely to produce cognitive biases, and are these desirable? In some cases, it may be desirable to emphasize the uncertainty so that viewers are sure to notice its presence or incorporate it into their decision-making [25]. In many domains, such as intelligence analysis or medicine, the costs of different types of errors are asymmetric, so exaggeration of uncertainty and a consequently biased reasoning strategy may be beneficial [58, 59].

Just as each of the methods for visualizing uncertainty has specific strengths and weaknesses in terms of human perception and comprehension, they also vary in their compatibility with different types of data and different types of uncertainty. Examples of how these methods are used in different domains and for different types of uncertainty are presented in the section below, along with information about which methods were most effective in cases where different methods have been compared directly.

14.6 Applications of Uncertainty Visualization Techniques in Different Domains

In this section, we dive into more detail by comparing ways in which some of the techniques discussed in Sect. 14.5 have been applied in different domains. The efforts to evaluate the effectiveness of uncertainty visualizations for different data types and different domains have been very uneven. For example, there has been a substantial effort to formalize approaches to uncertainty visualization for geospatial data, and there have been prior reviews of the literature in this area, notably by Drecki [34], MacEachren and colleagues [99], and by Kinkeldey and colleagues [81] and [79]. Another very useful review that addresses statistical graphs and infographics is provided by Franconeri and colleagues [45]. However, in other domains, such as scientific visualization, there have been few attempts to evaluate the effectiveness of different visualization techniques, much less to review and consolidate information about how different techniques might impact comprehension and decision-making.

In this section, we attempt to address this unevenness by discussing uncertainty visualization techniques that have been well-studied in one domain while going unevaluated in others. Our goal is to identify areas where there are gaps in the existing literature that could be addressed by future research. In particular, our goal is to identify instances where findings from one domain set up testable hypotheses in other domains where the generality of different findings could be explored.

14.6.1 Intrinsic Representations of Uncertainty

14.6.1.1 Hue

As discussed in Sect. 14.5, color is one of the most common ways of representing uncertainty in data, so it is also one of the techniques that has been most widely studied. However, it has not been thoroughly evaluated across all domains. There have been multiple studies that evaluate visualizations that use hue to encode uncertainty for geospatial data, but almost none that have evaluated the effectiveness of similar encodings for 3D spatial data or spatiotemporal data. Yet the findings from the geospatial domain suggest that color coding is not always as intuitive as it may seem, which could have important ramifications for visualizations in those other domains.

In many geospatial applications, hue is used to encode aspects of the data, so the saturation of the colors is a salient channel for indicating uncertainty [80].

Although this approach is widely used and participants believe it to be intuitive, direct comparisons have shown that color saturation is often less effective than other representations of uncertainty, such as transparency [80, 100]. A study of 2D spatial data from the scientific visualization domain also found that mapping uncertainty to color could produce undesirable perceptual effects. A two-color mapping for different levels of probability within a visualization produced visually salient boundaries that emphasized one particular change in probability while obscuring other, equally important changes [126]. These studies indicate that there are drawbacks to using color hue or saturation to encode information about uncertainty.

Despite these findings for geospatial data and 2D scientific data, color is still widely used to encode uncertainty in 3D spatial visualizations (cf. [11, 56, 90, 93]). Representing uncertainty in 3D visualizations creates additional challenges because of the increased visual complexity of the data and the increased potential for obscuring part of the visualization (cf. [123]). Praßni et al. [126] point out that 2D visualizations can depict details about uncertainty, whereas 3D visualizations are better suited to provide an overview of which regions of the volume are uncertain. Similarly, Potter and colleagues [125] note that uncertainty visualization techniques used for 3D data typically show the location and relative size of the uncertainty instead of more detailed information.

Given the difficulty of visualizing uncertainty in three dimensions, many researchers have mapped uncertainty information to hue, saturation, and/or brightness [22, 55, 95, 122, 131], operating under the assumption that color mappings are highly intuitive. However, there have been few studies to test the effectiveness of these encodings for 3D spatial data. Many of the studies on uncertainty visualizations for scientific data propose different approaches but do not test the impact of those approaches on comprehension or cognition. This is a gap in the current literature. Given that color mapping can produce undesirable perceptual effects in 2D visualizations, such as spurious boundaries [126], the same may be true for 3D visualizations, but this question and related questions about how color mappings impact viewers' perception and understanding of 3D visualizations have not been adequately explored.

14.6.1.2 Transparency and Texture

A less widely used, but potentially more effective alternative for conveying uncertainty in geospatial data is the use of transparency. Several studies have shown that manipulations of transparency and texture attributes have been shown to be effective for conveying uncertainty in geospatial data [34, 87, 89, 113]. Influential work in this domain has advocated for using the fog metaphor, implemented via manipulations of hue and edge crispness [99]. When using the fog metaphor, areas with higher certainty have more saturated colors and crisp edges, making them appear bright and clear. Meanwhile, areas with uncertain data use less saturated colors and blurred edges, making them look as if they are obscured by fog. This metaphor draws on naturalistic cues, and it rated by viewers as being highly intuitive [100].

Manipulations of transparency have also been used for 3D visualizations. A common application of this approach is the use of fuzzy boundaries to represent uncertainty about the spatial extent of some object or phenomenon (cf. [44, 93]). However, the visual metaphors used for 3D visualizations are often the opposite of those used for 2D visualizations. For example, instead of making uncertain areas appear to be foggy and visually obscured in 3D visualizations, it is more common for areas with high uncertainty to be represented by increased transparency, making them less visible altogether. Using transparency to encode uncertainty can be particularly useful in 3D visualizations, because increased transparency can reduce occlusion [94, 130].

User studies focused on specific domains have indicated that transparency manipulations can improve viewers' ability to accurately segment images [126] and to determine whether a marker is inside or outside of the error margin for a surface [53]. However, transparency does not provide precise quantitative information [94], which can impact viewers' understanding of the uncertainty. In addition, manipulating the transparency of uncertain data points or regions is often perceived as a change in texture when there is other data behind those regions in the 3D space [52, 53, 126]. This can produce undesirable perceptual effects. For example, Djurcilov and colleagues experimented with mapping the uncertainty for variables of interest to transparency. This mapping produced a speckly or textured appearance for regions of the visualization where the data were more uncertain [30, 31]. They found that these discontinuities in opacity could be misleading, as many small discontinuities in one region of the visualization could be perceived as a change in color rather than a change in transparency or texture. In addition, it was difficult for people to distinguish more than a few levels of uncertainty in this scenario.

Manipulations of transparency and texture can also be used to produce a blurring effect [10, 115]. Several prior studies have found that blur is an intuitive representation of uncertainty [12, 98], although it is difficult to quantify the difference between different levels of blurriness [12, 85]. Blur might be a useful technique for reducing some of the biases invoked by visual boundaries. As discussed earlier in this chapter, sharp visual boundaries can cause people to treat things inside of a boundary as being categorically different from things outside of the boundary, even when this is not actually the case or when the placement of the boundary is somewhat arbitrary [117]. It is reasonable to hypothesize that blurring may help to reduce this effect, but this hypothesis has not yet been thoroughly tested across different tasks and data types.

Blur is also a naturalistic metaphor for motion. For example, Roessing and colleagues [132] developed a framework to enhance the user's ability to make assessments of motion, speed, and distance of approaching cars in a rearview camera. They found that artificial motion blur supported more accurate assessments of speed and distance, aiding safer lane changes. In contrast, there was not a significant benefit from a visualization of risk potential in which a color overlay indicated the estimated time to impact for the approaching car. This provides an example where a more natural visual mapping (more blur for increased speed instead of color coding based on risk) supported more accurate interpretations.

14.6.1.3 Summary

In summary, although color coding is often perceived to be an intuitive representation of uncertainty, it often falls short when compared to other intrinsic representations. Most comparisons of this type have been done using 2D spatial data. It seems likely that similar patterns would hold for 3D spatial data, but there have been very few direct comparisons between different uncertainty visualization techniques for 3D data. Many 3D spatial visualizations use color to encode uncertainty under the assumption that this will be an intuitive representation but fail to test that assumption. The lack of research identifying effective representations of uncertainty for 3D data is an important gap in the existing literature. This question will become increasingly important as visualizations that incorporate augmented or virtual reality technology become more common.

Manipulations of transparency or texture that use fog or blurriness to convey uncertainty tend to perform well for 2D spatial visualizations and for representations of motion. However, these techniques may produce undesirable perceptual effects in 3D datasets. In addition, fog and blur are not appropriate metaphors for all types data. The fog metaphor might be misleading in the context of weather data, for example. In addition, some datasets are so large that blurring of certain data points would be imperceptible [56]. However, given the relative success of these metaphors relative to other intrinsic representations of uncertainty, they warrant further research. It would be useful to develop a systematic understanding of when the fog and blur metaphors support human comprehension and when they do not, across a wider variety of data types and domains.

14.6.2 Extrinsic Representations of Uncertainty

As discussed in Sect. 14.5, the mostly commonly used extrinsic representations of uncertainty are glyphs, and the most common glyphs for representing uncertainty are error bars. Bar charts with error bars are one of the most common methods of data visualization in scientific papers. Although error bars are widely used, even experienced scientists have difficulty with interpreting them correctly. Several studies have demonstrated that researchers who produce and consume statistical graphs frequently misinterpret the relationship between error bars and statistical significance as well as the meaning of different types of error bars, such as confidence interval (CI) and standard error (SE) bars [3, 101].

Several studies have found that there are design tradeoffs for statistical graphs because the visual encodings chosen for the mean and error in a dataset change viewers' interpretation of the data. Divis and colleagues [29] found that participants were overly generous with SE error bars, interpreting the difference between two datasets as significant even when it was not. In contrast, the participants were overly conservative with CI error bars, interpreting significant differences as if they were not significant. Several studies have found that bar graphs cause cognitive biases even when they do not have error bars. Participants who made decisions based on bar graphs consistently demonstrate "within-the-bar bias," interpreting values within the bar as being more likely than values outside of the bar [24, 112].

However, other work has found that participants who have experience with producing and interpreting statistical graphs have better performance when using error bars relative to using less familiar plot types, such as violin plots [101]. While these experienced participants exhibited consistent biases in how they interpreted the error bars, they were faster, more accurate, and more confident in their assessments of the datasets when using error bars than when using any other type of visualization. These findings indicate that familiarity and experience with visual encodings of uncertainty play an important role in their effectiveness (see also [57]).

For some types of data, error bars may provide an especially effective visual metaphor. In one of the few studies to study visualizations of temporal uncertainty, Gschwandtner and colleagues [54] used a variety of representations, including gradient plots, violin plots, accumulated probability plots, and error bars, to represent the start and end times of intervals. This study found that participants performed best when using error bars and variants of error bars, such as using a lighter hue for the uncertain regions of the intervals. In this case, the visual cues provided by the error bars may have been the best match for the cognitive demands of the task. In addition, the manipulation of hue might reduce some of the known biases that are associated with error bars by reducing the salience of the visual boundaries that error bars can create.

Glyphs are also widely used in the domain of weather forecasting, where there is also evidence that experience plays a role in how people interpret uncertainty visualizations. For novice viewers, different representations of uncertainty produce different patterns of biases in novice viewers. When viewing hurricane forecasts, participants shown the typical "cone of uncertainty" that often appears in media coverage of hurricanes believed that the cone represented an increase in the strength and size of the hurricane over time, rather than information about the uncertainty of the hurricane's path [120, 134]. People also tend to think that areas outside of the visual boundary created by the cone visualization are categorically different than areas inside of the boundary, producing biases in their assessments of risk [117]. When shown ensemble displays, or "spaghetti plots" in which multiple possible hurricane paths are each represented with a single line, participants believed that the storm was less intense when the lines were farther apart. They also believed that locations "touched" by one of the lines in the ensemble display would receive more damage from the storm [120]. In both cases, visually salient features of the uncertainty visualizations were interpreted in unintended ways by the viewers. The visual–spatial biases revealed by these studies show the value of using rigorous cognitive science methods to study the impact of different visual representations on human biases and decision-making.

Glyphs have also been used to convey uncertainty in 3D visualizations. They have the benefit of condensing information so that more information can be displayed while minimizing the occlusion of other features of the visualization [61, 70, 93, 94, 96, 113, 153]. Glyphs have been shown to have advantages over other representations of uncertainty in some studies. Vector glyphs (arrows) can be particularly useful because they provide an intuitive representation of flow, such as the movement of air or water. While the arrows show the direction and magnitude of flow, features such as their width or color can be used to depict measurement uncertainty [148]. In this case, a user study indicated that expert participants, who had prior experience with interpreting glyphs, were equally successful in terms of interpreting the glyphs themselves and interpreting the uncertainty represented by manipulating the glyphs.

Newman and Lee [113] asked participants to rate their perception of different types of glyphs as well as other representations of uncertainty, such as transparency and color coding. The glyphs included cylinders, cones, balls, arrows, and multi-point glyphs, where the uncertainty was encoded by one aspect of the geometry of each type of glyph. For example, the local uncertainty was represented by the height of the cylinder glyphs, the base diameter of the cone glyphs, or by the radius of the ball glyphs. The participants' ratings were strongly dependent on the type of glyph used. They preferred the glyphs over color coding and transparency when asked how easy it was to identify the information about uncertainty in the visualization. However, some glyphs received poor ratings when the participants were asked about the ease of identifying the data (rather than the uncertainty) and the visual clutter in the images. Essentially, the glyphs made the uncertainty information highly salient but obscured some of the data itself. This study provides another illustration of the point that glyphs that do not naturally map to the information in question might be more distracting than helpful.

Lodha and colleagues [95] found that glyphs helped viewers to understand the differences between two overlaid surfaces. In this case, glyphs were used to fill the space between the two surfaces, highlighting where the location and relative magnitude of the differences. They found this technique to be more effective than other methods, including color coding, transparency, and texture manipulations.

14.6.2.1 Summary

Glyphs may have advantages over other representations of uncertainty in some scenarios, particularly for 3D visualizations. However, as with intrinsic representations of uncertainty, using glyphs that provide an appropriate visual metaphor for the domain may be important. This point has received relatively little attention and warrants further research. Are there ways to manipulate the visual attributes of glyphs in ways that make them more intuitive or reduce cognitive biases? We return to this point in the section on statistical graphs, below.

Importantly, experience may play a critical role in how well viewers are able to interpret glyphs. Many of the studies that have found advantages for using glyphs used participants with domain experience [101], while those that used novice participants have often found that glyphs are misinterpreted [120, 134]. On the other hand, numerous studies have shown that even experienced researchers fail to interpret error bars correctly, so there are clearly limits to the benefits of experience when interpreting glyphs [3, 24, 101]. The balance between familiarity and the potential for misinterpretation is a tricky one, but drawing more heavily on the cognitive science literature to investigate this issue could be useful for developing principled approaches to designing glyphs. For example, can we use information about visual perception to identify cases where glyphs are likely to be misleading? Can training on how to interpret unfamiliar glyphs overcome the advantages of familiar glyphs, and if so, how much training is necessary?

14.6.3 Multiple Visualizations

As discussed in Sect. 14.5, one approach to representing uncertainty is the creation of multiple visualizations. Once again, this is a technique that has primarily been tested in the geospatial domain. In GIS data and other types of maps, it can be particularly difficult to distinguish the data from the uncertainty information and to visualize the uncertainty information in a way that does not occlude the data. Researchers in this area have tested a variety of approaches to this problem, including side-by-side visualizations, overlays, toggling, flickering, and animation (see [81] for a review). While some studies have found benefits to dynamic displays that use flickering or animation [39, 99], participants often prefer static displays [1] and find flickering or animation to be annoying or too difficult to understand [7, 107]. Animations may increase the participants' cognitive load if they are forced to remember and integrate information from different time points in the animation. This can be detrimental to task performance, as demonstrated by multiple studies that have found better performance for static displays than for animations [81].

Animations can also be more distracting than they are useful for understanding three-dimensional data [95]. However, at least one study has found that animations supported increased speed and accuracy for radiologists assessing volume renderings in a medical decision-making context [97]. Lodha and colleagues [95] also tested side-by-side views, although in this case the two views were showing different surfaces rather than a visualization of the data and a visualization of the uncertainty. They found that side-by-side views were adequate for identifying large differences between surfaces, but not for more subtle differences.

Static side-by-side displays may be one of the more effective ways of representing uncertainty when occlusion is a concern. In the geospatial domain, side-by-side displays often show some data in one image and the uncertainty associated with the data in an adjacent image (cf. [1]). For many tasks, participants perform equally well when using side-by-side views as when using coincident (overlaid) views of data and uncertainty [1]. However, there is also evidence that people can have trouble integrating the data and the uncertainty if the uncertainty is presented in a legend or in an adjacent image. Integrating information across multiple images or text-image combinations places a burden on the viewer and his or her cognitive load, increasing the chances that the uncertainty information will be misinterpreted or ignored [36].

For 3D data, side-by-side displays have been used to show errors produced by integration algorithms in fluid flow models. This approach is effective if the errors are large, but subtle errors are much harder to detect when comparing two images [16, 95]. In this case, superimposing the images has been shown to work better, although that raises the issue of occlusion once again (Lodha et al., 1996).

Interactivity can provide some of the benefits of multiple visualizations while mitigating some of the drawbacks. For example, viewers who are annoyed by a display that flickers between two or more different representations might be appeased if they can control when the information about uncertainty appears and disappears. However, when people are given the option of how to interact with the visualization, they may choose not to view the information about uncertainty at all, or they may struggle to manage the increased complexity of the visualization tool [9, 21]. Although interactivity has considerable promise for helping people to better understand uncertainty, there have been surprisingly few studies on how to design interactive visualizations that effectively convey information about uncertainty. What types of tasks and datasets benefit from interactivity and which do not? Are there methods of introducing interactivity that encourage people to grapple with uncertainty rather than ignoring it? How do the pros and cons of interactivity relate to human cognitive processing? Can we use aspects of cognition such as the limits of attention and working memory capacity to develop a scaffold for principled approaches to incorporating interactivity?

14.6.4 Statistical Graphs

Although we included some discussion of statistical graphs earlier in this section, they have some unique properties that warrant additional attention. Statistical graphs provide information about a dataset, including information about the variability in the data. For example, scientific papers commonly show the trends and variability in datasets using visualizations such as scatterplots, bar plots, box plots, violin plots, or line plots. Since these statistical graphs are so widely used, they have also been subject to a great deal of research investigating how different representations of the same underlying dataset can impact human comprehension.

Statistical graphs are more abstract than many of the visualization types discussed above, so the forms that they can take are less constrained. They may represent multi-dimensional data that cannot easily be visualized in 2D or 3D space without the use of dimension reduction techniques. Thus, there is rarely an obvious visual metaphor that relates the data shown in statistical graphs to things that people are familiar with in the real world.

When a dataset has three or fewer dimensions and a relatively small number of data points, scatterplots or dot plots can be used typically to show the variability of the data directly, with a symbol to represent each data point (cf. [84, 147]). This type of direct representation avoids many potential sources of cognitive bias. The symbols used to represent each data point can provide information about the uncertainty of that information. For example, imputed data, instances where missing data are replaced with a reasonable estimate, can be plotted using different symbols than data that were measured directly [84]. Cognitive studies have shown that people can quickly integrate and make judgments about scatterplots, even when multiple datasets are shown in one figure [20, 47, 129]. They can rapidly extract correlations and trends [32, 88, 106, 129] and make comparisons between different datasets or different classes of data points [47, 91, 92].

However, scatterplots are subject to cognitive biases of their own. Viewers consistently underestimate the correlation between variables [129], and they often use simple heuristics to make comparisons between datasets, such as judging the cluster with the highest overall point to have the highest mean value [29, 47]. Changes in the visual representation of a scatterplot can alter viewers' estimates of the correlation between variables [32, 88], and changes in the size, shape, and density of clusters can impact viewers' judgments about the clusters [38]. Furthermore, scatterplots are not a feasible visualization for large datasets or for datasets in which overplotting (multiple data points plotted in the same location) can obscure portions of the dataset. While there are a number of techniques to reduce visual clutter when large datasets are visualized (cf. [37, 136]), those techniques move away from direct representations of the data, introducing new layers of abstraction. This raises some of the same design challenges that are common for uncertainty visualizations: how should designers select abstractions that support human cognition without introducing confusion or biases?

Other types of statistical graphs provide methods for abstracting away from direct representations of the data.

Box plots were developed to provide a summary of a dataset at a glance, including its median, upper and lower quartiles, upper and lower extremes, and outliers [104]. An important drawback of traditional box plots is that datasets with very different distributions, such as a normal distribution and a bimodal distribution, can produce box plots that look identical [19]. Several researchers have proposed variants of the box plot that use color, shape, or shading to indicate the density of the underlying distribution [4, 19]. Similar approaches have been developed for bivariate box plots (cf. [133]). Violin plots, which encode the distribution of the data via the width of the bar [60], have been the most widely adopted variant of the box plot. See Fig. 14.3 for examples.

Correll and Gleicher [24] found that changing the visual encodings of statistical graphs could improve viewer performance, even for members of the general public with no background in statistics. They recommend using encodings that are visually symmetric, such as violin plots, and visually continuous, such as gradient plots. These types of visual encodings reduce the cognitive biases that have been observed for bar plots and error bars.

These techniques tie back to the implicit representations of uncertainty discussed above. For example, gradient plots manipulate visual attributes of the visualization,

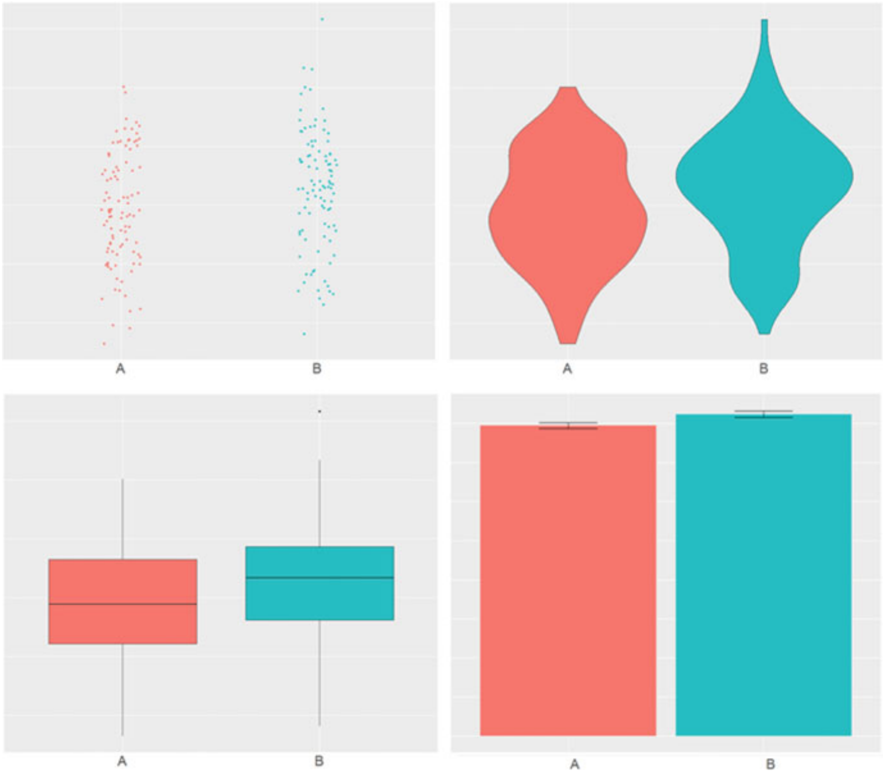

Fig. 14.3 The same data plotted as a scatterplot (top left), violin plot (top right), box plot (bottom left), and bar chart with error bars showing the mean standard error (bottom right)

such as transparency, based on the statistical distribution of the underlying dataset. Other techniques, such as kernel density estimation, can be used to generate density plots that manipulate intrinsic visualization variables to emphasize the most certain information and de-emphasize the most uncertain information [41, 84].

As discussed above, error bars are extrinsic representations of uncertainty that are widely used in statistical graphs. While error bars have disadvantages, they have the advantage of familiarity, which can lead to better performance than manipulations of intrinsic attributes [29]. In the case of statistical graphs, the familiarity of the representation and the effectiveness of the visual metaphor being deployed are often in conflict. This is another area that warrants continued research. In particular, are there encodings that consistently support better comprehension of uncertainty for both scientists and lay people? Can training on newer methods of representing uncertainty in statistical graphs overcome the effect of familiarity for people who have considerable experience with creating and consuming bar charts?

Finally, research questions related to the use of multiple visualizations also apply to statistical graphs. For many years, there have been calls to focus on the development of interactive statistical graphs (cf. [141]). Interactive visualizations have become relatively common in the media and can be very useful for guiding viewers to comparisons of interest (cf. [45]). However, interactive infographics or statistical graphs are still fairly rare in scientific publications. Printed papers must have static graphs, but online scientific communications could incorporate interactivity. What lessons can we draw from the use of interactive visualizations on news websites to apply to scientific communication, for both communications with peers and communications with the general public?

In the realm of statistical graphs, hypothetical outcome plots (HOPs) represent a variant of the approach of using multiple visualizations to convey uncertainty or probability. As discussed above, HOPs have shown a great deal of promise for helping viewers to understand the variability or uncertainty in a dataset [117]. HOPs draw from a probability distribution multiple times and visualize each draw. Hullman et al. [67] showed that HOPs supported better reasoning about distributions than more abstract representations such as error bars and violin plots. Similarly, Kale and colleagues [77] showed that participants were better able to infer the underlying trend in a dataset when using HOPs than when using error bars or line ensembles. However, there were some circumstances in which participants struggled when using the HOPs, such as when estimating the mean of a dataset with high variance [67]. While dynamic displays of uncertainty in datasets can mitigate the cognitive biases that are induced by static representations, more research will be needed to understand the circumstances under which animated HOPs may produce biases of their own.

An interesting direction for future research would be the application of a HOPs-like approach to other types of data. There has been some work along these lines, such as developing HOPs that show different instantiations of a network distribution that has probabilistic edges [152]. However, the HOPs approach to representing uncertainty is different from the approaches that have typically been used for geospatial or flow data. HOPs have typically been used to represent the frequency of different outcomes, while related approaches for spatial data, such as flickering, animation, and side-by-side displays, typically show one visual representation of the data and another showing which aspects of the data have higher uncertainty than others. These two different approaches to showing uncertainty through multiple visualizations would benefit from direct comparisons between then. For example, can a HOPs-like approach be developed for geospatial data or 3D scientific data? Would that improve comprehension of uncertainty relative to other techniques that have been used in these domains? Or is the type of metaphor used by HOPs a poor fit for spatial data? These would be fruitful areas for future research.

14.7 Discussion

Uncertainty visualizations present thorny problems for designers and viewers, and data visualization researchers do not yet have a good grasp on which techniques are most effective and why. Researchers are struggling to make sense of the relatively small number of rigorous evaluation studies that have been done in this area, with the result the same groups of studies have appeared in review papers and taxonomies over and over [8, 15, 80, 99, 121, 125]. The present paper is yet another attempt at making sense of this literature and identifying common threads across domains. Recent work bridging visualization and cognitive science has begun to make more progress in elucidating *why* some uncertainty visualization techniques are more effective than others [45, 117]. In this vein, we have attempted to identify the uncertainty visualization techniques that are used most commonly for different types of data, which methods have been compared with one another, and which were found to best support human decision-making.

Several common threads have emerged from this effort. First, and most importantly, the studies reviewed indicate that visual representations of uncertainty often improve comprehension and decision-making performance, so long as the way in which the uncertainty is expressed is a good fit to the participants' task (cf. [72]). Of course, we do not know how many unpublished experiments have found that visualizations of uncertainty *harmed* human performance, but the fact that there are published studies across many domains and data types that found benefits to visualizing uncertainty indicates that it can be done successfully.

The studies that demonstrated success point to a few important themes. First, the participants' task matters. Visualizations that performed well provided support for the participants' cognitive processes without overloading their cognitive resources. Although participants may prefer to ignore uncertainty information, visualizations can help people to understand uncertainty better and can push them to take uncertainty into account. The difficulty of the participants' task matters as well. When the task is relatively easy, the way in which the uncertainty is represented may have little or no impact on task performance (cf. [6, 18, 82]). However, when the task becomes more difficult, the match between the task and the representation becomes increasingly important.

Second, the participants' experience matters. If a particular visual metaphor is common in a particular domain, it may be best to stick with that metaphor so as not to confuse domain experts. In domains where there are standard representations of error that are widely used (such as error bars in scientific publications) or where efforts have been made to formalize visualization approaches (such as GIS and cartography), paying attention to the existing conventions is particularly important. While newer visualization techniques might show better performance for novice viewers, they may have negative impacts on expert viewers and vice versa [7, 101]. However, as discussed in Sect. 14.6, there are cases (such as error bars) where the dominant visual metaphor for uncertainty is also widely misunderstood and misinterpreted, even by domain experts. This finding opens the door for additional research at the intersection of visualization and cognition to help find ways to address this problem, either through modification of visual cues or through training.

Third, if a dataset is amenable to naturalistic metaphors for conveying uncertainty, such as the metaphor of fogginess, those metaphors can be highly effective. However, visualization designers must ensure that the metaphor is implemented in a way that matches the viewers' expectations [99, 138]. An important sidenote is that viewer preferences do not necessarily correspond to their performance, and mappings that viewers believe to be intuitive do not necessarily support better performance (cf. [34, 99, 113]). More empirical research is needed to explore the relationships between manipulations of visual cues and viewers' interpretations of those cues across different tasks and data types. Specifically, we need more studies that use objective measures of task performance rather than just gathering viewers' subjective opinions about different methods of representing uncertainty.

Fourth, many of the studies reviewed here suggest that intrinsic mappings of uncertainty are often better for providing an overview of the uncertainty, which extrinsic mappings are better for providing details. In particular, glyphs can provide a great deal of detail, which might be preferable for domain experts or decision-makers who need more than just an overview. However, it is also crucial that viewers are able to distinguish the data from the uncertainty and that their view of the data is not obscured by the information about uncertainty. While side-by-side images, animations, and dynamic displays can all help to address this problem, all can contribute to an increase in cognitive load for viewers.

Finally, visualization designers need to be aware of unintentional perceptual effects that can arise in uncertainty visualizations, such as salient boundaries [126]. The presence of a visual boundary, whether intentional or unintentional, impacts cognition [117] and should be used with care.

Thinking about data visualizations from the perspective of how well they support human cognition and decision-making has advantages for both designers and researchers. Commonalities in human perception and attention can provide insights into why some visualization designs are more effective than others, allowing for more generalizable evaluation strategies. There are also many unanswered questions about the impact of different visualization techniques on human cognition. This provides a wealth of opportunities for cognition researchers to advance our understanding of visual cognition while also having an impact on real-world problems. There are a growing number of collaborations between cognitive scientists and visualization researchers, adding to the growing body of literature in visualization psychology. Visualizations of uncertainty provide a particularly rich environment for these types of collaborations. There are many unanswered questions about how people process and make sense of information about uncertainty, and many domains where effective visualizations of uncertainty are needed. Although some uncertainty visualization techniques have been compared and evaluated, there is much work left to be done.

Acknowledgments This chapter describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. This work was funded by Sandia National Laboratories' Laboratory-Directed Research and Development program. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

1. J. C. Aerts, K. C. Clarke, and A. D. Keuper. Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science*, 30(3):249–261, 2003.
2. S. Basapur, A. M. Bisantz, and T. Kesavadas. The effect of display modality on decision-making with uncertainty. *Proc. Human Factors and Ergonomics Society Annual Meeting*, 47(3):558–561, 2003.
3. S. Belia, F. Fidler, J. Williams, and G. Cumming. Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, 10(4):389, 2005.
4. Y. Benjamini. Opening the box of a boxplot. *The American Statistician*, 42(4):257–262, 1988.
5. A. M. Bisantz, D. Cao, M. Jenkins, P. R. Pennathur, M. Farry, E. Roth, and J. Pfautz. Comparing uncertainty visualizations for a dynamic decision-making task. *Journal of Cognitive Engineering and Decision Making*, 5(3):277–293, 2011.
6. A. M. Bisantz, S. S. Marsiglio, and J. Munch. Displaying uncertainty: Investigating the effects of display format and specificity. *Human Factors*, 47(4):777–796, 2005.
7. S. Blenkinsop, P. Fisher, L. Bastin, and J. Wood. Evaluating the perception of uncertainty in alternative visualization strategies. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 37(1):1–14, 2000.
8. G. P. Bonneau, H. C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz. Overview and state-of-the-art of uncertainty visualization. In *Scientific Visualization*, pages 3–27. Springer, London, 2014.
9. A. Bostrom, L. Anselin, and J. Farris. Visualizing seismic risk and uncertainty: A review of related research. *Annals of the New York Academy of Sciences*, 1128(1):29–40, 2008.
10. R. P. Botchen, D. Weiskopf, and T. Ertl. Texture-based visualization of uncertainty in flow fields. In *Proc. IEEE Visualization (VIS 05)*, pages 647–654. IEEE, 2005.
11. R. P. Botchen, D. Weiskopf, and T. Ertl. Interactive visualization of uncertainty in flow fields using texture-based techniques. In *Proc. Intl. Symp. on Flow Visualization*, volume 2, 2006.
12. N. Boukhelifa, A. Bezerianos, T. Isenberg, and J. D. Fekete. Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2769–2778, 2012.
13. M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.
14. S. Bresciani and M. J. Eppler. The pitfalls of visual representations: A review and classification of common errors made while designing and interpreting visualizations. *Sage Open*, 5(4):2158244015611451, 2015.
15. K. Brodlie, R. Allendes Osorio, and A. Lopes. A review of uncertainty in data visualization. In *Expanding the Frontiers of Visual Analytics and Visualization*, pages 81–109, 2012.
16. P. G. Buning. Sources of error in the graphical analysis of CFD results. *Journal of Scientific Computing*, 3(2):149–164, 1988.
17. K. Bürger, R. Fraedrich, D. Merhof, and R. Westermann. Instant visitation maps for interactive visualization of uncertain particle trajectories. In *Proc. Visualization and Data Analysis*, volume 8294, page 82940P, 2012.
18. L. Cheong, S. Bleisch, A. Kealy, K. Tolhurst, T. Wilkening, and M. Duckham. Evaluating the impact of visualization of wildfire hazard upon decision-making under uncertainty. *International Journal of Geographical Information Science*, 30(7):1377–1404, 2016.
19. C. Choonpradub and D. McNeil. Can the box plot be improved. *Songklanakarin Journal of Science and Technology*, 27(3):649–657, 2005.
20. W. S. Cleveland, P. Diaconis, and R. McGill. Variables on scatterplots look more highly correlated when the scales are increased. *Science*, 216(4550):1138–1141, 1982.

21. D. C. Cliburn, J. J. Feddema, J. R. Miller, and T. A. Slocum. Design and evaluation of a decision support system in a water balance application. *Computers & Graphics*, 26(6):931–949, 2002.
22. A. Coninx, G. P. Bonneau, J. Droulez, and G. Thibault. Visualization of uncertain scalar data fields using color scales and perceptually adapted noise. In *Proc. ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, pages 59–66, 2011.
23. C. E. Connor, H. E. Egeth, and S. Yantis. Visual attention: bottom-up versus top-down. *Current Biology*, 14(19):R850–R852, 2004.
24. M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, 2014.
25. M. Correll, D. Moritz, and J. Heer. Value-suppressing uncertainty palettes. In *Proc. CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2018.
26. M. Daradkeh. Exploring the use of an information visualization tool for decision support under uncertainty and risk. In *Proc. International Conference on Engineering & MIS 2015*, pages 1–7, 2015.
27. S. Deitrick. Evaluating implicit visualization of uncertainty for public policy decision support. In *Proceedings AutoCarto*, page available online, 2012.
28. E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri. Mitigating the attraction effect with visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):850–860, 2018.
29. K. M. Divis, L. E. Matzen, M. J. Haass, and D. A. Cronin. Perceptual biases in scatterplot interpretation. In D. Albers Szafir, R. Borgo, M. Chen, D. Edwards, B. Fisher, and L. Padilla, editors, *Visualization Psychology*. Springer, 2022.
30. S. Djurcilov, K. Kim, P. Lermusiaux, and A. Pang. Visualizing scalar volumetric data with uncertainty. *Computers & Graphics*, 26(2):239–248, 2002.
31. S. Djurcilov, K. Kim, P. F. Lermusiaux, and A. Pang. Volume rendering data with uncertainty information. In *Data Visualization 2001*, pages 243–252. Springer, Vienna, 2001.
32. M. E. Doherty, R. B. Anderson, A. M. Angott, and D. S. Klopfer. The perception of scatterplots. *Perception & Psychophysics*, 69(7):1261–1272, 2007.
33. X. Dong and C. C. Hayes. Uncertainty visualizations: Helping decision makers become more aware of uncertainty and its implications. *Journal of Cognitive Engineering and Decision Making*, 6(1):30–56, 2012.
34. I. Drecki. Visualisation of uncertainty in geographical data. In *Spatial Data Quality*, pages 140–159, 2002.
35. C. Eaton, C. Plaisant, and T. Drizd. Visualizing missing data: Classification and empirical study. In *Proc. IFIP International Conference on Human-Computer Interaction*, pages 861–872, September 2005.
36. L. D. Edwards and E. S. Nelson. Visualizing data certainty: A case study using graduated circle maps. *Cartographic Perspectives*, 38:19–36, 2001.
37. G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1216–1223, 2007.
38. R. Etemadpour, B. Olk, and L. Linsen. Eye-tracking investigation during visual analysis of projected multidimensional data with 2D scatterplots. In *Proc. International Conference on Information Visualization Theory and Applications (IVAPP)*, pages 233–246, 2014.
39. B. J. Evans. Dynamic display of spatial data-reliability: Does it benefit the map user? *Computers & Geosciences*, 23(4):409–422, 1997.
40. J. S. B. Evans. The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13(3):378–395, 2006.
41. D. Feng, L. Kwock, Y. Lee, and R. Taylor. Matching visual saliency to confidence in plots of uncertain data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):980–989, 2010.

42. M. Fernandes, L. Walls, S. Munson, J. Hullman, and M. Kay. Uncertainty displays using quantile dotplots or CDFs improve transit decision-making. In *Proc. CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.
43. R. Finger and A. M. Bisantz. Utilizing graphical formats to convey uncertainty in a decision-making task. *Theoretical Issues in Ergonomics Science*, 3(1):1–25, 2002.
44. N. Fout and K. L. Ma. Fuzzy volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2335–2344, 2012.
45. S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman. The science of visual data communication: What works. *Psychological Science in the Public Interest*, 22(3):110–161, 2021.
46. R. Fuller, N. Dudley, and J. Blacktop. How informed is consent? Understanding of pictorial and verbal probability information by medical inpatients. *Postgraduate Medical Journal*, 78(923):543–544, 2002.
47. M. Gleicher, M. Correll, C. Nothelfer, and S. Franconeri. Perception of average value in multiclass scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 12(19):2316–2325, 2013.
48. D. G. Goldstein and D. Rothschild. Lay understanding of probability distributions. *Judgment & Decision Making*, 9(1), 2014.
49. M. Greis, E. Avci, A. Schmidt, and T. Machulla. Increasing users' confidence in uncertain data by aggregating data from multiple sources. In *Proc. CHI Conference on Human Factors in Computing Systems*, pages 828–840, 2017.
50. H. Griethe and H. Schumann. The visualization of uncertain data: Methods and problems. In *Proc. SimVis*, pages 143–156, 2006.
51. D. Griffin and A. Tversky. The weighing of evidence and the determinants of confidence. *Cognitive Psychology*, 24(3):411–435, 1992.
52. G. Grigoryan and P. Rheingans. Probabilistic surfaces: Point based primitives to show surface uncertainty. In *Proc. IEEE Visualization (VIS 02)*, pages 147–153. IEEE, 2002.
53. G. Grigoryan and P. Rheingans. Point-based probabilistic surfaces to show surface uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 10(5):564–573, 2004.
54. T. Gschwandtner, M. Bögl, P. Federico, and S. Miksch. Visual encodings of temporal uncertainty: A comparative user study. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):539–548, 2016.
55. C. Hansen, S. Zidowitz, F. Ritter, C. Lange, K. Oldhafer, and H. K. Hahn. Risk maps for liver surgery. *International Journal of Computer Assisted Radiology and Surgery*, 8(3):419–428, 2013.
56. S. Haroz, K. L. Ma, and K. Heitmann. Multiple uncertainties in time-variant cosmological particle data. In *Proc. IEEE Pacific Visualization Symposium*, pages 207–214. IEEE, 2008.
57. M. Harrower. Representing uncertainty: Does it help people make better decisions. In *Proc. UCGIS Workshop on Geospatial Visualization and Knowledge Discovery*, pages 18–20, 2003.
58. M. G. Haselton, D. Nettle, and P. W. Andrews. The evolution of cognitive bias. In D. M. Buss, editor, *The Handbook of Evolutionary Psychology*, pages 724–746. John Wiley & Sons Inc, Hoboken, NJ, 2005.
59. R. J. Heuer. *Psychology of intelligence analysis*. Center for the Study of Intelligence, 1999.
60. J. L. Hintze and R. D. Nelson. Violin plots: a box plot-density trace synergism. *The American Statistician*, 52(2):181–184, 1998.
61. M. Hlawatsch, P. Leube, W. Nowak, and D. Weiskopf. Flow radar glyphs–static visualization of unsteady flow with uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1949–1958, 2011.
62. J. M. Hofman, D. G. Goldstein, and J. Hullman. How visualizing inferential uncertainty can mislead readers about treatment effects in scientific results. In *Proc. CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020.
63. D. Howard and A. M. MacEachren. Interface design for geographic visualization: Tools for representing reliability. *Cartography and Geographic Information Systems*, 23(2):59–77, 1996.

64. J. Hullman. Why evaluating uncertainty visualization is error prone. In *Proc. 6th Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pages 143–151, 2016.
65. J. Hullman. Why authors don't visualize uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):130–139, 2019.
66. J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):903–913, 2018.
67. J. Hullman, P. Resnick, and E. Adar. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PloS One*, 10(11), 2015.
68. ISO/IEC Guide 98–3:2008. *Guide to the Expression of Uncertainty in Measurement*. The International Organization for Standardization, 1997.
69. C. R. Johnson and A. R. Sanderson. A next step: Visualizing errors and uncertainty. *IEEE Computer Graphics and Applications*, 23(5):6–10, 2003.
70. D. K. Jones. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magnetic Resonance in Medicine*, 49(1):7–12, 2003.
71. S. Joslyn and D. Jones. Strategies in naturalistic decision making: A cognitive task analysis. In *Naturalistic Decision Making and Macrocognition*, pages 183–201. CRC Press, 2008.
72. S. Joslyn and J. LeClerc. Decisions with uncertainty: The glass half full. *Current Directions in Psychological Science*, 22(4):308–315, 2013.
73. S. Joslyn, K. Pak, D. Jones, J. Pyles, and E. Hunt. The effect of probabilistic information on threshold forecasts. *Weather and Forecasting*, 22(4):804–812, 2007.
74. S. L. Joslyn and J. E. LeClerc. Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of Experimental Psychology: Applied*, 18(1):126, 2012.
75. D. Kahneman, P. Slovic, and A. E. Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, 1982.
76. D. Kahneman and A. Tversky. On the study of statistical intuitions. *Cognition*, 11(2):123–141, 1982.
77. A. Kale, F. Nguyen, M. Kay, and J. Hullman. Hypothetical outcome plots help untrained observers judge trends in ambiguous data. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):892–902, 2019.
78. M. Kay, T. Kola, J. R. Hullman, and S. A. Munson. When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proc. CHI Conference on Human Factors in Computing Systems*, pages 5092–5103, 2016.
79. C. Kinkeldey, A. M. MacEachren, M. Riveiro, and J. Schiewe. Evaluating the effect of visually represented geodata uncertainty on decision-making: systematic review, lessons learned, and recommendations. *Cartography and Geographic Information Science*, 44(1):1–21, 2017.
80. C. Kinkeldey, A. M. MacEachren, and J. Schiewe. How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4):372–386, 2014.
81. C. Kinkeldey, J. Mason, A. Klippel, and J. Schiewe. Evaluation of noise annotation lines: using noise to represent thematic uncertainty in maps. *Cartography and Geographic Information Science*, 41(5):430–439, 2014.
82. S. S. Kirschenbaum and J. E. Arruda. Effects of graphic and verbal probability information on command decision making. *Human Factors*, 36(3):406–418, 1994.
83. C. Konold. Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, 3(1), 1995.
84. E. L. Korn and B. I. Graubard. Scatterplots with survey data. *The American Statistician*, 52(1):58–69, 1998.
85. R. Kosara, S. Miksch, H. Hauser, J. Schrammel, V. Giller, and M. Tscheligi. Useful properties of semantic depth of field for better F+C visualization. In *ACM International Conference Proceeding Series*, volume 22, pages 205–210, 2002.

86. C. Kreuzmair, M. Siegrist, and C. Keller. High numerates count icons and low numerates process large areas in pictographs: results of an eye-tracking study. *Risk Analysis*, 36(8):1599–1614, 2016.
87. M. Kunz, A. Grêt-Regamey, and L. Hurni. Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system. *Natural Hazards*, 59(3):1735–1751, 2011.
88. D. M. Lane, C. A. Anderson, and K. L. Kellam. Judging the relatedness of variables: The psychophysics of covariation detection. *Journal of Experimental Psychology: Human Perception and Performance*, 11(5):640–649, 1985.
89. M. Leitner and B. P. Buttenfield. Guidelines for the display of attribute certainty. *Cartography and Geographic Information Science*, 27(1):3–14, 2000.
90. H. Li, C. W. Fu, Y. Li, and A. Hanson. Visualizing large-scale uncertainty in astrophysical data. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1640–1647, 2007.
91. J. Li, J. B. Martens, and J. J. van Wijk. A model of symbol size discrimination in scatterplots. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pages 2553–2562. ACM, 2010.
92. J. Li, J. B. Martens, V. Wijk, and J. J. Judging correlation from scatterplots and parallel coordinate plots. *Information Visualization*, 9(1):13–30, 2010.
93. S. K. Lodha, N. M. Faaland, A. P. Charaniya, P. K. Varshney, K. Mehrotra, and C. Mohan. Visualization of uncertain particle movement. In *Proc. Computer Graphics and Imaging Conference*, pages 226–232, 2002.
94. S. K. Lodha, A. Pang, R. E. Sheehan, and C. M. Wittenbrink. UFLOW: Visualizing uncertainty in fluid flow. In *Proc. IEEE Visualization (VIS 96)*, pages 249–254. IEEE, 1996.
95. S. K. Lodha, B. Sheehan, A. T. Pang, and C. M. Wittenbrink. Visualizing geometric uncertainty of surface interpolants. In *Proc. Graphics Interface*, pages 238–245, 1996.
96. A. Lopes and K. Brodlie. Accuracy in 3D particle tracing. In *Mathematical Visualization*, pages 329–341. Springer, Berlin, 1998.
97. C. Lundström, P. Ljung, A. Persson, and A. Ynnerman. Uncertainty visualization in medical volume rendering using probabilistic animation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1648–1655, 2007.
98. A. M. MacEachren. Visualizing uncertain information. *Cartographic Perspectives*, 13:10–19, 1992.
99. A. M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3):139–160, 2005.
100. A. M. MacEachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan. Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2496–2505, 2012.
101. L. E. Matzen, K. M. Divis, M. J. Haass, and D. Cronin. Variable biases: A study of scientists? interpretation of plot types commonly used in scientific communication. Technical Report (No. SAND2020–8777C), Sandia National Laboratories, Albuquerque, NM, 2020.
102. L. E. Matzen, M. J. Haass, K. M. Divis, Z. Wang, and A. T. Wilson. Data visualization saliency model: A tool for evaluating abstract data visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):563–573, 2017.
103. N. McCurdy, J. Gerdes, and M. Meyer. A framework for externalizing implicit error using visualization. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):925–935, 2018.
104. R. McGill, J. W. Tukey, and W. A. Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, 1978.
105. A. McNutt, G. Kindlmann, and M. Correll. Surfacing visualization mirages. In *Proc. CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2020.
106. J. Meyer and D. Shinar. Estimating correlations from scatterplots. *Human Factors*, 34(3):335–349, 1992.

107. M. Monmonier and M. Gluck. Focus groups for design improvement in dynamic cartography. *Cartography and Geographic Information Systems*, 21(1):37–47, 1994.
108. T. Munzner. *Visualization Analysis and Design*. CRC Press, 2014.
109. T. Mussweiler and F. Strack. Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, 35(2):136–164, 1999.
110. L. Nadav-Greenberg and S. L. Joslyn. Uncertainty forecasts improve decision making among nonexperts. *Journal of Cognitive Engineering and Decision Making*, 3(3):209–227, 2009.
111. National Research Council. *Completing the forecast: Characterizing and communicating uncertainty for better decisions using weather and climate forecasts*. The National Academies Press, Washington, DC, 2006.
112. G. E. Newman and B. J. Scholl. Bar graphs depicting averages are perceptually misinterpreted: The within-the-bar bias. *Psychonomic Bulletin & Review*, 19(4):601–607, 2012.
113. T. S. Newman and W. Lee. On visualizing uncertainty in volumetric data: techniques and their evaluation. *Journal of Visual Languages & Computing*, 15(6):463–491, 2004.
114. A. O'Hagan, C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons, 2006.
115. A. Osorio, R. S., and K. W. Brodlie. Uncertain flow visualization using LIC. In *Proc. Theory and Practice of Computer Graphics, Eurographics UK Chapter Conference*, pages 215–222. Eurographics Association, 2009.
116. L. Padilla, S. C. Castro, and H. Hosseinpour. A review of uncertainty visualization errors: Working memory as an explanatory theory. *Psychology of Learning and Motivation*, 74:275–315, 2021.
117. L. Padilla, M. Kay, and J. Hullman. Uncertainty visualization. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, and J. Teugels, editors, *Wiley StatsRef: Statistics Reference Online*. https://doi.org/10.1002/9781118445112.stat08296, 2021.
118. L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, 3(1):29, 2018.
119. L. M. Padilla, G. Hansen, I. T. Ruginski, H. S. Kramer, W. B. Thompson, and S. H. Creem-Regehr. The influence of different graphical displays on nonexpert decision making under uncertainty. *Journal of Experimental Psychology: Applied*, 21(1):37, 2015.
120. L. M. Padilla, I. T. Ruginski, and S. H. Creem-Regehr. Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive Research: Principles and Implications*, 2(1):1–16, 2017.
121. A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, 1997.
122. T. Pfaffelmoser, M. Reitinger, and R. Westermann. Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. *Computer Graphics Forum*, 30(3):951–960, 2011.
123. K. Pothkow and H. C. Hege. Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1393–1406, 2010.
124. K. Potter, P. Rosen, and C. R. Johnson. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In *Uncertainty Quantification in Scientific Computing: IFIP Working Conference on Uncertainty Quantification*, pages 226–249. Springer, Berlin, 2011.
125. K. Potter, P. Rosen, and C. R. Johnson. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In *IFIP Advances in Information and Communication Technology*, volume 377, pages 226–249, 2012.
126. J. S. Prassni, T. Ropinski, and K. Hinrichs. Uncertainty-aware guided volume segmentation. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1358–1365, 2010.

127. P. S. Quinan, L. M. Padilla, S. H. Creem-Regehr, and M. Meyer. Towards ecological validity in evaluating uncertainty. In *Proc. Workshop on Visualization for Decision Making Under Uncertainty (at VIS'15)*, 2015.
128. C. B. Read. Five? number summaries. In *Encyclopedia of Statistical Sciences*, volume 4. Wiley, 2004.
129. R. A. Rensink and G. Baldridge. The perception of correlation in scatterplots. *Computer Graphics Forum*, 29(3):1203–1210, 2010.
130. P. Rheingans and S. Joshi. Visualization of molecules with positional uncertainty. In *Proc. Data Visualization'99*, pages 299–306. Springer, Vienna, 1999.
131. P. J. Rhodes, R. S. Laramee, R. D. Bergeron, and T. M. Sparr. Uncertainty visualization methods in isosurface rendering. In *Proc. Eurographics (Short Papers)*, 2003.
132. C. Roessing, A. Reker, M. Gabb, K. Dietmayer, and H. P. Lensch. Intuitive visualization of vehicle distance, velocity and risk potential in rear-view camera applications. In *Proc. IEEE Intelligent Vehicles Symposium*, pages 579–585, 2013.
133. P. J. Rousseeuw, I. Ruts, and J. W. Tukey. The bagplot: a bivariate boxplot. *The American Statistician*, 53(4):382–387, 1999.
134. I. T. Ruginski, A. P. Boone, L. M. Padilla, L. Liu, N. Heydari, H. S. Kramer, and S. H. Creem-Regehr. Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition & Computation*, 16(2):154–172, 2016.
135. E. J. Saccone, O. Landry, and P. A. Chouinard. A meta-analysis of the size-weight and material-weight illusions. *Psychonomic Bulletin & Review*, 26(4):1195–1212, 2019.
136. A. Sarikaya and M. Gleicher. Scatterplots: Tasks, data, and designs. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):402–412, 2018.
137. S. Savelli and S. Joslyn. The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology*, 27(4):527–541, 2013.
138. S. C. Sibrel, R. Rathore, L. Lessard, and K. B. Schloss. The relation between color and spatial structure for interpreting colormap data visualizations. *Journal of Vision*, 20(12):7, 2020.
139. A. L. Simpson, B. Ma, E. M. Vasarhelyi, D. P. Borschneck, R. E. Ellis, J. Stewart, and A. Computation and visualization of uncertainty in surgical navigation. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 10(3):332–343, 2014.
140. M. Skeels, B. Lee, G. Smith, and G. Robertson. Revealing uncertainty for information visualization. In *Proc. Working Conference on Advanced Visual Interfaces*, pages 376–379, 2008.
141. D. Spiegelhalter, M. Pearson, and I. Short. Visualizing uncertainty about the future. *Science*, 333(6048):1393–1400, 2011.
142. D. A. Stapel and W. Koomen. Social categorization and perceptual judgment of size: When perception is social. *Journal of Personality and Social Psychology*, 73(6):1177, 1997.
143. B. N. Taylor and C. E. Kuyatt. Guidelines for evaluating and expressing the uncertainty of NIST measurement results. Technical Report Technical Note 1297, NIST, 1994.
144. J. Thomson, E. Hetzler, A. MacEachren, M. Gahegan, and M. Pavel. A typology for visualizing uncertainty. In *Proc. Electronic Imaging 2005*, page 146–157. International Society for Optics and Photonics, 2005.
145. A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.
146. E. T. Watkins. Improving the analyst and decision-maker's perspective through uncertainty visualization. Technical Report No. AFIT/GCS/ENG/00M-24, Air Force Inst of Tech Wright-Patterson Afb OH School of Engineering, 2000.
147. L. Wilkinson. Dot plots. *The American Statistician*, 53(3):276–281, 1999.
148. C. M. Wittenbrink, A. T. Pang, and S. K. Lodha. Glyphs for visualizing uncertainty in vector fields. *IEEE Transactions on Visualization and Computer Graphics*, 2(3):266–279, 1996.
149. J. M. Wolfe. Guided Search 4.0. In W. Gray, editor, *Integrated Models of Cognitive Systems*, pages 99–119. Oxford University Press, New York, 2007.

150. Y. Wu, G. X. Yuan, and K. L. Ma. Visualizing flow of uncertainty through analytical processes. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2526–2535, 2012.
151. C. Xiong, V. Weelden, L., and S. Franconeri. The curse of knowledge in visual data communication. *IEEE Transactions on Visualization and Computer Graphics*, 26(10):3051–3062, 2019.
152. D. Zhang, E. Adar, and J. Hullman. Visualizing uncertainty in probabilistic graphs with network hypothetical outcome plots (NetHOPs). *IEEE Transactions on Visualization and Computer Graphics*, 28(1):443–453, 2022.
153. T. Zuk, J. Downton, D. Gray, S. Carpendale, and J. D. Liang. Exploration of uncertainty in bidirectional vector fields. In *Proc. Visualization and Data Analysis*, volume 6809, page 68090B, 2008.

Chapter 15
Analysis of Sensemaking Strategies: Psychological Theories in Practice

Margit Pohl, Johanna Doppler Haider, Patrick Seidler, Neesha Kodagoda, and B. L. William Wong

Abstract Sensemaking processes are regarded as a relevant conceptualization of how users interact with information visualizations. Nevertheless, there is little research about the specific sensemaking strategies users adopt when they work with visualizations. Psychological theories about human thinking and reasoning and theories from the area of graph comprehension are relevant approaches that should be taken into account when investigating sensemaking processes with information visualizations. In these areas, there is more detailed research about problem-solving strategies (e.g., in mathematical problem-solving) that could be relevant for information visualization. We provide an overview of interesting approaches and in which way they are relevant for interactions with visualizations. We describe an exploratory investigation with 18 computer science students performing a realistic task using a visual analytics system. The result of this investigation was a set of eleven sensemaking strategies. We discuss whether these strategies can be generalized across different visualizations and compare the results to results from other studies we have conducted in this area. We also present examples for recommendations based on such research.

M. Pohl (✉) · J. D. Haider
Vienna University of Technology, Vienna, Austria
e-mail: margit@igw.tuwien.ac.at; johanna.haider@tuwien.ac.at

P. Seidler · N. Kodagoda
Genetec UK, London, UK
e-mail: pseidler@genetec.com; gnkodagoda@genetec.com

B. L. William Wong
Auckland University of Technology, Auckland, New Zealand
e-mail: william.wong@aut.ac.nz

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2_15

15.1 Introduction

In information visualization and visual analytics, the issues of sensemaking and insight generation have been addressed occasionally [3, 21, 29, 34]. Nevertheless, we still have no full understanding of the cognitive processes taking place, while users interact with visualizations. Some approaches in cognitive psychology may help to clarify open questions in this area. Making sense of visualizations may be seen as a problem-solving activity. This has already been suggested by Mayr et al. [23].

Research on problem-solving has investigated problem-solving processes in great detail and has shown what kinds of strategies are used in such processes [11, 22, 25, 39]. In this context, the relationship between the usage of more general strategies and domain-specific strategies has been discussed. While much research on problem-solving has been conducted in laboratories, it has been argued that there might be differences between reasoning processes in the laboratory and "in the wild" [41]. Research in everyday thinking and reasoning has indicated that people use a wide variety of strategies, and the strategy usage is highly dependent on domain knowledge [41]. Research in graph comprehension has addressed the topic of different levels of insight generation when interacting with graphs [8]. Naturalistic decision-making [14–16] has been discussed in the visualization community already, and some research has been done based on this approach [18].

These approaches are also relevant for the design of visualizations. It has already been argued by visualization researchers that a detailed analysis of the interaction processes of users with visualizations is necessary [28]. A result of such research could be the identification of strategies or heuristics that users adopt. These strategies (or a combination of them) can be more or less effective. Design of visualization interfaces could be adapted to more effective strategies. Research on everyday thinking and reasoning has indicated that people use strategies very flexibly [41]. This flexibility has to be taken into account also in the design of visualizations. The everyday thinking and reasoning approach also emphasizes the importance of domain knowledge. The importance of background knowledge is commonly accepted in the visualization community, but it is not entirely clear how this should be reflected in the interface design. The theory of graph comprehension [8] has indicated that novices tend to create more simple forms of insights than experts. It is necessary to identify those types of visualizations that support more simple or more complex types of insight generation. In addition, all these approaches may be used to inform teaching the understanding and creation of visualizations and to increase visualization literacy in general.

In this chapter, we want to provide a brief overview of these theories and then describe a study we conducted that can show the relevance of the application of these theories. The main goal of this chapter is to identify sensemaking strategies adopted by users of visualizations. In addition, we try to show how existing theories of problem-solving and cognitive strategies can be applied in this kind of research.

15.2 Related Work

In cognitive psychology, there is considerable amount of research on how people solve problems. Simon and Newell [39] and Newell and Simon [25] conducted influential research on problem-solving. They assume that there is a large problem space with all possible states that can be reached during the problem-solving process. It is necessary to reduce the number of possible states to make this process manageable. They suggest to use heuristics to achieve this. There are some general-purpose heuristics that can be used (e.g., hill-climbing or means-end analysis) [31]. Hill-climbing describes a process where problem-solvers move forward one step at a time, and every step aims to get him or her nearer to the goal. Occasionally, this forces the problem-solver to do some backtracking. Means-end analysis is a more sophisticated strategy. The problem-solver analyzes the current state and the goal state and tries to identify actions that reduce the differences between those two. In this context, the problem is split into subproblems. Other general-purpose heuristics that have been described in the literature are "Less is (sometimes) more" or the "Take the Best" heuristic [10, 11]. "Less is (sometimes) more" means that human beings can be overwhelmed by information, and, therefore, sometimes prefer to rely on the most relevant information items and ignore the rest. This heuristic is relevant for information visualization because users often are overwhelmed by the amount of information they get. Designers of visualizations have to take care to offer as much information as necessary, but not more. "Take the Best" tries to balance the search for the best solution with restrictions in time. Human beings often fairly quickly choose alternatives that satisfy their needs to a certain degree, although they know that this alternative is maybe not the best one (satisficing).

There is some overlap between the concept of heuristics and the concept of strategies. A heuristic is usually a rule of thumb that sometimes works unconsciously, while strategies are seen as conscious step-wise processes, although there is some controversy about this definition [22]. Lemaire and Fabre [22] also discuss the relative importance of general problem-solving strategies that are valid across domains and specific problem-solving strategies. They point out that novices tend to adopt general problem-solving strategies as long as they lack relevant domain knowledge. It is an open question whether users of visualizations rather use general strategies or domain-specific strategies. In our research, we found that users of information visualizations use both general and domain-specific strategies. It is, however, not clear which of these approaches is more successful.

Roberts and Newton [32] argue that people often do not use the best strategy because it is not available to them. Nevertheless, good strategies can be learned. Recently, visualization literacy has been discussed within the visualization community [1, 21]. It has been pointed out that many people lack visualization literacy [2]. Analyzing strategies of how people infer insights from visualizations and comparing them to successful strategies can help to inform the teaching of visualization literacy and identify problems with the visualization that might be reduced by an appropriate design. Roberts and Newton [32] point out that there is some variability in strategy selection because different individual strategies may be appropriate for different parts of a problem. Strategy selection is also context dependent.

There is also some interesting research in the area of “Everyday Reasoning” [9, 41]. Woll [41] argues that there is some difference between everyday reasoning and formal reasoning. Context and background knowledge are highly relevant for everyday reasoning. Scribner [37] conducted research about strategies used in everyday reasoning and found that people are highly flexible in their usage of strategies. They tend to adapt their strategies to the situation at hand. This conforms to research done by Lave [20] who argued that formal processes learned at school are often not applied in practice. She found out that mathematics used in supermarkets or other similar contexts differs from formal mathematical procedures. In our research, we also found that in many cases there is not one optimal cognitive strategy but several strategies that can be used flexibly.

Evans [7] also points out that there is a difference between the strategies used in everyday problem-solving processes and those strategies used in rational decision-making based on principles of formal logic. He assumes that there are two systems of decision-making, one that is fast and often unconscious and another one that is slow, deliberative, and analytic (dual-process model). In everyday decision-making processes, people usually only consider one alternative at a time and are often content with a decision that reaches a certain threshold (satisficing). Padilla et al. [27] showed that there is evidence that the dual-process model is also valid for interactions with visualizations. Based on this idea, they developed an integrated cognitive framework for studying the cognitive processes underlying the use of visualizations.

Much of the research concerning strategy use has been conducted in the educational domain [4]. One of the major issues discussed in this context is whether there is a relationship between strategy use and learning outcomes. There is some indication that activities that students engage in while they learn are a better predictor of learning outcomes than their abilities or other individual differences [6]. Nevertheless, the relationship between strategy use and performance is still not very well-understood [4]. In our research, we also addressed the issue of the relationship between strategy use and performance. We found some weak relationship, but we would like to point out that we looked at very specific strategies related to the interaction with visualizations.

Another approach that also investigates how users interact with visual material is graph comprehension. The goal of graph comprehension is to clarify how people make sense of graphs and how they draw inferences from them [8]. Viewers are supposed to develop a coherent mental model of the application domain represented by graphs. Typical strategies that are adopted by viewers are connecting elements of a graph or predicting future behavior of the systems represented by the graph. Research in graph comprehension is also relevant for educational issues. The aim of much of the research in that area is to improve graphs, so that students will be better able to gain knowledge about the domain represented in the graph.

Research on graph comprehension, e.g., Kosslyn [19], Tversky [40], deals with sensemaking processes using visualizations. Friel et al. [8] developed a model consisting of three different levels: (1) reading the data (i.e., extracting data, locating data), (2) reading between the data (i.e., finding relationships, integrating data), and (3) reading beyond the data (i.e., extrapolating from the data and generating hypotheses). This model conforms to our research findings concerning insights. We found low-level insights reporting simple facts without explanations, moderate insights including explanations, and high-level insights including recommendations based on new hypotheses.

Probably, the most influential theoretical approach for the study of sensemaking in the visualization domain is the data–frame model developed by Klein et al. [14, 15]. The data–frame model is inspired by the approach of naturalistic decision-making (NDM). NDM aims to model decision processes in realistic situations by domain experts. Much of the research concerning problem-solving strategies is conducted using the so-called puzzle problems. Using puzzle problems helps to achieve easily controllable experimental conditions. Research under realistic conditions, on the other hand, trades validity and practical relevance for experimental rigor. The data–frame model assumes that people develop schematic representations of the phenomena they encounter in their daily lives called frames. These frames are dynamic models that can be elaborated, questioned, or rejected. Klein [13] later on extended his model to include specific sensemaking processes: The Triple Path Model of Insight that explains how scientists gain their insights. This model incorporates three processes: (1) Connection, (2) Contradiction, and (3) Creative desperation. When people connect information, they try to identify patterns or find relationships between elements. Contradiction implies that people are confronted with information that is contradictory and does not allow to form a coherent mental model. Creative desperation happens when people cannot make sense of the information they encounter. The model of Klein et al. has already been used for the evaluation of visualizations before (see, e.g., Kodagoda et al. [18]).

The work described in the following is based to a large extent on Klein's models using the coding scheme of previous work [5] as a starting point. Refinements and changes on the coding scheme were made to create more precise and selective categories.

These approaches are, to a certain extent, related to each other. Some approaches within the research on problem-solving address problems in ill-structured domains when neither the route to a solution nor the solution itself can be defined easily [23]. In such cases, heuristics are often applied [11, 22, 25, 39]. Some of them also conceptualize problem-solving as an activity happening in realistic situations [14–16, 41]. All approaches we will describe focus on cognitive activities that are exploratory and dependent on previous knowledge. Especially research on cognitive strategies [4, 7] and on graph comprehension [8] is often motivated by issues from educational research where previous knowledge and level of expertise play an important role. Visualizations often represent data in ill-structured domains, they require the generation of insights in realistic situations, interaction with visualizations often relies on expertise of the users, and exploratory activities are essential for gaining insights. Approaches already used in the visualization community are partly based on theories mentioned above, especially Klein et al.'s data–frame theory [14, 15] that was very influential for the adoption of the sensemaking approach.

15.3 System Description

The VALCRI system was designed to support and augment the intelligence analysis process based on user requirements from focus groups with 20 intelligence analysts using Cognitive Task Analysis [35] and one-to-one interviews with seven intelligence analysts using the retrospective interview technique Critical Decision Method [12, 17]. The goal of the system is to support intelligence analysis in police forces to identify criminal offenders. It contains several visualizations showing place and time of crimes, the modus operandi, and possible relationships between offenders.

The system supports the analysis of events happening in space and time through multiple visualizations that can be used as filters in connected views. The results can be explored via ten visualizations: Search, Timeline, Map, Bar Chart, List, Statistical Process Chart (SPC), Space Similarity Selector (S3), Clusters, Crime Classification Table (CCT), and Crime Cards. The timeline represents how many crimes were committed in a period of time. The map shows aggregated crime event locations in an overlaying grid for results containing more than 300 events. The bar chart provides a distribution of different characteristics of those events, e.g., the offence type is shown per default, see Fig. 15.1. The SPC shows the mean crime rate of a selected time period and its standard deviation to detect anomalies in the data, such as very large or sudden shifts. Intelligence analysts interpret various types of deviation from the mean over consecutive points typically as a new trend, which are

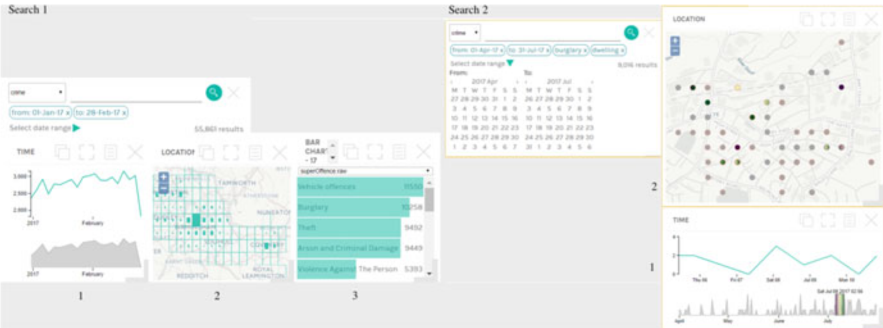

Fig. 15.1 Multiple, interactive, and connected visualizations provide insight on spatiotemporal–thematic event data. Search queries can be run in parallel (search 1 and search 2) to compare different result sets or to use different perspectives on a large canvas. The standard views are a timeline (1), a map (2), and a bar chart (3), which show the data on different aggregation levels; The cluster overview in search 1 on the left shows more than 300 events; In search 2 on the right, detailed events are shown with overlaid temporal information in a map based on a selected time frame in the timeline

encoded as turquoise points in the chart. Outliers are highlighted with red crosses. The CCT summarizes the information from the crime reports showing all features in columns (e.g., time, crime type, etc.) and crimes per row. Clusters can be visualized via the S3 and cluster view. The S3 view groups crimes into clusters based on their similarities, providing controls for the k-means algorithm to, e.g., define the number of clusters. Each crime report is also depicted in a summarizing crime card. Finally, multiple visualizations can be grouped into lists.

15.4 Study

We analyze the use of multiple visualizations with regard to insight and sensemaking to extract strategies participants employ while working with the VALCRI system to make sense of information provided by the visual analytics tool. We expect participants to use different visualizations to answer different kinds of questions and follow up on investigational hypotheses during an analytical task. Our research questions are:

- R1: How do users generate insights with intelligence-specific visualizations and which visualization tools do they use to achieve this?
- R2: Which sensemaking strategies do participants use most?
- R3: Are sensemaking strategies related to the number and/or quality of insights?

To answer them, we observe sensemaking strategies and insights during an analytic task. The outcome of the analytic task can be compared in terms of the number of insights and depth of intelligence analysis, i.e., the quality of insights gained.

15.4.1 Methodology

We conducted an extensive think-aloud study for the formative evaluation of a visual analytics prototype. For the purpose of analyzing cognitive processes, the think-aloud method proved to be useful [30]. An essential part of such an analysis is the development of an appropriate coding scheme to get systematic results from these protocols [33]. The development of such coding schemes can either be based on the literature (top-down) or on the repeated analysis of the protocols (bottom-up). Frequently, both approaches are combined. The development of such coding schemes can take place in the context of a qualitative content analysis [24, 33, 36]. The coding scheme for sensemaking strategies was developed in an iterative approach combining bottom-up and top-down elements, in which two authors studied the content together and refined a coding scheme to eleven content-specific sensemaking codes through repeated in-depth discussions. Consequently, the entire think-aloud protocols were coded with these categories. A second evaluator coded two minutes each from the first and the second half of the screen capture at random (i.e., 12.4% of the protocols). The result of the coding process was the identification of 11 sensemaking strategies. The sensemaking strategies were analyzed by their occurrences using two measures:

- Frequency: How often is a sensemaking strategy used?
- Duration: How long does a sensemaking strategy take?

Our research questions with respect to insight were addressed by looking at three aspects:

- Number of insights: How many insights could be gained during the analysis?
- Quality of insights: On which level are the reported insights? How many conclusions can be made for future actions?

We counted the number of these insights and assessed the quality of insights based on the following criteria: whether the participants provided a textual explanation of their insight and whether the text was only an explanation or contained a recommendation in addition.

15.4.2 Participants

We recruited eighteen computer science students with a Bachelor's degree (5 Female, 13 Male; $N = 18$). Participants were trained in strategic analysis goals and in using the system with an introduction and a video tutorial explaining the context as well as specific terminology. No color vision impairments were reported.

The duration of the experiment was 100–120 min with two short breaks after one hour and just before a questionnaire about their experience. The study consisted of a training task, an analysis task, and a short, semi-structured interview to follow up on issues that arose during the analysis. Participants worked on the analysis task for 45–55 min.

15.4.3 Dataset and Task

The dataset comprises more than 1.1 million records of burglary crime in the time frame of three years in the United Kingdom. The records describe property stolen and damaged as well as people affected by and associated with the crime.

Analysis Task: "Analyze offences that occurred between April 1st and July 31st 2017 and prepare a short report that summarizes your findings."

Participants should report insights, assessing relevant crime data by time, location, and crime type. We think that the task implicitly defines the insights that should be the outcome. This makes it more obvious how insights should be measured. In our case, an insight is a policing recommendation accompanied by one or two screenshots showing the visualization(s) that helped the participants to get that insight. This screenshot was in most cases accompanied by a written explanation.

15.5 Results

The main goal of the investigation was to identify sensemaking strategies. This was a qualitative research process. The coding process in such a study is very time-consuming. Therefore, we only used a relatively small sample size. We also quantitatively analyzed some results, but these quantitative results have to be checked in future research with larger samples.

The analysis of sensemaking strategies during the task reveals differences in the insight generation process (Research question R1). A detailed description of the sensemaking strategies observed (Research question R2) including coding criteria and exemplary statements is provided in the following. Insights that got documented during the analysis by the participants and their relation to the analysis processes help to identify more successful strategies (Research question R3).

15.5.1 Sensemaking Strategies

The coding scheme includes eleven sensemaking strategies, which showed high inter-coder reliability (Cohen's Kappa $\kappa = 0.871$, $N = 115$). The agreement of two coders was high for all strategies, except the scarce Coincidental Aha, compare Table 15.1. We also add statements by participants (P1–P18). The strategies are described in the following:

Table 15.1 Frequency of codes (sensemaking strategies) assigned by the first experimenter and agreement rate of the second experimenter

Code	Frequency	Agreement
Pattern	286	93%
Trend	158	90%
Profiling	121	86%
Elimination	116	85%
Storytelling	75	100%
Elimination incl. trend	57	100%
Creative desperation	33	100%
Contradiction	26	100%
Verification	25	75%
Pattern incl. profiling	15	67%
Coincidental Aha	12	50%

15.5.1.1 Pattern: Looking for Similarities Across Several (Groups of) Actors

Similarities in our data were crime types, criminals' time intervals (when and why actors were active), direct and indirect relationships, etc.

Example statements from the participants are: "So what I am interested in is how they got in, window or door [looking at clusters in CCT] (P4)," and "A lot of them are unsolved [map] (P15)." The definition of this strategy is based on the work of Klein [13].

15.5.1.2 Trend: Looking for Trends in the Data

Attempt to identify the change over time, here, detect an increase or decrease in criminal activity. Criteria for coding were: (1) Discussing crime rate change in timeline and (2) Comparing development over certain periods of time.

Example: "In this time period it [crime rate] goes up (P2)."

The definition of this strategy is based on the work of Klein [13].

15.5.1.3 Profiling: Characterizing Crimes or Criminals Based on Features

Features in our data were links (identify gangs), crime types (what type of crimes is most critical), or time intervals (when and why actors were active). Criteria for coding were: (1) Inspecting specific individual, group of actor(s) or crime, and (2) Considering various aspects about suspects and going into detail, which may contain trend assessment. Profiling aims at details, not at comprehending relationships and trends. It is a necessary activity but does not support the ability to go beyond the data, as defined in the model of Friel et al. [8]. This is a domain-specific strategy.

Examples: "I have a crime with a victim, I will have a look at [the victim], maybe he was even mugged twice (P16)," and "I will read the description [of a crime report] (P18)."

15.5.1.4 Pattern Incl. Profiling: Combination of Pattern and Profiling

Criteria for coding were if the activity could not be coded separately, e.g., comparing crime reports with each other. Our aim was that codes should be mutually exclusive. In a few cases (e.g., in pattern and profiling), this was not possible. Therefore, we introduced a combination of two codes.

Example: [putting several crime reports next to each other] (P18)

15.5.1.5 Elimination: Generating New Understanding by Eliminating Data Considered as Not Relevant

The activity of filtering out data that do not fit or are not interesting.

Example: "I will select burglary dwelling in the bar chart [which adds a filter on the crime type] (P10)."

15.5.1.6 Elimination Incl. Trend: Reducing the Search Space Due to Time

Reducing possibilities that do not fit a temporal requirement.

Example: "I will look at 16th July (P10)."

15.5.1.7 Storytelling: Constructing a Story by Explaining the Behavior of Crimes and Relationships

The observations within the data were given meaning by using one's experience or imagination to add information that subjectively makes sense, which followed often after profiling. Criteria for coding were: (1) Added information that is not visibly obtained from the data and (2) "Made-up" or subjectively enhanced information depending on the previous life experience of the participants.

Example: "People come home from the Easter holidays, kids go back to school and parents work and then we have all the dwellings (P2)."

This category is based on the work by Segel and Heer [38].

15.5.1.8 Creative Desperation: Not Knowing What to Do Next and the Feeling of Being Stuck in an Impasse

Expressing to be stuck and to not know how to continue.

Examples: "Now I am not sure of what best to look for (P4)," and "At the moment I am a bit at a loss what I could try else (P9)."

The definition of this strategy is based on the work of Klein [13].

15.5.1.9 Verification: Consulting Both Representations for Verification

Looking up information in another visualization with the intent to verify results from the analysis.

Example: "I suppose I can see this here [in this visualization] as well (P1)."

15.5.1.10 Contradiction: Realizing a Mismatch of What Was Hypothesized

Noticing contradictory information and that previous thoughts or assumptions were not right. Coding criteria: Realization of contradictory information of what was assumed to be true.

Examples: "I don't see a clear pattern as I would expect from ... (P3)," and "Now they don't seem to be connected [anymore] (P16)."

The definition of this strategy is based on the work of Klein [13].

15.5.1.11 Coincidental Aha's: Seemingly Coincidental Insights That Are Not Conscious

This activity was coded if it was not clear why an action was taken, or where a new idea came from. The insights are not based on structured, conscious reasoning processes.

Examples: "Ah! That's a good cluster (P3)," and "Ah, that's interesting (P13)."

The definition of this strategy is based on the work of Klein [13].

15.5.2 Reported Insights

We asked participants to report their insights in a tool of their choice (MS Word/PowerPoint) using screenshots to capture the state of the tool when an insight was gained and annotations for explanation. It was up to the individual what to report, and no minimum length per explanation was required. The task description suggested to include time, location, and crime type for a recommendation. We analyzed the screenshots to see which visualizations are included to convey an insight and the text descriptions to assess the insight quality. In total, 89 insights were reported using 1.7 visualizations per insight on average. Two thirds of the participants reported more than three insights. One insight was the minimum and eleven the maximum. We grouped the participants according to the number of insights they produced (few, moderate, many). The participants are spread evenly among those groups (6 participants in every group). Few insights ranged from 1 to 3, a moderate number of insights from 4 to 6, and many insights from 7 to 11.

We evaluated the quality of the insight reports by assessing the depth of the given recommendation. Low-level insights are statements about single facts, such as crime rates at certain dates shown via screenshots. Moderate reports describe possible connections between data. They are used as explanations for observed phenomena. These connections or patterns are used to develop hypotheses. High-level insights are conclusion that can be drawn from discovered problematic crime scenes. A high-level insight report includes at least one policing recommendation.

This coding scheme is based on the three-layered model developed by Friel et al. [8].

15.5.3 Employment of Strategies and the Number of Insights

We identified *Looking for Patterns* as the predominant strategy in our evaluation, coded twice as much as most of the other strategies. This activity includes comparing characteristics of crime events, such as location or offence type, and often describes manual clustering. Hence, participants were mostly looking for similarities within crimes. The second most frequent strategy *Looking for Trends* is similar to *Looking for Patterns* in that connections or similarities are perceived. The task description contained much temporal information, such as bank and school holidays, which were utilized by all participants during their analysis. In contrast, the specifications of the burglary type and town were sometimes disregarded. The offence type was disregarded in eight cases and the fictional town in five cases.

There is a difference in how long the individual strategies were used. We measured how long the strategies were employed during every occurrence, compare Fig. 15.2. *Patterns incl. Profiling* took longest with 48 s, i.e., reading crime reports is a very time-consuming strategy. Participants searched longer for Trends than, e.g., for *Patterns or Verification*. *Profiling* and *Patterns incl. Profiling* took longer than

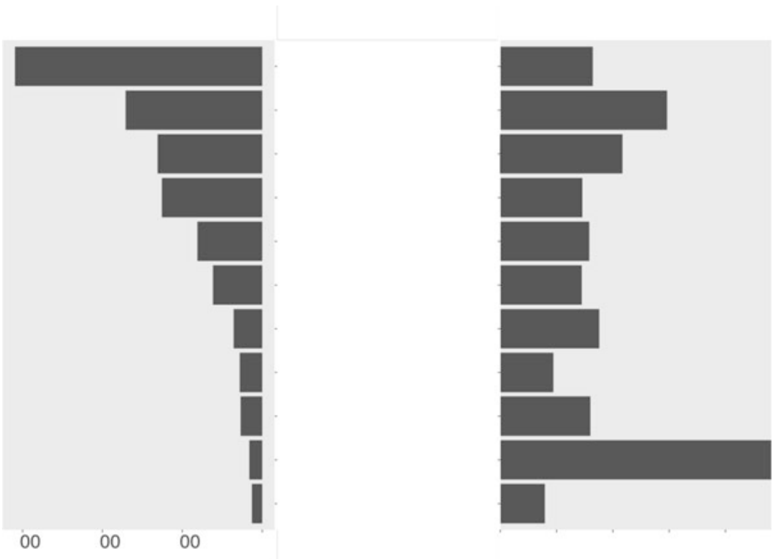

Fig. 15.2 Relative duration (sec) per sensemaking strategy in the order of strategy frequency. Looking for patterns was used most often and moderately long, while *Coincidental Aha's* could be observed as short moments, occurring least often

looking for Patterns. Our results further confirm that the *Aha moment* is a quick process and it turned out to be a shorter moment than the one of a *Contradiction*, realizing that something does not fit.

We analyzed the relation of sensemaking strategies to the number of reported insights. The number of employed strategies is moderately associated with the number of reported insights (Pearson's $r = -0.441$, $p = 0.066$). Participants who reported more insights in general used fewer strategies. To follow up on our research question if sensemaking strategies are related to insight (R3), we grouped participants into three equally sized groups by the number of reported insights.

Participants reporting the smallest number of insights used the *Profiling* strategy the most. A Kruskal–Wallis test revealed a significant difference in the use of Profiling ($\chi 2 = 6.106$, $p = 0.0473$) across three different groups (Group 1: 1–3 insights, $n = 6$; Group 2: 4–6 insights; $n = 6$, Group 3: 7–11 insights, $n = 6$). Post-hoc pairwise comparison of groups using a Bonferroni adjustment revealed that the difference is significant only between Group 1 and Group 3 ($p < 0.05$) with a large effect size ($r = 0.58$). Many insights were reported when *Trends* and *Patterns* were used (compare Fig. 15.3). It is particularly interesting that participants with few insights in absolute numbers (Group 1) use the strategies of *Verification*, *Creative Desperation*, *Coincidental Aha's*, *Contradiction*, and *Pattern incl. Profiling*. This indicates that all these strategies, not only *Creative Desperation*, are related with struggle and a lack of a systematic approach. The (coincidental) Aha moment, for example, sometimes seems to indicate a contradiction that participants were not completely aware of.

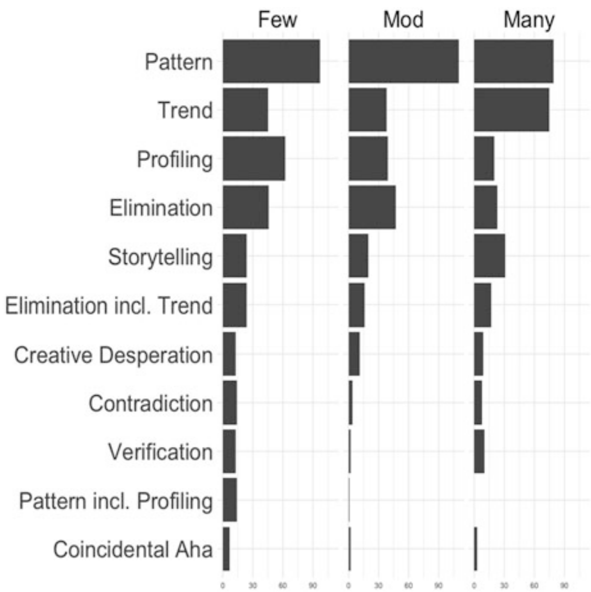

Fig. 15.3 Number of insights: *Pattern* and *Pattern incl. profiling* led to relatively few insights; *Trend* and *Storytelling* were used most when many insights got reported

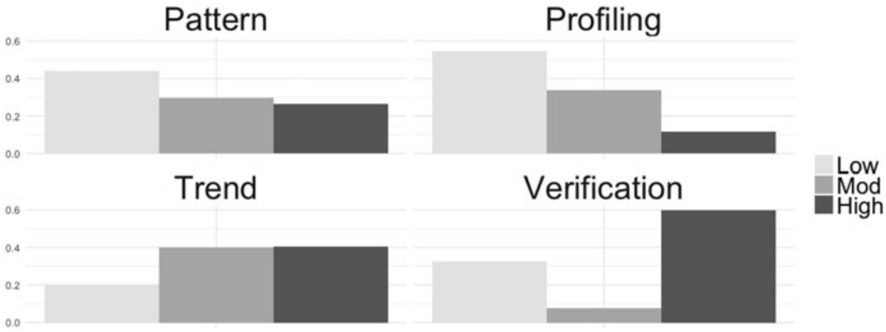

Fig. 15.4 Insights quality: *Pattern* and *Profiling* show a clear tendency of leading to low-level insights; Higher-level insights were gained using *Trend* and *Verification*

15.5.4 The Quality of Insights

A qualitative analysis reveals that applying fewer strategies, and, therefore, spending more time on one strategy, yields better insights.

The correlation between the number of strategies and the assigned insight quality is, however, not significant (Pearson's $r = -0.454$, $p = 0.058$).

Trends were assessed more often in the better groups (moderate- and high-level insights) than in then low-level group. *Patterns* and *Profiling*, on the other hand, were more often used in the low-level insight group than in the moderate- and high-level insight group. Participants with high-level insights *verified* the most. The distribution of mean usage is shown in Fig. 15.4. Pairwise comparisons using Wilcoxon rank sum test showed significant differences between high- and low-level insights through *Profiling* ($p < 0.05$) with a large effect size ($r = 0.73$) and between moderate- and high-level insights through *Verification* ($r = 0.57$). The remaining strategies were used more equally between the groups.

15.6 Discussions

We observed eleven sensemaking strategies during an explorative analysis task in the domain of criminal intelligence analysis. The analysis of strategy choices indicates that concentrating on fewer strategies leads to more hypotheses and higher-level insights. This does not imply that users should adopt as few strategies as possible. Nevertheless, we could observe that participants who changed strategies very often were confused and tried out different strategies arbitrarily. One indication of this behavior is that the time used for each strategy is very brief. Looking for connections in the data, i.e., patterns, was the predominantly used strategy. It was, however, less beneficial for insight generation than we assumed. The tools seem to support *Looking for connections (patterns)* better than *Looking for trends*. *Coincidental Aha's* occurred rather seldom.

The connected views in the system allowed to look at the data from different perspectives without further effort, i.e., without the need to repeat queries. This feature was appreciated and used often in the context of testing hypotheses. We could observe that the *Verification* strategy performed significantly better than the *Profiling* strategy in our system. *Verification* was used by participants who generated more hypotheses and consequently validated them more by using another view. Participants who did not formulate many hypotheses could not follow them up with *Verification*. This indicates a guideline for the design of visual analytics systems, so that the process of *Verification* is supported to a greater extent and, thus, to support the insight generation process. From the literature on cognitive biases, we also know that verification can help to overcome such biases [26]. Participants who concentrated on the *Profiling* strategy, on the other hand, tended to focus on details and could not create more general recommendations. The quality of their insights was, therefore, not as high. This result further suggests that visualization systems should rather emphasize looking for connections, patterns, and trends than identification of single events.

References

1. J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1963–1972, 2014.
2. K. Börner, A. Maltese, R. N. Balliet, and J. Heimlich. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization*, 15(3):198–213, 2016.
3. R. Chang, C. Ziemkiewicz, T. M. Green, and W. Ribarsky. Defining insight for visual analytics. *IEEE Computer Graphics and Applications*, 29(2):14–17, 2009.
4. D. L. Dinsmore, L. K. Fryer, and M. M. Parkinson. *Handbook of Strategies and Strategic Processing*. Routledge, 2020.
5. J. Doppler Haider, B. Gastecker, M. Pohl, P. Seidler, N. Kodagoda, and B. L. W. Wong. Sensemaking strategies in explorative intelligence analysis of network evolutions. *Behaviour & Information Technology*, 38(2):198–215, 2019.
6. D. Dumas. Strategic processing within and across domains of learning. In *Handbook of Strategies and Strategic Processing*, page 11. Routledge, 2020.
7. J. S. B. Evans. *Hypothetical thinking: Dual processes in reasoning and judgement*. Psychology Press, 2019.
8. S. N. Friel, F. R. Curcio, and G. W. Bright. Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, page 124–158, 2001.
9. A. Garnham and J. Oakhill. *Thinking and reasoning*. Basil Blackwell, 1994.
10. G. Gigerenzer. *Adaptive thinking: Rationality in the real world*. Oxford University Press, 2000.
11. G. Gigerenzer. *Gut Feelings: The Intelligence of the Unconscious*. Penguin Books, London, 2008.
12. R. R. Hoffman, B. Crandall, and N. Shadbolt. Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors*, 40(2):254–276, 1998.

13. G. Klein. *Seeing What Others Don't: The Remarkable Ways We Gain Insights*. PublicAffairs, a Member of the Perseus Book Group, New York, 2013.
14. G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, 21(4):70–73, 2006.
15. G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 2: Alternative perspectives. *IEEE Intelligent Systems*, 21(5):88–92, 2006.
16. G. Klein, J. K. Phillips, E. L. Rall, and D. A. Peluso. A data-frame theory of sensemaking. In *Expertise out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*, page 113–155, 2007.
17. G. A. Klein, R. Calderwood, and D. Macgregor. Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(3):462–472, 1989.
18. N. Kodagoda, S. Pontis, D. Simmie, S. Attfield, B. L. W. Wong, A. Blandford, and C. Hankin. Using machine learning to infer reasoning provenance from user interaction log data: Based on the data/frame theory of sensemaking. *Journal of Cognitive Engineering and Decision Making*, 11(1):23–41, 2017.
19. S. M. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3):185–225, 1989.
20. J. Lave. *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge University Press, 1988.
21. S. Lee, S.-H. Kim, and B. C. Kwon. VLAT: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):551–560, 2016.
22. P. Lemaire and L. Fabre. Strategic aspects of human cognition: Implications for understanding human reasoning. In *Methods of Thought: Individual Differences in Reasoning Strategies*, page 11–56. Psychology Press, 2005.
23. E. Mayr, M. Smuc, and H. Risku. Many roads lead to Rome: Mapping users' problem-solving strategies. *Information Visualization*, 10(3):232–247, 2011.
24. P. Mayring. *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Beltz Verlag Weinheim, 2003.
25. A. Newell and H. A. Simon. *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ, 1972.
26. A. Nussbaumer, K. Verbert, E.-C. Hillemann, M. A. Bedek, and D. Albert. A framework for cognitive bias detection and feedback in a visual analytics environment. In *Proc. European Intelligence and Security Informatics Conference (EISIC)*, page 148–151, 2016.
27. L. M. Padilla, S. H. Creem-Regehr, M. Hegarty, and J. K. Stefanucci. Decision making with visualizations: a cognitive framework across disciplines. *Cognitive Research: Principles and Implications*, 3(1):1–25, 2018.
28. W. A. Pike, J. Stasko, R. Chang, and T. A. O'Connell. The science of interaction. *Information Visualization*, 8(4):263–274, 2009.
29. P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, volume 5, pages 2–4, 2005.
30. M. Pohl, G. Wallner, and S. Kriglstein. Using lag-sequential analysis for understanding interaction sequences in visualizations. *International Journal of Human-Computer Studies*, 96:54–66, 2016.
31. D. Reisberg. *Cognition: Exploring the science of the mind.* WW Norton & Co, 1997.
32. M. J. Roberts and E. J. Newton. Strategy usage in a simple reasoning task: Overview and implications. In *Methods of Thought: Individual Differences in Reasoning Strategies*, chapter 6, page 30 pages. Psychology Press, 2005.
33. J. Saldaña. *The coding manual for qualitative researchers*. Sage, 2015.
34. P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):443–456, 2005.
35. J. M. Schraagen, S. F. Chipman, and V. L. Shalin. *Cognitive task analysis*. Psychology Press, 2000.

36. M. Schreier. *Qualitative Content Analysis in Practice*. Sage Publications, London, 2012.
37. S. Scribner et al. Thinking in action: Some characteristics of practical thought. In R. J. Sternberg and R. K. Wagner, editors, *Practical Intelligence: Nature and Origins of Competence in the Everyday World*, volume 13, page 60. Cambridge University Press, 1986.
38. E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010.
39. H. A. Simon and A. Newell. Human problem solving: The state of the theory in 1970. *American Psychologist*, 26(2):145, 1971.
40. B. Tversky. Visuospatial reasoning. In *The Cambridge Handbook of Thinking and Reasoning*, page 209–240. Cambridge University Press, 2005.
41. S. Woll. *Everyday thinking: Memory, reasoning, and judgment in the real world*. Psychology Press, 2001.

Index

D. Albers Szafir et al. (eds.), *Visualization Psychology*,
https://doi.org/10.1007/978-3-031-34738-2